Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros de Telecomunicación

Improving Mental and Emotional Wellbeing based on Intelligent Techniques. Application to the case of Smart Offices

Tesis Doctoral

Sergio Muñoz López
Ingeniero de Telecomunicación

2023

Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros de Telecomunicación

Improving Mental and Emotional Wellbeing based on Intelligent Techniques. Application to the case of Smart Offices

Tesis Doctoral

Sergio Muñoz López
Ingeniero de Telecomunicación

2023

Departamento de Ingeniería de Sistemas Telemáticos

Escuela Técnica Superior de Ingenieros de Telecomunicación

Universidad Politécnica de Madrid

Improving Mental and Emotional Wellbeing based on Intelligent Techniques. Application to the case of Smart Offices

Autor:
Sergio Muñoz López
Ingeniero de Telecomunicación

Tutor:
Carlos A. Iglesias
Doctor Ingeniero de Telecomunicación

2023

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día ________ de ________________ de __________.

Presidente: ____

Vocal: ____

Vocal: ____

Vocal: ____

Secretario: ____

Suplente: ____

Suplente: ____

Realizado el acto de defensa y lectura de la Tesis el día ________ de ________________ de __________. en la E.T.S.I. Telecomunicación habiendo obtenido la calificación de ____.

EL PRESIDENTE LOS VOCALES

EL SECRETARIO

A mi familia.

Agradecimientos

En primer lugar: a mis padres, todo lo que soy y lo que tengo se lo debo a ellos. También a mis hermanas, que han sido el ejemplo a seguir durante toda mi vida y me han allanado el camino, haciendo que todo sea más fácil. A mis sobrinos, que llenan mi vida de emoción e ilusión. A Linda, por su apoyo y por «mettermi in riga», tener al lado a una persona así es un regalo.

A Carlos Ángel, por su extraordinaria paciencia y dedicación. Su papel como director ha sido fundamental para poder realizar esta tesis, y es un lujo poder aprender y admirar cada día su trabajo. También quiero darles las gracias a mis compañeros y amigos del Grupo de Sistemas Inteligentes: Álvaro, Fernando y Óscar. Estar con ellos es un aprendizaje y divertimiento continuo.

Durante mi estancia en Trento, tuve la suerte de conocer a personas increíbles que han sido muy importantes en esta tesis. Quiero agradecer a Óscar y a Venet, por su acogida en la Fondazione Bruno Kessler, donde tuve la oportunidad de disfrutar y aprender tanto durante los meses de estancia. También quiero agradecer a todos los amigos que conocí allí y que contribuyeron a que la experiencia me cambiara la vida: Lorenzo, Marco, Flavio, Sara, Hugo, Pietro, Alessandro, Giulio, Nantine, Tommy, Toni, Elly, Alex, Sarah, Stefano, y todos los demás. Y por último, doy las gracias también a mi familia trentina: Mauro, María, e Isabella, que con su hospitalidad me hacen sentir como en casa.

Resumen

El lugar de trabajo es uno de los entornos que más repercuten en la salud psicológica de la población, y se ha demostrado que la inversión en el bienestar laboral produce beneficios sustanciales para empleados y empresas. La evolución de la tecnología ha promovido el desarrollo de nuevos enfoques para la promoción de la salud emocional y mental, y esta tesis pretende avanzar en la aplicación de estos enfoques al lugar de trabajo. En concreto, se ha tratado de avanzar en tres líneas diferentes: la detección automática del estado psicológico mediante métodos no intrusivos; el análisis efectivo de estrategias de promoción del bienestar mental y emocional; y la adaptación automática e inteligente del entorno.

En cuanto a la detección automática del estado psicológico, esta tesis contribuye con el diseño de soluciones no intrusivas basadas en aprendizaje automático para el reconocimiento de trastornos mentales. Por un lado, se ha avanzado en la detección de trastornos mentales a partir del texto con una solución para detectar el estrés que combina características extraídas de léxicos con representaciones distribucionales. Por otro lado, se ha diseñado un sistema para predecir el nivel de estrés actual de un individuo utilizando datos de su estrés circundante, es decir, sus niveles de estrés anteriores junto con el de sus compañeros cercanos. Ambas soluciones mostraron resultados positivos durante la evaluación experimental.

En cuanto a la segunda línea, esta tesis contribuye con el diseño de un modelo de agentes para el estrés laboral basado en las condiciones ambientales y laborales. El modelo se ha integrado en un sistema de simulación, permitiendo la evaluación de diferentes políticas de promoción de la salud mental. La fiabilidad del sistema ha sido validada mediante varios experimentos que incluyen la experimentación en el laboratorio y el análisis de sensibilidad.

Por último, esta tesis avanza en la investigación sobre la adaptación inteligente del entorno. Se ha diseñado una arquitectura para una plataforma de automatización de tareas que integra mediante tecnologías semánticas sensores y actuadores orientados a la detección y regulación de emociones. De esta forma, la plataforma mejora el bienestar de los empleados mientras que garantiza la interoperabilidad y la escalabilidad de todos sus componentes. Se ha observado que dicha plataforma tiene resultados positivos en la evaluación experimental, aumentando el bienestar y la productividad de los usuarios.

Abstract

The workplace is one of the environments that most impacts on population psychological health and well-being, and there is growing evidence that investment in work-related well-being derives substantial benefits for employees and companies. The rapid development of technology has fostered the development of new approaches to the promotion of emotional and mental health. Thus, this thesis aims to advance in the research of this field and its application to the workplace. Specifically, we sought to advance the research in three different lines: automatic psychological awareness using non-obtrusive methods; analysis of mental and emotional well-being promotion strategies; and smart environment adaptation.

Regarding automatic psychological awareness, this thesis contributes with advancements in unobtrusive and economical solutions for automatic mental illness recognition based on machine learning. On the one hand, we further early mental disorder detection from text with an approach to detect psychological stress combining a lexicon-based feature framework with distributional representations. On the other hand, we have designed a machine learning approach to predict the current stress level of an individual using their surrounding stress-related data, that is, their previous stress levels along with stress levels from their close colleagues. Both solutions showed positive results during the experimental evaluation.

As for the second line, this thesis contributes with the design of an agent model for occupational stress based on ambient and work conditions. The model has been integrated into a simulation system, enabling the evaluation of different mental health promotion policies at work. The reliability of the system has been validated through several experiments including in-lab experimentation and sensitivity analysis.

Finally, this thesis advances in the research on the intelligent adaptation of the environment. An architecture has been designed for a task automation platform that integrates through semantic technologies sensors and actuators oriented to the detection and regulation of emotions. In this way, the platform promotes the well-being of employees while ensuring the interoperability and scalability of all its components. It has been observed that this platform has positive results in the experimental evaluation, increasing the well-being and productivity of users.

Contents

Reader's guide

This thesis is presented as a compilation of publications. As a result, the purpose of this document is not to serve as a comprehensive monograph, but rather as a collection of the author's academic works that meet specific criteria. In order to comply with the guidelines set by the institution, at least three of these publications must have the Ph.D. student listed as the first author and must have been published in Q1 or Q2 ranked journals, as indexed by JCR or Scopus. The reader can find a list of these publications and their compliance with these requirements in Chapter 3 and Appendix A.

The structure of the document is also in accordance with the guidelines established by the Universidad Politécnica de Madrid. This section aims to provide a guide for the reader to navigate through the contents of the document.

Chapter 1 serves as an introduction to the context in which the research was conducted. It describes the motivation (Section 1.1) and the background of the research (Section 1.2), including a review of relevant literature and an examination of current practices in the field. The research hypotheses and objectives for this study are also presented (Sections 1.3 and 1.4). This chapter sets the stage for the subsequent chapters, which will delve into the methodology and results of the thesis.

Chapter 2 provides a detailed and systematic description of the research approach followed in order to arrive at the research findings and conclusions presented in the subsequent chapters. It presents the methodology used to conduct and evaluate the outcomes of this thesis, the phases of the research, some strategies followed on research methods, and how and under which criteria the different outcomes are evaluated.

Chapter 3 presents the core papers published as part of this Ph. D. thesis. These papers represent the primary research contributions of this work and demonstrate its novelty and significance. Each paper is presented in a separate section, where the full text is preceded by a short table with relevant information (title, authors, etc.). Those publications by the author that do not directly contribute to the objectives of this thesis have not been included in this chapter, but in Appendix A, where the complete list of relevant publications can be found.

Chapter 4 intends to provide a comprehensive overview of the work presented in the preceding chapters. It begins by providing a general overview of the solutions proposed in the thesis and an analysis of the results obtained (Section 4.1). It then presents a discussion of the main findings of the research describing their implications and significance. Finally, the chapter concludes with a description of possible lines for future research that can build upon the work presented in this thesis 4.5. Overall, the purpose of this chapter is to reflect upon the research conducted and to provide a summary of the contributions made to the field.

Finally, the acronyms and abbreviations used along the document are described in Chapter 4.5.

CHAPTER 1

Introduction

The first chapter of this Ph. D. thesis serves as an introduction to the context in which the research was conducted. It describes the motivation and the background for this research, including a review of relevant literature and an examination of current practices in the field. The research hypotheses and objectives for this study are also presented. This chapter sets the stage for the subsequent chapters, which will delve into the methodology and results of the thesis.

1.1 Motivation

Well-being is essential to population health and furthers the functioning and life quality of individuals, families, communities, and society. This term encompasses today, in addition to physical health, aspects related to the psychological conditions such as emotional and mental health of individuals (De Simone, 2014). Actually, the assessment of psychological health plays a key role in the promotion of population well-being (Slade, 2009). Current psychological health assessment strategies focus on fostering supportive environments in everyday contexts or settings where mental and emotional conditions can be promoted, such as homes, schools, or workplaces (Barry et al., 2019).

Among these contexts, this thesis focuses on the promotion of psychological well-being in the workplace. The workplace is one of the environments that most impacts the population's psychological health and well-being (Hafner et al., 2015). Consequently, the analysis of the role of work in promoting mental and emotional well-being has gained momentum. There is growing evidence that investment in work-related well-being derives substantial benefits for employees and companies, such as increased commitment, job satisfaction, and improved performance (Krekel, Ward, and De Neve, 2019). However, strategies in the workplace specifically aimed at addressing the promotion of well-being and the prevention of emotional and mental health issues are not very common yet (Barry et al., 2019). In addition, the recent changes in the work nature have increased workload demands and job insecurity, which contribute to adverse emotional and mental health outcomes (Appelbaum et al., 2000).

Fortunately, the major advances in the fields of ambient intelligence and affective computing have opened up a range of possibilities for the improvement of emotional and mental health. Indeed, the application of these technologies to enhance well-being has received significant momentum in recent years and has gained widespread traction in the field of psychological health. Nevertheless, existing solutions often have documented limitations (Peters, Calvo, and Ryan, 2018). Current wellness regulation strategies based on existing technologies are often impersonal and limited in terms of interaction and personalization. They usually lack crucial elements such as simplicity and ease of use, integrated transdiagnostic capabilities, real-time interactive engagement, or personalized assessment (Santarossa et al., 2018).

This thesis aims to tackle this issue and advance the research of technology-assisted strategies for enhancing mental and emotional well-being in the workplace. In this context, three fundamental foundations have been identified: recognition and awareness of psychological well-being; effective implementation of psychological well-being promotion policies;

and support for early interventions (Harnois et al., 2000).

These bases pose, however, some strong challenges for companies. Firstly, psychological health awareness can not always be guaranteed. The stigma associated with mental illness is still a problem, and many employees are diffident to expose their concerns at work (Szeto and Dobson, 2010). This can delay or even hamper the detection of work-related mental and emotional disorders. Secondly, knowledge regarding the effectiveness and suitability of interventions is still debouching. Existing gaps in the study of robust workplace well-being promotion and prevention interventions can hinder their effective implementation and make them vulnerable to misleading design or cost increases (Barry et al., 2019). Finally, early interventions require the seamless integration of the recognition and regulation processes. Nevertheless, traditional methods for the detection and regulation of mental disorders usually entail significant delays (Alberdi, Aztiria, Basarab, and Diane J Cook, 2018).

Therefore, we aspire to advance the research on technology-assisted well-being promotion at work in three different lines: automatic psychological awareness using non-obtrusive methods, analysis of mental and emotional well-being promotion strategies, and smart environment adaptation.

Concerning the first line, methods for automatic detection and recognition of psychological issues have traditionally consisted of self-reports in response to standardized questionnaires such as PHQ-9 (Spitzer et al., 1999) or PSS (Cohen, Kamarck, Mermelstein, et al., 1994). Whereas the validity of these methods has been established (Gilbody, House, and Sheldon, 2001), they present two main shortcomings: their subjectiveness and reliance on individuals' recall and condition awareness, and their typical retard in diagnosis, which makes them inadequate for timely detection (Alberdi, Aztiria, Basarab, and Diane J Cook, 2018). Hence, there is still a need for objective procedures that use physical and physiological information to automatically recognize psychological-related issues or disorders (Greene, Thapliyal, and Caban-Holt, 2016).

The enormous strides in affective computing unlock a variety of potential directions to fill this gap. A significant deal of research effort has been put in recent years into techniques and systems that employ smart devices and affective computing algorithms for automatic mental health assessment. Among the most popular approaches are those consisting of the analysis of physiological signals (Rahim et al., 2019), facial expression (Giannakakis et al., 2017), or speech (Liu et al., 2015). The reliability of these methods has been proved (Graham et al., 2019), and they succeed in avoiding the subjective response bias from the subjects, which is commonly present in self-report questionnaires. Nonetheless, their intrusiveness and elevated implementation costs are frequently a significant limitation (Novais and Carneiro, 2016).

In this regard, this thesis intends to further the research on non-obtrusive and economical solutions for automatic mental awareness. With this aim, two different approaches are proposed. On the one hand, this thesis aims to exploit the vast quantity of publicly available textual data to detect the psychological state. Social media platforms where users express and share their reflections or experiences have become widely used and offer vast information about people's feelings and worries (Pang et al., 2019). Thus, textual data from these services can be used as an information source for the detection of mental disorders (Chancellor and De Choudhury, 2020). In this line, this thesis contributes to the topic of early mental disorder detection from text with an approach to detect psychological stress based on the combination of a lexicon-based feature framework with distributional representations.

On the other hand, existing evidence regarding the influence that affective states emanated from the individual have in the mood of close persons is examined (Petitta et al., 2021). With this purpose, a machine learning approach to predict the current stress level of an individual using their historical data along with data from their close colleagues is proposed. In this manner, this thesis contributes with advancements in unobtrusive and economical solutions for automatic mental illness recognition based on machine learning.

As for the second line, the analysis of mental and emotional well-being promotion strategies often entails a major challenge: the evaluation of their effectiveness in a real scenario. This challenge arises from two main issues: the complexity of human behaviour and the high economic costs usually associated with the deployment of these interventions in a real scenario (McDaid, Curran, and Knapp, 2005).

This thesis aims at contributing to the cost-reduced analysis and study of emotional and mental health promotion policies at work using agent-based simulation. Agent-based simulation technologies have a well-established role in the study of human behaviour (Sobkowicz, 2016), as they offer a cost-effective and smooth approach to analyzing a wide range of behavioural and psychological dimensions or parameters. This is especially helpful when there are heterogeneous populations or complex behaviours (Bonabeau, 2002). However, the reliable design of agent behaviour models is still an open challenge, given the vast number of factors involved in human behaviour. These variables are frequently challenging to calibrate, quantify, and even justify (Elkosantini, 2015a).

To solve these challenges, this thesis aims at progressing the study of work-related factors influencing mental illness and integrating this knowledge into an agent-based simulation system. To this aim, a model for occupational stress estimation is designed taking into account environmental, work, personal and social-related conditions. The reliability of the model is validated through several experiments, including the use of public datasets, in-lab experimentation, and sensitivity analysis. Besides, the development of an agent-based simulation system integrating this model enables the cost-effective and effortless evaluation of different mental health promotion strategies in the workplace.

Finally, this thesis aims to advance the research on smart environment adaptation. Recognition and intervention processes must be seamlessly and continuously integrated to guarantee the efficacy and efficiency of well-being promotion strategies. To this end, the environment must be adaptive to workers' moods and needs. A popular approach to address this issue is the application of ambient intelligence principles to the work environment. The equipment of everyday objects and physical devices with Internet connection capabilities opens endless possibilities to enhance people's welfare through the reaction and adaptation of the environment in response to the context (Sovacool and Del Rio, 2020). Accordingly, the application of these technologies in workplaces allows companies to promote the satisfaction and productivity of their workers by supporting and meeting their needs (Prasetyo, Zaliluddin, and Iqbal, 2018).

However, the ability to seamlessly exchange data and functionality between diverse platforms and components has become a major issue in intelligent environments (Mouromtsev, 2021). The information in these environments is transferred among multiple services and devices through heterogeneous communication technologies and network protocols (H. Rahman and Hussain, 2020). Therefore, smart environments require a method to clearly represent information in an unambiguous and well-defined manner that permits its sharing and correct understanding among independently developed services or applications (Palavalli, Karri, and Pasupuleti, 2016).

A popular approach to address the interoperability challenge is the use of semantic modeling. This approach involves the creation of formal definitions of concepts and relationships known as ontologies that substitute ad-hoc application-centric representation models. Ontologies provide a standardized representation of inferred data, allowing for the seamless exchange of information among diverse systems and promoting a deeper comprehension of the environment. Furthermore, semantic representation can significantly enhance the scalability of a system by providing a machine-readable format that is rich in detail and can be easily understood, reasoned about, and reused (Bikakis et al., 2007).

These capabilities have led to a growing trend of integrating semantic technologies into IoT-based ecosystems, and more particularly into smart environments (Shi et al., 2018). The benefits of semantic technologies have been exploited in a variety of research works related to smart environments, even for the particular case of smart offices (Coronato, De Pietro, and

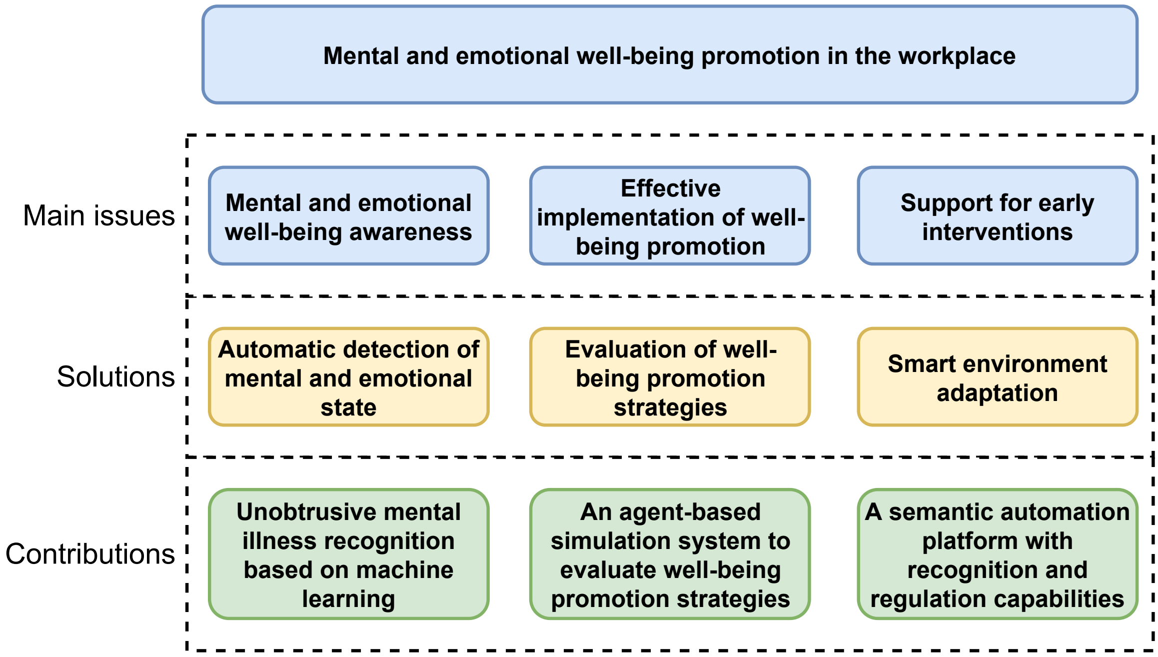


Figure 1.1: Overview of the context of this thesis outlining the main issues involved in well-being promotion at work along with the proposed solutions and contributions.

Esposito, 2006). Nevertheless, there is a gap in the literature concerning the integration of semantic technologies with psychological awareness and promotion into a unified approach for intelligent workplaces.

To fill this gap is the direction in which the contributions of this thesis point. With this purpose, we propose an emotion-aware task automation platform that integrates through semantic technologies sensors and actuators oriented to the detection and regulation of emotions. The platform is based on a semantic vocabulary built on the basis of popular ontologies regarding emotions (Sánchez-Rada and Carlos A. Iglesias, 2016), intelligent devices (Janowicz et al., 2019), and automation (Coronado, Carlos A Iglesias, and Serrano, 2015). In this manner, the intelligent platform not only can help promote the well-being of employees thanks to its mental and emotional aware capabilities but also guarantees the interoperability and scalability of all system components.

To sum up, this thesis aims at exploring the use of technology to assist psychological well-being promotion in the workplace. To this aim, the thesis proposes different solutions to address the main issues involved. An overview of the context and issues to address proposed solutions and contributions is given in Fig. 1.1.

1.2 Background

In this thesis, we build upon recent advancements in the field of artificial intelligence, specifically within the subfield of ambient intelligence, as our starting point for our research. Ambient Intelligence (AmI) consists of the use of hardware devices and software technologies to create intelligent and adaptive environments that are sensitive and responsive to people. The environments dotted with ambient intelligence support their inhabitants in carrying out their daily activities in an intelligent, interconnected, flexible, dynamic, and unobtrusive fashion (Altieri, Ceccacci, and Mengoni, 2019). This can include detecting and responding to human needs, activity, or even psychological factors to increase comfort and well-being (Acampora, Diane J. Cook, et al., 2013).

AmI integrates two major research lines: recognition and intervention. The former involves gathering context and user-related data and the application of affective computing technologies to recognize the psychological state of the users. The latter consists of acting on the environment and adapting it to the users' preferences and needs in order to promote their well-being. This can be approached from different perspectives, such as personalization and tailoring of experiences, smart social support and assistance, or automatic interventions for mental and emotional regulation.

Besides, AmI draws on different areas and technologies, such as multiagent systems and semantic technologies (Sadri, 2011). On the one hand, agent-based simulation is commonly used in AmI applications to help with distributed information and problem-solving. On the other hand, semantic technologies are very useful for the representation of meaning and knowledge of the environments, enabling the semantic modeling of entities, activities, and their relations (Gams et al., 2019). As a work based on ambient intelligence, this thesis is also grounded in these areas and technologies: mental and emotional awareness, automatic interventions, agent-based simulation, and semantic technologies. The following sections provide further insight into the specific background of each of these concepts.

1.2.1 Mental and emotional awareness

One key aspect of AmI is the ability to recognize and respond to the affective state of the people in the environment. To this end, AmI systems must be equipped with emotional capabilities and be able to adapt and respond to users' mental and emotional states (Acampora and Vitiello, 2013). This has led to the notion of emotion-aware ambient intelligence (AmE) (J. Zhou and Kallio, 2005), which aims at fostering the development of emotion-aware services in pervasive AmI environments to provide people with automatic psychological support. AmE draws upon principles from psychology and social sciences to identify the affective state of individuals, enabling the development of more effective and tailored ambient intelligence systems and empowering users with a cluster of services to promote and improve their psychological well-being (Castillo et al., 2016).

The first step in making a space adapt to the psychological state of the users is sensing, that is, the detection and recognition of the psychological state. Sensing involves the process of gathering affective data from individuals and the application of techniques that exploit these data to identify their psychological state. Over the last few years, mental and emotional awareness has been gaining the attention of the scientific community, given its broad applications in several fields, such as health, commerce, and businesses. Traditional approaches in the field mainly consisted of self-reports in response to standardised questionnaires. However, the emergence of affective computing techniques has opened a range of possibilities for automatic awareness of emotional and mental states (Picard, 2000).

So far, research in the field has evolved in two primary directions: the analysis of physiological data and the analysis of behavioural data (Alberdi, Aztiria, and Basarab, 2016; K. Yang et al., 2021). The notion of wearables that detect the affective state of an individual from biometrical signals goes back to the beginnings of affective computing (Picard and Healey, 1997). Physiological data can furnish objective insight into the psychological condition of an individual, and a broad range of physiological signals have been explored (Shu et al., 2018). These signals emanate from the central nervous and autonomic nervous systems of the human body, both of which are linked with inner mental and emotional states.

Among them, Electro-Dermal Activity (EDA) (Pakarinen, Pietilä, and Nieminen, 2019), Heart Rate Variability (Castaldo et al., 2019), and Electroencephalogram (EEG) (Song et al., 2018) have yielded the most positive outcomes. For instance, sweating skin is related to high-stress levels and other affects and leads to skin conductance changes that can be detected through the appropriate sensors. Thus, Electro-Dermal Activity can be used as an indicator of arousal. Heart rate is another example of a physiological signal linked to mental conditions: its variability varies depending on the arousal level. In general terms, heart rate is lower for pleasant and low arousal stimuli and higher for unpleasant and high arousal stimuli (Brouwer et al., 2013). Some other examples of signals include blood pressure (BP) (Gordon and Mendes, 2021), respiration (Sadat-Mohammadi et al., 2021), blood volume pulse (BVP) (Handouzi et al., 2014), eye gaze, blinking or pupil diameter (PD) (Skaramagkas et al., 2021).

As for behavioral data analysis, this approach exploits variations in individuals' behaviour to assess mental or emotional conditions (Areàn, Ly, and Andersson, 2022). These methods comprise the analysis of a wide range of behavioral signals such as facial expressions (J. Zhang et al., 2019), speech (Koolagudi and Rao, 2012), or activity (Giakoumis et al., 2012).

Whereas the reliability of these methods for recognizing emotions and mental issues has been proved (Graham et al., 2019; Saxena, Khanna, and D. Gupta, 2020), their implementation in real-life settings entails further challenges. Analysing these data often requires optimum technological conditions, but everyday activities can decrease the signal quality or lead to altered body responses (Han et al., 2020). As a consequence, errors due to detached equipment, movements, or incorrect placement are common in daily life and lead to corrupted data (Can, Arnrich, and Ersoy, 2019). Additionally, some solutions require a large quantity of data or the installation of costly and obtrusive sensing equipment (Novais and Carneiro, 2016). This decreases the feasibility of these approaches being implemented in real-world settings.

This challenge has encouraged research on more economical and unobtrusive solutions, including the use of computer patterns (e.g., mouse or keystroke dynamics) (Dacunhasilva, Z. Wang, and Gutierrez-Osuna, 2021), smart textiles (Scheuermann et al., 2020), the use of mobile phones (Ferdous, Osmani, and Mayora, 2015), or textual data (Lin et al., 2016). In this thesis, we are particularly interested in using textual data, given the wide variety of such data available in the workplace (e.g., in the form of e-mails, messaging systems and forums, or reports). The use of textual data has received increasing traction in recent years, and algorithms to predict emotions from text are mature and considered accurate. This has been partly motivated by the increasing popularity of social media services that offer a huge quantity of information related to people's emotions, feelings, and worries (Pang et al., 2019). Textual data from these services have been employed with positive results in a broad range of text classification applications such as detection of sentiment (Yue et al., 2019) or mental illness (Chancellor and De Choudhury, 2020; Banerjee and Shaikh, 2021; Cao et al., 2021). The most prominent approaches in the literature targeting text classification are lexicon-based solutions and machine learning methods (Bandhakavi et al., 2017).

The lexicon-based approach relies on predefined lists of words or phrases that are associated with specific classes. This approach compares words in the text with a dictionary to compute the frequency of appearance of certain bearer terms. Lexicons are able to provide an overall indication of specific characteristics (e.g., emotion, sentiment, or topic) on the basis of the words they contain (Khoo and Johnkhan, 2018). This approach tends to be computationally faster and easier to implement than machine learning methods. Besides, it has achieved positive results on a broad range of text classification applications based on texts from social media (Giuntini et al., 2020; Almatarneh and Gamallo, 2018). Nevertheless, their reliance on predefined lists of words makes them less flexible, as they are often limited by the words included in the lexicon (Hartmann et al., 2019). Also, their performance is usually lower than that achieved with machine learning methods, as they may not be able to handle variations in language (e.g., spelling mistakes or idiomatic expressions) or successfully identify sarcasm or irony (Weitzel, Prati, and Aguiar, 2016). These factors can be more easily handled by machine learning methods able to learn patterns and features in the text.

Machine learning methods involve training a model on a labeled dataset and using it to predict the class of new, unseen texts. To this end, they require a significant amount of labeled training data. Also, these methods are able to automatically learn patterns and features in the text, what makes them highly flexible, and can handle large amounts of complex data (Hartmann et al., 2019). Popular techniques to implement machine learning methods include the use of deep neural networks, word embeddings, or even lexicon-based features. As previously described, lexicon-based features can help to provide a general indication of a text in specific terms, but they fail to capture more subtle qualities and contextual cues inherent to the human language (Giatsoglou et al., 2017). The use of word embedding techniques helps to overcome this challenge, as they enable the encoding of semantic and syntactic features present in words and represent them in a vector space. These representations are called pre-trained word vectors and are very useful for the representation of textual data in text classification applications (S. Wang, W. Zhou, and Jiang, 2020). Word2Vec (Mikolov et al., 2013), GloVe (Pennington, Socher, and Manning, 2014) and FastText (Joulin et al., 2016) are some examples of widespread word embedding techniques.

Furthermore, previous research has demonstrated that the combination of word embeddings with lexicon-based features can improve the quality of text representations and can lead to better prediction performance (Fu et al., 2018). This combination is exploited in this thesis for the detection of stress in texts from social media, whose results are published in **"A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representation"** (Section 3.2.1). Such work hypothesizes that the combination of word embeddings with a framework of lexicon-based features can yield positive results for stress detection from texts and validates it on public English datasets.

Finally, there are certain aspects related to the nature of affective states that existing approaches have not fully exploited. There is a considerable amount of research on the influence of affective states such as stress (Dimitroff et al., 2017) and emotions (Petitta et al., 2021) on an individual's mood, as well as the role that data related to past emotions can play in predicting future moods (Hollis et al., 2017). Our work presented in **"Prediction of stress levels in the workplace using surrounding stress"** (Section 3.2.2) exploits the contagious nature of stress to predict stress levels of an individual. Such a system opens additional possibilities for multimodal approaches that combine different methods for mental illness or emotion recognition, as it decreases the quantity of data required and improves the effectiveness in the presence of scarce data.

1.2.2 AmI-based interventions for psychological well-being promotion

Affective factors are crucial for everyday human functioning, as they support the response of individuals to certain situations. Positive mental and emotional conditions are correlated with healthy patterns and are associated with better life quality. However, when these conditions are not suitable for a given situation, they can lead to adverse outcomes or even be harmful (Gross, 2013). For example, negative affects such as anger, stress, or anxiety may increase human susceptibility and vulnerability to illness (Cohen, Doyle, et al., 1995).

In addition, people's mental and emotional condition will affect other aspects of their life, such as social relationships or professional performance. It is widely recognized in the working environment today that affective factors can positively or negatively influence employees' well-being and performance (Zapf et al., 2021). In general terms, positive emotions tend to be beneficial and contribute to a productive work environment, while negative emotions tend to worsen well-being and performance (Lyubomirsky, King, and Diener, 2005). The importance of psychological factors has fostered research into pervasive environments with mental and emotional regulation capabilities. These environments aim at improving people's well-being by providing affective services intended to promote positive mental and emotional conditions (J. Zhou and Kallio, 2005).

Emotion regulation consists of the alteration or management of the processes involved in the generation, experience, or expression of emotions (Campos, Frankel, and Camras, 2004). It implies the ability to recognize and understand one's own emotions and manage and express them in an effective and healthy way. Given the impact of emotions on life, their regulation is considered essential to health and well-being throughout the life span. Furthermore, it is a critical component of psychological well-being and effectual social functioning, as it can affect an individual's ability to perform tasks and make decisions.

Historically, research on emotion regulation has been based on the use of static self-report measurement techniques. Nevertheless, in recent years, the focus has shifted to the dynamic nature of emotion regulation processes. The most prevalent framework for the analysis of emotion regulation is the process model proposed by Gross (2015). This model involves four phases that constitute an ongoing cycle: identification of the mental conditions to regulate, selection of an appropriate regulation strategy, implementation of the strategy, and monitoring of the outcomes.

Therefore, there has been a growing body of literature that addresses which strategies are more effective in helping people to better regulate their mental and emotional conditions. The strategies proposed range from ancient relaxation methods based on mindfulness (e.g., breath control or yoga) to more modern techniques of emotion induction (Pedrini et al., 2022). The most studied emotion induction techniques involve the use of music (Moore, 2013), color (Sokolova et al., 2015), and lighting (Sroykham, Wongsathikun, and Wongsawat, 2014). Regarding music, the most effective and frequently studied musical cues in terms of their influence on emotions are mode and tempo. With regard to color and lighting, research highlights chroma and light temperature to produce an impact on emotions (Sroykham, Wongsathikun, and Wongsawat, 2014). Finally, the influence of ambient conditions such as temperature on emotion regulation has also been studied (T. Zhang et al., 2021).

The recent proliferation of digital technologies employed in the field of mental health research has facilitated the implementation and monitoring of emotion regulation strategies. Technology-based interventions are based on the theoretical frameworks of emotion regulation but have a unique potential to overcome the major challenges existing in emotion-regulation research (Bettis et al., 2022). A popular approach is the use of applications that enable mood monitoring and direct intervention by goading users to be involved in a specific activity or offering assistance and support. The positive results demonstrated by these solutions for the treatment of mental health-related issues such as anxiety or stress (Loo Gee, Griffiths, and Gulliver, 2016) have boosted the development of mental health-related apps (Wasil et al., 2019). The popularity of some of these apps (e.g., Calm or Pause) speaks for itself in relation to their effectiveness in promoting mental and emotional well-being.

A step forward in these solutions is to leverage recent advances in natural language processing and language generation models for the development of social conversational agents. For example, the capabilities of social robots or chatbots to provide mental health assistance and support or improve users' mood have been demonstrated (Šabanović et al., 2013), and solutions exploiting these technologies have proliferated. For example, the work proposed by Pham et al. (2021) presents an emotion regulation system consisting of a robot assistant that engages in an interactive conversation to distract and support the user in managing negative emotions.

In addition, technological developments have enabled the delivery of automatic interventions that do not require direct participant involvement. These approaches are based on smart systems that benefit from ambient intelligence technologies to detect the mental and emotional conditions of an individual and automatically adapt and adjust the environment to regulate them. Adjustments in ambient color, lighting, and music are likely to be the basis on which these solutions are founded. For instance, Fernández-Caballero et al. (2016) proposed a framework for a smart environment that automatically recognizes and regulates individuals' emotions through music and color. Similarly, Ortiz-García-Cervigón et al. (2015) proposed an emotion regulation system based on RGB LED strips whose color and intensity are adjusted to control the ambiance. Their study found that warm colors were appraised as more tense, hot, and less preferable for lighting, while cold colors were appraised as more pleasant.

The extensive research in the field demonstrates that smart environments have an outstanding potential to promote mental and emotional well-being through the automatic detection and regulation of psychological conditions. In these systems, there exists a variety of techniques for implementing the adaptation and modification of the environment in response to the context, such as ECA (event-condition-action) rules, decision trees, and case-based reasoning (Sadri, 2011). Among them, ECA rules constitute one of the most popular approaches in the development of intelligent environments to improve well-being and health, given their straightforward interpretation (Augusto et al., 2008; Leong, Ramli, and Perumal, 2009). These rules consist of events and actions in a way that certain actions should be taken upon the detection of certain events if certain conditions are satisfied.

The use of smart environments to automatically regulate mental and emotional conditions is a key part of this thesis. Accordingly, the concepts described in this section are used extensively in the following publications: **"An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices"** (Section 3.2.6), **"An Emotion-Aware Learning Analytics System Based on Semantic Task Automation"** (Section 3.2.5) and **"An agent based simulation system for analyzing stress regulation policies at the workplace"** (Section 3.2.4).

1.2.3 Semantic technologies

The previous sections have described the potential of AmI-powered environments to improve mental and emotional well-being through the detection and regulation of psychological conditions. Research has shown the effectiveness of smart environments in enhancing the comfort and well-being of people. However, a traditional problem when developing this kind of system lies in finding a common way to represent information and knowledge. The lack of a unified model to represent the knowledge these systems manage and share makes it challenging to integrate new components (e.g., devices or services) or adapt them to new scenarios.

Thus, the establishment of a standard set of protocols or rules that delineate information sharing and exchange among independent systems is essential. These protocols should outline expected schemas, formats, and behaviour. One of the most popular solutions to address this issue is the adoption of semantic technologies to express domain information. Semantic technologies help to design a standard definition of concepts and relationships that substitutes application-centric representation models and formats. Such definitions furnish a machine-readable format that is both robust and easily understandable and can be reasoned about and reused.

As the Semantic Web has gained popularity, there has been a notable growth in related technologies such as languages, standards, and reasoners (Gams et al., 2019). The most popular standard to define how data should be organized and interchanged is Resource Description Framework (RDF) (Decker et al., 2000). The RDF provides a straightforward and general-purpose model for representing information based on subject-predicate-object statements. For example, let us consider the expression "Edward feels sad". Here, "Edward" is the subject, "feels" is the predicate, and "sad" is the object. We could represent this information along with other related information using RDF triples as:

```
<http://example.com/edward> <http://example.com/born-in> <http://
    example.com/
usa> .
<http://example.com/edward> <http://example.com/work-as> <http://
    jobs.test/
artist> .
<http://example.com/edward> <http://example.com/feels> <http://
    emotions.test/
sad> .
```

Listing 1.1: Example of RDF

Listing 1.1 encodes the following information: Edward is an artist born in the USA who feels sad. Note that to adhere to the Linked Data principles, each element of triples must have a URI that is addressable and unique to each specific concept.

Besides, RDF model can be extended to support more sophisticated ontological representations. A popular extension is Notation 3 (also known as N3), an assertion and logic language that extends the RDF data model by adding formulae, variables, logical implication, and functional predicates (Berners-Lee and Connolly, 2008). In addition, it also provides a textual syntax alternative to RDF/XML that significantly enhances its readability. The previous example could be represented using Notation3 in a more clear way, as shown in Listing 1.2.

```
@prefix ex: <http://example.com/> .
@prefix jobs: <http://jobs.test/> .
@prefix emotions: <http://emotions.test/> .

ex:edward ex:born-in ex:usa ;
    ex:work-as jobs:artist ;
    ex:feels emotions:sad .
```

Listing 1.2: Example of RDF in Notation3

The formulae and logical and functional capabilities provided by Notation3 enable the representation of rules. This allows semantic reasoners to automatically infer new knowledge based on existing information. For example, in the previous example, it would be logical to assume that if Edward was born in the USA, he speaks English. We could benefit from the logical capabilities of Notation3 to express this rule, as shown in Listing 1.3.

```
@prefix ex: <http://example.com/> .

{
    ex:edward ex:born-in ex:usa .
}
=>
{
    ed:edward ex:speaks ex:english.
}
```

Listing 1.3: Example of rule in Notation3

In this way, a semantic engine could automatically infer that Edward speaks English, given that he was born in the USA. A step forward in this example would be the representation of more generic rules, which is also permitted by Notation3. For example, we can represent a rule indicating that any person who was born in the USA speaks English, as shown in Listing 1.4.

```
@prefix ex: <http://example.com/> .

{
    ?person ex:born-in ex:usa .
}
```

```
=>
{
    ?person ex:speaks ex:english.
}
```

Listing 1.4: Example of a generic rule in Notation3

The above examples have shown how to use RDF to define a representation model and to express information for our example domain. However, the potential of semantic technologies can not be fully realized if different data sources represent the same concepts employing vastly unlike predicates and URIs. Therefore there is a need to agree on a set of definitions grouping common predicates and entities. These definitions are referred to as ontologies and provide a mechanism for sharing knowledge (Gruber, 1993). Ontologies contribute helpful concepts such as classes, relationships, and properties and facilitate the exchange of semantically annotated information among heterogeneous applications. This leads to significant enhancements in interoperability and scalability. An ontology typically represents a particular domain in detail and inherits concepts from other ontologies when required to foster reusability and composability. Subsequently, systems are able to combine parts or modules from different ontologies to represent their entire knowledge base. Indeed, Linked Data principles encourage reusing existing models rather than creating new ones.

This thesis aims at semantic modeling the process of automatic regulation of affective states. Three different domains are involved in this process: affect, intelligent environments, and rule-based automation. The first comprises the emotions and mental factors that are being recognized and regulated. The second includes all the issues related to sensing and actuating in the environment. Finally, the rule-based automation domain covers the procedure of inducing the corresponding actions to adapt the environment in response to the assessed circumstances and accordingly to the defined rules.

Consequently, a review of the main existing ontologies has been conducted for each involved application domain. Fig. 1.2 shows an overview of the different domains where we have applied semantic modeling and the most prominent ontologies existing for each domain. These ontologies are described in the following lines.

Semantic modeling has emerged as a widely-used method for representing data in smart environments. One of the most outstanding ontologies within this field was proposed by the W3C Semantic Sensor Network Incubator Group: the Semantic Sensor Network (SSN) ontology (Compton et al., 2012). SSN was released in 2012 and enables the representation of notions related to ambient intelligence, such as sensor, actuator, and observation. In the

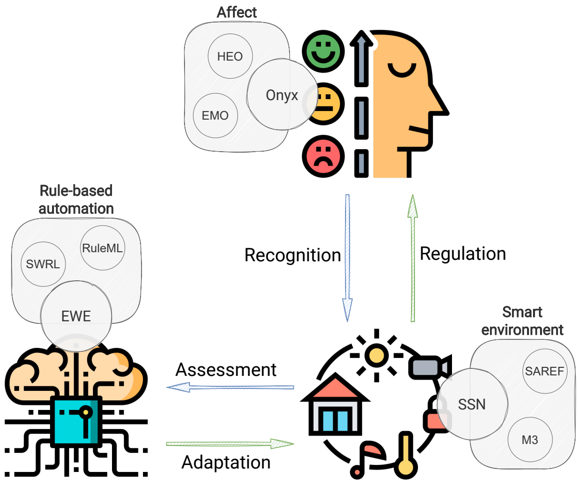


Figure 1.2: Overview of the main ontologies existing in the different areas.

ensuing years, the SSN ontology was upgraded and gave rise to the Sensor, Observation, Sample, and Actuator (SOSA) ontology (Janowicz et al., 2019).

The SSN ontology is structured into several modules Haller et al., 2019. Its core module is the so-called SOSA (Sensor, Observation, Sampling, Actuator), an independent and lightweight module that contains the primary classes and properties. SOSA provides a flexible framework that enables the representation of all the concepts concerned with sensing and actuation in addition to the relationships among them (Janowicz et al., 2019). Consequently, it can be employed as a standalone ontology. It is possible to explore SOSA from two different points of view: observation and actuation.

From the point of view of observation, we discover classes intended to be employed in sensing and measuring activities. The observation of a property is considered an event, which is concluded once the result is available. Specifically, SOSA defines the class *Sensor*, which represents a physical device that performs an *Observation* on some *ObservableProperty*. From the perspective of actuation, we find classes intended to model variations and actuations. In this regard, SOSA provides the class *Actuator*, which represents a physical device performing an *Actuation* on some *ActuatableProperty*. *ObservableProperty* and *ActuatableProperty* subclass *ssn:Property*, which represents a quality of an entity. Lastly, SOSA offers two models to link the value of a property to observation or actuation activities. The *sosa:hasSimpleResult* property can be used to represent results consisting of a simple literal, while the object property *sosa:hasResult* is needed to represent more complex results as it offers the possibility to link an individual to the observation.

The aforementioned ontologies enable the modeling of devices, platforms, and communication systems present in a smart environment but do not address the modeling of the rule-based automation process. Rule-based knowledge can be modeled and implemented with a wide variety of ontologies, such as RuleML (Boley, Paschke, and Shafiq, 2010), Semantic Web Rule Language (O'Connor et al., 2005), Rule Interchange Format (RIF) (Kifer, 2008), and Evented Web (EWE) (Coronado, Carlos A Iglesias, and Serrano, 2015). Among these, this thesis focuses on EWE, which is specifically designed for the modeling of Task Automation Services (TASs).

EWE ontology has been designed based on an analysis of some of the most prominent TAS [1] with the purpose of enabling the modeling of the rule-based automation processes in a smart environment. It defines the key aspects of task automation services in a descriptive manner and offers a standard model for their representation. The primary elements of the ontology were defined and formalized, founded on a set of key standpoints identified: configurability, privacy, integration, and communication. Furthermore, EWE is based on OWL 2 classes, and enables inferences employing a SPARQL Inferencing Notation (SPIN) Engine[2] and Notation 3 (N3) Logic[3].

The core of EWE is composed of four major classes: *Channel*, *Event*, *Action*, and *Rule*. The class *Channel* defines entities such as smart services or devices that can generate events, provide actions, or both. For instance, in the context of an intelligent environment, sensors such as a stress level detector and actuators such as a smart speaker are described as channels. The class *Event* represents a specific occurrence of a process and can be used to define the conditions under which a series of actions should be triggered. These conditions support the configuration of certain parameters to provide additional details. For example, the recognition of a high-stress level produced by the stress detector sensor could be an entity belonging to this class. The class *Action* defines an operation to be triggered under certain conditions. An action provides effects that are specific to the action itself and can be configured to react accordingly to the data collected from an event. Following the smart environment context, playing relaxing music is an example of action generated by the smart speaker channel. Finally, the class *Rule* defines an *Event-Condition-Action (ECA)* rule, such as "*If a high-stress level is detected, then play some relaxing music.*".

Finally, with regard to the representation of affects, the literature also presents interesting alternatives, among which we emphasize the Human Emotion Ontology (HEO) (Grassi, 2009), and Onyx (Sánchez-Rada and Carlos A. Iglesias, 2016). Specifically, this thesis employs Onyx due to its more generic approach. Onyx is a publicly available ontology for

[1]Task Automation Services study: `http://www.gsi.upm.es/ontologies/ewe/study/full-results.html`

[2]TopBraid (Available at `https://www.w3.org/2001/sw/wiki/TopBraid`

[3]Euler Yet another proof Engine (EYE) Available at `http://eulersharp.sourceforge.net/`

emotion representation that integrates with the Provenance Ontology [4], which provides the capability to reason about the origin of data annotations. Furthermore, it provides a meta-model for emotions, allowing the publication of new emotion models while preserving semantic validity, thereby facilitating the detachment between representation and psychological models. This has great relevance, given the lack of a standard wide-adopted model for emotions.

The Onyx ontology contains three primary classes that are worth explaining. Namely, they are: *Emotion*, *EmotionAnalysis* and *EmotionSet*. *Emotion* represents the affective state of an individual. *EmotionAnalysis* contains several details regarding the process followed to annotate an entity (e.g., a text or a video segment), such as the source or the algorithm used. Finally, the *EmotionSet* represents a group of emotions found in a given entity.

Semantic technologies are used frequently in smart environments. However, their application to emotion-aware systems aimed at enhancing and promoting mental and emotional well-being has yet to be thoroughly studied. In this line, further research is needed in the integration of affective sensors and semantic modelling into a unified smart automation platform. This idea serves as motivation for the work described in **"An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices"** (Section 3.2.6), **"An Emotion-Aware Learning Analytics System Based on Semantic Task Automation"** (Section 3.2.5) and **"Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices"** (Section 3.2.3).

1.2.4 Agent-based simulation for well-being promotion

Comprehension and understanding of a complex well-being promotion ecosystem is a challenge. The reconditeness of human behaviour coupled with the economic costs associated with the design, implementation, and validation of interventions are known obstacles to the promotion of well-being. Furthermore, the application of technology to such an ecosystem may entail additional levels of complexity (Kalton et al., 2016). Therefore, there is a need for solutions that enable the suitability and efficacy of promotion strategies before their implementation. These solutions aim at assisting the development, evaluation, and prediction of the impact of well-being promotion strategies without interfering with the real-world environment (Kalton et al., 2016). In this context, a popular approach is the utilization of agent-based simulation and modeling techniques.

Agent-based simulation is a computational method for visualizing and analyzing complex dynamic systems (Tracy, Cerdá, and Keyes, 2018). It consists of agents with a predetermined

[4]Provenance Ontology. Available at: `https://www.w3.org/TR/prov-o/`

set of attributes that make decisions and interact with each other and their surroundings. Agents can represent any entity of interest, from human individuals to organizations, services, or devices. The behaviour of these simulated entities can be adapted in response to their learning, experience, and interactions in accordance with a set of programmable rules. These rules are intended to represent important aspects of the real world, including a variety of individual or community-level traits and other social affects. The individual's personal qualities, their social and physical surroundings, and the interactions agent-agent and agent-environments are defined by these rules (Badham et al., 2018).

In this way, agent-based simulation systems can potentially yield great fidelity between the real environment and the model, supporting the extrapolation from the virtual simulated scenario to the real world. The insights and conclusions drawn from the outcomes obtained with the simulation can be used to understand the real environment. Therefore, agent-based simulation allows designers to analyze the impact and outcomes of strategies through simulations in virtual scenarios with the same characteristics as the real environment.

The benefits of agent-based simulation have made it the ideal solution for a wide range of applications and has been applied with positive outcomes in several domains, including emergency evacuation (Hawe et al., 2012), commerce (Chen et al., 2008), or traffic management (Bazzan and Klügl, 2014). These systems have also been used in the health field, mainly in the modeling of infectious disease transmission and control in populations (Shamil et al., 2021). The susceptible-infected-recovered (SIR) framework has become the most popular and widely utilized model for simulating infectious disease transmission (Siettos and Russo, 2013). This framework utilizes differential equations to govern the flow between susceptible, infected, and recovered states within a population. The SIR framework has been successfully applied to a variety of infectious diseases, including influenza and COVID-19, and has provided valuable insights into the spread and potential control measures for these diseases.

In the last decades, agent-based simulation has been increasingly applied to noncommunicable diseases and health behaviours (Tracy, Cerdá, and Keyes, 2018). This has also been reflected in an increase in the number of papers related to the use of agent simulation for the analysis of mental and emotional conditions. One of the most relevant works in this field is the one proposed by Silverman (2001). Silverman proposed a framework integrating theories and models related to human behaviour modeling literature, including physiological, cognitive, and emotive processes. This model has been particularly defined for modeling human behaviour in agents and has served as a starting point for related works based on agent simulation for the analysis of mental issues as stress (Duggirala et al., 2016; Singh et al., 2016). There are also some works related to the analysis of depression. For example, Abdelhamid et al. (2016) developed an agent-based model for analyzing the evolution of depression that included agent-to-agent interactions and within-agent interactions among depression symptoms.

Emotional agents have been considered in a wide range of works related to crowd evacuation, analyzing the evolution of emotions during emergencies, their impact on the execution of emergency procedures, and even their propagation and contagion among agents (S. Yang et al., 2020). Furthermore, emotion contagion is also a ubiquitous target for agent simulation-based analysis, as demonstrated by the large number of works conducted in this field (Haeringen, Gerritsen, and Hindriks, 2023).

Finally, regarding the efficacy of mental health interventions, some studies have been performed to analyze the influence of diverse mindfulness programs on specific aspects. For example, García-Magariño and Plaza (2017) designed a system to simulate the effect of mindfulness on specific mental and emotional conditions. However, agent-based simulation is still relatively uncommon in the areas of mental and emotional well-being promotion, and its application to psychological aspects is still challenging.

One of the most challenging aspects in developing an agent-based system is the specification of the model, specifically to decide an appropriate level of detail that finds the suitable balance between simplicity and realism (Tracy, Cerdá, and Keyes, 2018). This level of detail falls on a continuum whose position is defined by the main target of the model. In general terms, the level of detail increases according to the importance of the level of realism required by the system (Badham et al., 2018). The difficulty in addressing this challenge is exacerbated in the context of agent-based models where there is a lack of theoretical consensus or empirical data, as is often the case when modeling human behaviour (Jager, 2017).

The scarcity of empirical data also leads to challenges related to the validation of the models. There exist, however, validation techniques that not only increase confidence in model results but also help to elucidate the contribution of specific parameters to the explanatory capability of the model. An example of this technique is sensitivity analysis, which is an essential practice in ABM (Saltelli and Annoni, 2010). Sensitivity analysis consists of systematically varying the model inputs and analyzing the effect on the simulation results. In addition, it can be combined with qualitative methods or specific experiments to better assess the validity of the system (Chalabi and Lorenc, 2013).

As commented before, approaches to psychological well-being promotion based on agent simulation are scarce. A more detailed and realistic model for this domain can benefit research, and that has been the motivation for the work presented in **"An agent based simulation system for analyzing stress regulation policies at the workplace"** (Section 3.2.4).

1.3 Hypotheses

To drive the research described in this thesis, a set of hypotheses have been defined in view of the described background and context. These hypotheses aim to narrow down and precisely introduce the challenges to address in order to clearly verify and validate the contributions of this work. The defined hypotheses are:

- **Hypothesis 1 (H_1). Social, environmental, and personal-related signals can help to predict work-related mental disorders using non-obtrusive techniques.** With the first hypothesis, we assume that work-related mental disorders such as occupational stress can be predicted using non-obtrusive methods that take into account social, environmental and personal-related information. Therefore, this hypothesis guides us to the study of which techniques based on the exploitation of users' information can enhance the assessment of mental disorders, and to which extent.
- **Hypothesis 2 (H_2). The development of mental disorder models can help in understanding the effects of each factor for proposing suitable regulation techniques.** With this hypothesis, we theorize that the definition of mental disorder models can help to identify the influence of different factors on mental and emotional welfare, as well as provide further insight into their foundations and management. Besides, we raise a hypothesis that a model promoting the understanding of these factors can also help in proposing suitable regulation techniques.
- **Hypothesis 3 (H_3). Mental and emotional well-being in the workplace can benefit from the seamless integration of social and personal sensors and actuators.** With the last hypothesis, we put forth the assumption that the smooth interoperation between sensors and actuators which are aware of users' personal and contextual information can enhance mental and emotional well-being in the workplace. This hypothesis leads us to an approach for a intelligent system that takes care of the users' preferences, emotions, mental state and context to improve their welfare and comfort.

To address each of the above three hypotheses, this thesis proposes a number of approaches.

1.4 Objectives

Motivated by the above-described hypotheses and to fulfill the motivation of this thesis, the main goal of this thesis has been outlined. The primary objective is to deliver a solution that will promote mental and emotional welfare in the workplace using intelligent techniques. However, as depicted in the previous sections the challenges related to the enhancement of mental and emotional well-being in the workplace arise from a number of matters. This, the thesis global objective has been broken down into a number of more particular ones that contribute to progressively build the final solution:

- **Objective 1 (O_1). Specification of an approach for stress prediction using non-obtrusive methods.** This approach should utilize context and personal-related data to predict the stress level of an individual. The approach will achieve this by the exploitation of textual information and analysis of surroundings, in order to ensure its unobtrusiveness.
- **Objective 2 (O_2). Design of an agent-based simulation model for understanding stress-related factors and regulation policies in the workplace.** This model should help to identify and understand the different elements and factors which interfere on occupational stress, enabling the definition and analysis of regulation policies. The reliability and robustness of the model should be guaranteed by its validation with real data.
- **Objective 3 (O_3). Development of a reference architecture for an emotion-aware automation platform based on semantic technologies.** The architecture should seamlessly integrate a variety of connected devices and services able to capture users' information and adapt the environment accordingly. This architecture should incorporate a defined semantic model to guarantee the interoperability and scalability of all system components.

1.5 Document outline

The remainder of the document is organized as follows:

- **Chapter 2** presents the methodology used to conduct and evaluate the outcomes of this thesis. It describes the phases of the research, some strategies followed on research methods, and how and under which criteria the different outcomes are evaluated.

- **Chapter 3** is a compilation of the full text of the publications that form the core of this thesis. First, a list of the publications is provided. Then, each article is presented in a separate section, where the complete text is preceded by a table containing related useful information (title, authors, etc.).
- **Chapter 4** contains a general discussion about this thesis. It provides a general overview of the solutions proposed and an analysis of the results obtained. Subsequently, it presents a discussion of the main findings of the research describing their implications and significance and concludes with a description of possible lines for future research.

Finally, Appendix A contains a summary of all articles published during the course of this thesis, including those with an indirect relation to this thesis.

CHAPTER 2

Methodology

One of the fundamental principles of the scientific method is the importance of conducting rigorous and methodological research. This means carefully planning and conducting the research in a way that minimizes bias and maximizes the reliability of the results. This chapter provides a detailed and systematic description of the research approach followed in order to arrive at the research findings and conclusions presented in the subsequent chapters.

2.1 Introduction

The complex and multi-dimensional nature of mental and emotional well-being poses a series of challenges when conducting research in the field. One of these challenges is the necessity to integrate theoretical and empirical approaches. The combination of both approaches allows researchers to achieve a broader and more nuanced comprehension of the different factors contributing to well-being and how they can be adapted to improve it.

The theoretical foundations of this thesis are grounded in the area of artificial intelligence, more specifically in the fields of ambient intelligence, agent-based simulation, Linked Data, and machine learning. However, the work conducted in this thesis follows a multidisciplinary approach and is also linked to other areas, such as psychology or medicine. A range of theoretical perspectives in these areas has been exploited, including theories of emotion modeling, mental and emotional condition regulation, mental illness models, and propagation or contagion of moods. These theories provide a framework to understand the processes underlying well-being, so we have drawn upon these theoretical foundations to design and implement new technology-based approaches to mental and emotional well-being promotion.

Given the multiple areas of artificial intelligence involved in this thesis, we have leveraged a wide range of different technologies for the development of the proposed solutions. This led us to the use of empirical methods that allowed us to test the validity of our hypotheses and evaluate the proposed solutions. Nevertheless, assuring the reliability and robustness of the solutions requires the use of specific validation techniques depending on the technologies involved and the nature of the proposed approaches. In this regard, our thesis has used a set of validation techniques, including experiments in real scenarios, sensitivity analysis, statistical analysis, or machine learning performance metrics.

Concerning real-world experiments, it is worth mentioning that some of them involved real participants. In these cases, the experiment was designed according to the ethical standards of participation with humans, and the ethical approval for the experiment was obtained by the corresponding ethics committee when needed.

The following sections outline the phases of our research, along with the specific methodological considerations that have been followed for each research line in this thesis.

2.2 Phases

The organization of this thesis is built upon the foundation of three main fields of knowledge: automatic psychological awareness, agent-based simulation for analyzing well-being promotion strategies, and smart environment adaptation. These phases were further subdivided into different tasks:

1. Phase 1 (automatic recognition of psychological condition)
   (a) Study of the state of the art in automatic recognition of psychological condition
   (b) Definition of a non-obtrusive approach to recognize stress
   (c) Implementation of models for automatic stress detection
   (d) Evaluation of the proposed models
2. Phase 2 (agent-based simulation for analyzing well-being promotion strategies)
   (a) Study of the state of the art in agent-based simulation for well-being promotion
   (b) Definition of a model for occupational stress based on work and ambient conditions
   (c) Development of an agent-based simulation system to analyze mental well-being promotion strategies
   (d) Analysis of mental well-being promotion strategies with the developed system
   (e) Evaluation of the proposed system
3. Phase 3 (smart environment adaption)
   (a) Study of the state of the art in smart environment adaption for well-being promotion
   (b) Definition of a semantic vocabulary of an emotion-aware automation platform
   (c) Development of a reference architecture for a semantic emotion-aware automation platform
   (d) Evaluation of the proposed platform

2.3 Evaluation

The anticipated results of this thesis can be grouped into one of the following categories: machine learning models, agent-based models, semantic vocabularies, reference implementations, and software architectures. The validity of all these results must be assessed before they can be considered a scientific contribution.

Machine learning models are, a priori, the easiest to evaluate properly. The extensive literature in the field contains a series of well-established methodologies and guidelines for the implementation and evaluation of machine learning-driven research. In this way, each machine learning model is evaluated on multiple datasets, and different performance metrics like accuracy, precision, and F-Score are computed. These metrics provide some insights regarding the performance of the proposed model.

Also, to further study the impact and performance of the models, and to determine whether there is a difference between the classification methods, performing a statistical analysis is a good practice (Alpaydin, 2020). Such an analysis should be done when possible, as it helps to avoid subjective observations of the results. Accordingly, the machine learning models of this thesis have been evaluated through a number of statistical tools, providing exhaustive comparisons among them and also with existing baselines in the field. To this end, available public datasets have been used,

The validation of agent-based simulation systems refers to the process of assessing the accuracy, reliability, and robustness of these systems. This is a crucial step in the development and use of these systems, as it helps to ensure the significance and utility of the insights drawn from them. There are several popular approaches to validating agent-based simulation systems. The validation of these systems is also well-covered by the literature, and popular techniques include comparing the output of the simulation to experimental data and using sensitivity analysis to test the robustness of the simulation (Heath, Hill, and Ciarallo, 2009).

Sensitivity analysis is a popular and straightforward approach that consists in systematically varying the input parameters of the simulation and assessing the impact on the output. This can help identify which parameters are most important and to which extent the simulation is sensitive to small changes in these parameters. However, the comparison with experimental data presents more challenges due to the scarcity of available data. We have faced this problem during this thesis, given the scarce data available for the analysis of psychological well-being in the workplace. Thus, we have performed some experiments in a real scenario to collect data from the real-world system being studied and compare it to the output of the simulation. This evaluation provides the researchers with confidence in the accuracy and reliability of the results and analysis performed with the system.

Real-world experiments were also used to evaluate the proposed environment adaptation methods. These experiments were intended to assess the effectiveness of the proposed methods in the regulation of the psychological state and to measure the satisfaction and performance of users. A within-subject design was used, that is, all participants were exposed to every condition.

As for reference implementations and additional types of software results, we have followed standard software engineering practices. We have ensured that all software modules are documented, tested, and conform to specifications. Besides, the developed software has been published under Open Source licenses on public code repositories (GitHub and GitLab). In this manner, a community-driven approach is enabled where external researchers are fostered to find flaws or contribute with their adaptations or enhancements to the code. In addition, in order to simplify installation and ensure compatibility with as many platforms as possible, we have also provided containerized images of our software. Specifically, we used Docker, one of the most popular open-source containerization software.

Finally, we defined a set of competency questions to evaluate semantic vocabularies to check their selection capabilities. Competency questions represent the ontology requirements and consist of a set of queries related to the data that the ontology must be able to answer correctly (Wisniewski et al., 2019).

CHAPTER 3

Publications

This chapter presents the core papers published as part of this Ph. D. thesis. These papers represent the primary research contributions of this work and demonstrate its novelty and significance.

3.1 Introduction

The table below provides an overview of the publications that are included as part of this thesis. As this document is structured as a compilation of publications, only those that meet the established criteria are included in this section. A comprehensive list of the author's publications, including those not directly related to the research objectives of this thesis, can be found in A.

Table 3.1: List of publications directly related to this thesis

Page	Title	Year	Venue	Ranking
35	A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations	2022	Information Processing & Management	JCR 2021 Q1 (7.466)
55	Prediction of stress levels in the workplace using surrounding stress	2022	Information Processing & Management	JCR 2021 Q1 (7.466)
69	Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices	2022	Electronics	JCR 2021 Q3 (2.690)
89	An agent based simulation system for analyzing stress regulation policies at the workplace	2021	Journal of Computational Science	JCR 2021 Q1 (3.817)
103	An Emotion-Aware Learning Analytics System Based on Semantic Task Automation	2020	Electronics	JCR 2021 Q3 (2.690)
128	An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices	2018	Sensors	JCR 2021 Q1 (3.847)

3.2 List of central publications

3.2.1 A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations

Title	A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations
Authors	Muñoz, Sergio and Iglesias, Carlos A.
Journal	Information Processing & Management
Impact factor	JCR 2021 7.466
ISSN	03064573
Volume	59
Year	2022
Keywords	Stress detection, stress framework, distributional representations, text classification, affective computing
Pages	103011
Online	`https://linkinghub.elsevier.com/retrieve/pii/S0306457322001212`
Abstract	Nowadays, stress has become a growing problem for society due to its high impact on individuals but also on health care systems and companies. In order to overcome this problem, early detection of stress is a key factor. Previous studies have shown the effectiveness of text analysis in the detection of sentiment, emotion, and mental illness. However, existing solutions for stress detection from text are focused on a specific corpus. There is still a lack of well validated methods that provide good results in different datasets. We aim to advance state of the art by proposing a method to detect stress in textual data and evaluating it using multiple public English datasets. The proposed approach combines lexicon-based features with distributional representations to enhance classification performance. To help organize features for stress detection in text, we propose a lexicon-based feature framework that exploits affective, syntactic, social, and topic-related features. Also, three different word embedding techniques are studied for exploiting distributional representation. Our approach has been implemented with three machine learning models that have been evaluated in terms of performance through several experiments. This evaluation has been conducted using three public English datasets and provides a baseline for other researchers. The obtained results identify the combination of FastText embeddings with a selection of lexicon-based features as the best-performing model, achieving F-scores above 80%.

Information Processing and Management 59 (2022) 103011

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations

Sergio Muñoz *, Carlos A. Iglesias

Intelligent Systems Group, ETSI Telecomunicación, Universidad Politécnica de Madrid, Avenida Complutense 30, Madrid, 28040, Spain

ARTICLE INFO

Keywords:
Stress detection
Stress framework
Distributional representations
Text classification
Affective computing

ABSTRACT

Nowadays, stress has become a growing problem for society due to its high impact on individuals but also on health care systems and companies. In order to overcome this problem, early detection of stress is a key factor. Previous studies have shown the effectiveness of text analysis in the detection of sentiment, emotion, and mental illness. However, existing solutions for stress detection from text are focused on a specific corpus. There is still a lack of well-validated methods that provide good results in different datasets. We aim to advance state of the art by proposing a method to detect stress in textual data and evaluating it using multiple public English datasets. The proposed approach combines lexicon-based features with distributional representations to enhance classification performance. To help organize features for stress detection in text, we propose a lexicon-based feature framework that exploits affective, syntactic, social, and topic-related features. Also, three different word embedding techniques are studied for exploiting distributional representation. Our approach has been implemented with three machine learning models that have been evaluated in terms of performance through several experiments. This evaluation has been conducted using three public English datasets and provides a baseline for other researchers. The obtained results identify the combination of FastText embeddings with a selection of lexicon-based features as the best-performing model, achieving F-scores above 80%.

1. Introduction

Mental illness – and more specifically, stress – is a growing problem for modern society (Can, Chalabianloo, Ekiz and Ersoy, 2019). Stress has become part of our daily lives, impacting affected individuals and their families, health care systems, private and social insurers, employers, work colleagues, and the society at large (Kassymova et al., 2019).

The economic and health costs that stress entails worldwide are much more significant than expected. In Europe, a public survey identified stress as the second most prevalent work-related health problem (for Safety & at Work, 2013). According to this survey, the stress in the workplace is a frequent issue for 51% of European employees, and 40% of them feel that stress is not handled effectively (Parent-Thirion et al., 2012). In terms of economic impact, the annual cost to the European companies of work-related stress is estimated at 25 billion euros (Hassard et al., 2014). The problem is not particular to Europe since existing reports from countries worldwide show similar results. For example, in the United States, 40% of workers consider their job notably or vastly stressful, whereas 29% of them feel slightly or highly stressed at the workplace (Saute et al., 1999). Also, it is estimated that the

* Corresponding author.
E-mail address: sergio.munoz@upm.es (S. Muñoz).

https://doi.org/10.1016/j.ipm.2022.103011
Received 29 March 2022; Received in revised form 22 June 2022; Accepted 26 June 2022
Available online 1 July 2022

United States spends 300 billion USD per year on stress-related diseases (Can, Arnrich and Ersoy, 2019). Available reports from other countries such as China (Xiong, Skitmore, & Xia, 2015) or Australia (Society, 2015) show the global importance of the stress problem.

In the light of its high impact, preventing and regulating stress has become a critical health issue for populations (Greene, Thapliyal, & Caban-Holt, 2016). Early detection and monitoring of stress problems can significantly improve the efficiency of interventions, decreasing their costs and preventing stress from being chronic (Can, Chalabianloo et al., 2019). In this context, methods are needed to detect stress in time. Stress detection has been traditionally assessed using self-reports in response to standardized questionnaires (Andreou et al., 2011), such as Perceived Stress Scale (Chan & La Greca, 2020) or Depression Anxiety and Stress Scale (Osman et al., 2012). Although the validity of these methods has been proved, they suffer from two major drawbacks: their frequent delay in diagnosis, which makes them unsuitable for early detection, and their subjectivity and dependence on subjects' recall and situation awareness (Alberdi, Aztiria, Basarab, & Cook, 2018).

For this reason, there is still a need for objective measures to detect stress based on physical and physiological information (Greene et al., 2016). The great advances in affective computing open a range of possibilities for addressing these issues. In recent years, a great deal of effort has been devoted to research into methods and systems that use smart devices and affective computing algorithms for automatic stress detection. Some of the most popular approaches consist of the analysis of physiological signals (de Santos Sierra, Ávila, Casanova, & del Pozo, 2011), facial expression (Giannakakis et al., 2017), speech (Hansen & Patil, 2007), phone usage (Ferdous, Osmani, & Mayora, 2015), or keystroke-dynamics (Vizer, Zhou, & Sears, 2009). These methods avoid the subjective response bias from the individual introduced in self-report questionnaires, and their reliability has been proved. However, their intrusiveness and high implementation costs are often a major constraint (Novais & Carneiro, 2016).

This challenge has boosted the research on more economical and unobtrusive methods, such as the use of social media data (Lin, Jia, Nie, Shen, & Chua, 2016). Social media services where users communicate and share their thoughts or experiences have gained popularity and provide a vast amount of information related to people's emotions, daily moods, and worries (Pang et al., 2019). Textual data coming from these platforms have been successfully exploited by the research community in a wide variety of text classification applications such as detection of sentiment (Yue, Chen, Li, Zuo, & Yin, 2019), radicalization (Araque & Iglesias, 2020) and also mental illness (Banerjee & Shaikh, 2021; Chancellor & De Choudhury, 2020). The most popular approach in the literature concerning text classification is the use of machine learning techniques that represent texts as vectors in a feature space and classify them into categories (Bandhakavi, Wiratunga, Padmanabhan, & Massie, 2017).

Some studies have exploited these techniques for the task of stress detection from text (Cao et al., 2021; Lin et al., 2016; Winata, Kampman, & Fung, 2018). Existing works achieve great results, but they focus on specific data sources. Thus, there is still a lack of well-validated methods that show good results in different datasets. We aim to contribute to state of the art in stress classification from text by proposing a machine learning method to detect stress in corpora from different sources: personal interviews, Reddit social network, and Twitter social network. Also, our work aims at identifying which features and techniques perform better. Therefore, we focus on the following research questions (RQs) related to the task of stress classification from text:

- **RQ1**: Which kinds of lexicon-based features are more relevant and yield better results?
- **RQ2**: How do different machine learning models compare in terms of performance?
- **RQ3**: Can a machine learning approach achieve good results when evaluated on several corpora?

Motivated by these RQs, this paper proposes three different machine learning models for detecting stress in texts: the first consists of a lexicon-based feature extraction method, the second uses word embedding techniques for exploiting distributional representations, and the third combines distributional representations with lexicon-based features. For this purpose, our work proposes a lexicon-based feature framework that exploits affective, syntactic, social, and topic-related features. This framework aims at helping in the organization and characterization of features for stress detection in text. Besides, three different word embedding techniques are considered to analyze their suitability for the stress detection task. To evaluate the effectiveness of the proposed models, several experiments have been conducted using public English datasets from the three different sources mentioned before. In this way, a baseline for other researchers is provided. Furthermore, the relevance of the experiments is confirmed with a statistical study that allows us to further analyze the proposed models' performance. Finally, our best-performing model is compared with state-of-the-art stress detection methods in text. The obtained results identify the combination of FastText embeddings with a selection of lexicon-based features as the best-performing method. This method outperforms the existing works for every dataset, achieving F-scores that surpass 80%.

The rest of the paper is organized as follows. An overview of stress theory and text classification methods is given in Section 2. Following, Section 3 presents the used datasets along with a preliminary analysis performed on them. The feature framework for stress detection from text is presented in Section 4. In Section 5, the proposed stress detection models are described. Later, in Section 6, the experimental setup aimed at evaluating the proposed models is presented, along with the obtained results. A discussion regarding the main findings of the article is given in Section 7. Finally, the paper concludes in Section 8, where conclusions drawn from the work are depicted, and an outline of possible lines of future work is presented.

2. Related work

This section presents the background and related work of the concepts and technologies involved in the paper. First, Section 2.1 gives an overview of stress theory and detection methods. Then, Section 2.2 introduces popular approaches for text classification.

2.1. Psychological stress

Due to the subjectivity of stress and the different contexts where this concept is used, a universally recognized definition for stress is still lacking (The American Institute of Stress, 2013). One of the earlier and more generic stress definitions was proposed by Hans Selye, who defined stress as the non-specific response of the body to any demand (Selye, 1956). Over the last years, a vast number of extended and more specific stress definitions have been proposed (Burman & Goswami, 2018). For example, Kim and Diamond (Kim & Diamond, 2002) propose a stress definition that considers that stress requires three main components: heightened excitability or arousal, an experience perceived as aversive, and lack of control. Another perspective is proposed by Cox and Griffiths (1995a), Cox and Griffiths (1995b) who state that the definition of stress can be addressed from three different approaches: psychological, engineering, and physiological. From a psychological point of view, stress can be seen as a dynamic process that stems from the interaction between an individual and the environment. The engineering approach defines stress as a stimulus of the environment in the form of a demand level. Finally, regarding the physiological one, stress can be defined as the changes that occur in a human under pressure. According to this, stress can be non-formally defined as the reaction of the human body to any challenging or hazardous situation (Can, Arnrich et al., 2019).

The stress subjectivity and complexity have also resulted in numerous stress theories that form the basis for understanding stress (Dewe, O'Driscoll, & Cooper, 2012). One of the most relevant is the Person–Environment (P-E) fit theory (French, Caplan, & Van Harrison, 1982), founded on the studies by Lewin (1936) and Murray (1938). This theory has been the source for other approaches to stress and well-being and argues that stress arises from the lack of fit or congruence between the person and the environment. According to this theory, stress can be seen as a lack of match between a person's abilities and demands. In 1982, the Transactional Model of Stress (Holroyd & Lazarus, 1982) was proposed by Lazarus and Holroyd. This theory considers stress as a relationship between the person and the environment. According to the authors, the person appraises the environment as taxing, hence threatening well-being (Glanz, Rimer, & Viswanath, 2008).

These models have been the basis for understanding stress and have helped the development of prevention, detection, and regulation methods. Research in the field has shown excellent results in detecting stress using smart sensors and devices (Can, Arnrich et al., 2019). These devices measure physiological (e.g., brain or heart activity, skin response, and breath response) and physical features (e.g., facial expression, eye tracking, behavior, and gesturing). Greene et al. (2016) and Panicker and Gayathri (2019) performed exhaustive reviews about the usage of these techniques to detect stress. The most successful results have been yielded using Electro-Dermal Activity (EDA) (Affanni, Bernardini, Piras, Rinaldo, & Zontone, 2018), Electroencephalogram (EEG) (Vanitha & Krishnan, 2017), Blood Pressure (BP), Respiration, Blood Volume Pulse (BVP) (Widanti, Sumanto, Rosa, & Miftahudin, 2015), facial expressions (Gao, Yüce, & Thiran, 2014), speech (Tomba, Dumoulin, Mugellini, Abou Khaled, & Hawila, 2018), or mobile phone usage (Maxhuni et al., 2021). The accuracy obtained with these methods ranges between 70% and 90%, proving the reliability of these systems for detecting stress. However, they often have a major drawback when implementing these solutions in real scenarios: their intrusiveness and high costs (Novais & Carneiro, 2016). Besides, errors from incorrect placement, movements, or detached equipment are very common in daily life and lead to corrupted data (Can, Arnrich et al., 2019).

This fact, along with the significant advances in Natural Language Processing techniques, has given momentum to the approach of stress detection from text. Moreover, the vast amount of textual data contained nowadays in social networks (Statista, 2018) makes these techniques a promising approach.

2.2. Text classification methods applied to stress detection

Stress detection from text can be considered a text classification problem that aims to distinguish texts depending on whether they express stress. This can be shaped by exploiting certain syntactic and linguistic features using machine learning techniques or through lexicons. The lexicon-based approach compares words in the text with a dictionary to calculate the presence and frequency of specific bearer terms. Lexicons can provide an overall indication of specific features (e.g., sentiment, emotion, cognition, or topic) depending on the nature of the words they contain (Khoo & Johnkhan, 2018). This approach tends to be computationally fast and has yielded good performance on a wide variety of text classification applications such as detection of mental illness (Giuntini et al., 2020) or extreme opinions (Almatarneh & Gamallo, 2018) from social media. For example, Mike Thelwall used lexicons to develop TensiStrength (Thelwall, 2017), a system able to detect stress and relaxation in tweets. TensiStrength detects the expressions of stress and relaxation through a list of stress-related terms and a set of rules. This method yields reasonable accuracy levels, but the performance is lower than that obtained using machine learning methods.

In order to enhance the performance, a popular approach is to include lexicon-based features in machine learning methods. Lin et al. (2014) exploited this approach by proposing a deep neural model for detecting stress from Chinese micro-blogging posts like Sina Weibo or Tencent Weibo. Intending to overcome the problem of scarce stress-annotated data existing, they collected posts from these platforms. They used sentence patterns such as "The day was stressful" to obtain the ground truth labeled data. The social data extracted from the social network was combined with textual and image data obtained from each post and then used to train a deep neural network model.

Lexicon-based features can help to provide a general indication of a text in terms of sentiment, cognition, or topic. However, they fail to apprehend more refined attributes and contextual cues intrinsic to the human language (Giatsoglou et al., 2017). Word embedding-based approaches address this challenge, enabling the encoding of semantic and syntactic features present in words and their representation in a vector space as relation offsets. These vectors are named pre-trained word vectors and can be used for textual representation in text categorization tasks (Wang, Zhou and Jiang, 2020). Popular word embedding techniques are

Word2Vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), Global Vectors for Word Representation (GloVe) (Pennington, Socher, & Manning, 2014) and FastText (Joulin et al., 2016). Winata et al. (2018) exploited Word2Vec word embeddings for detecting stress in interview questions. They used data from social networks (Twitter) to extend the Natural Stress Emotion corpus (Zuo, Lin, & Fung, 2012). This corpus contains transcriptions of 12 interview questions designed to be progressively stress-provoking answered by 25 students. Then, they used these data for feeding a neural model based on bidirectional Long Short-Term Memory (Bi-LSTM) model with an attention mechanism, obtaining great results for stress classification.

Although word embedding-based approaches effectively capture syntactic and semantic features in texts, they do not exploit the individual affective or social value of the words. Previous studies have demonstrated that combining word embeddings with lexicon-based features can enhance the quality of text representations, yielding higher prediction performance (Fu, Yang, Li, Fang, & Wang, 2018). E. Turcan and K. McKeown exploited this combination to detect stress from social media data (Turcan & McKeown, 2019). They proposed a new text corpus of long social media data coming from Reddit to identify more implicit indicators than those available in micro-blogging posts. The typically longer length of Reddit posts compared to those from micro-blogging platforms allows researchers to deeply analyze the causes and indicators of stress in text. They evaluated different supervised learning methods for stress detection in the corpus, achieving the best results by combining lexicon-based features and Word2Vec word embeddings. This method establishes the current benchmark on the problem of binary text classification of stress at 79.80% F-score.

All these works have shown excellent results for classifying stress. However, they are evaluated only on data from a specific dataset (coming from micro-blogging platforms, interviews, or Reddit). We aim to provide a solution that is validated across multiple public English datasets. In this way, we propose a cross-dataset model for detecting psychological stress from text. Our approach combines word embeddings with different kinds of lexical-based features: affective, topic, social, and syntactic related. Different lexicons are evaluated, and a feature framework is proposed to identify which kinds of features drive better performance. Moreover, several word embedding techniques are considered to analyze their impact on classification performance. To the best of our knowledge, this is the first paper on stress detection from text that proposes several machine learning models combining surface and deep features and evaluates them on different stress-based text corpora (Skaik & Inkpen, 2020; Su, Xu, Pathak, & Wang, 2020; Thieme, Belgrave, & Doherty, 2020). The following sections will deeply describe our approach and its evaluation through a set of experiments. The performance reached by the proposed method is comparable to the performance of stress detection methods that use physiological or physical data (Panicker & Gayathri, 2019), but the entailed costs and complexity are considerably lower. The applications of stress textual classifiers are manifold, from detecting stress at work to business applications such as customer management or marketing.

3. Materials

To perform the evaluation, we have used three English language datasets from three different sources: Dreaddit (Reddit) (Turcan & McKeown, 2019), Natural Stress Emotion (personal interviews) (Zuo et al., 2012), and TensiStrength (Twitter) (Thelwall, 2017). These datasets and some statistics drawn from them are described in the following lines. It should be noted that even if the datasets have a different nature, for commodity, we will refer to the instances of each dataset as "posts" in the rest of the article.

- The *Dreaddit* dataset was collected by E. Turcan and K. McKeown. It contains 3549 Reddit posts annotated using Amazon Mechanical Turk, resulting in 47.75% of non-stress and 52.75% of stress posts. The average post length in the dataset is 88 words. Besides, they proposed a subset of this dataset containing only those posts that obtained confidence greater than 0.8 in the annotation (80% of agreement between annotators). The authors demonstrated that this high-agreement subset led to higher reliability in stress detection, so we have used this subset in our study.
- The *Natural Stress Emotion (NSE)* dataset, collected by Zuo et al. (2012) and extended by Winata et al. (2018), consists of a set of 38 interviews where students answered questions designed to be progressively stress-provoking. After the extension, the data results in a set of 2243 instances, where 63.70% are labeled as non-stress and 36.30% as stress. This labeling was carried out by three judges, taking as ground truth label the majority vote between them.
- The *TensiStrength* dataset, collected by M. Thelwall, contains 6142 tweets labeled manually. First, the tweet collection was carried out using keywords from various sources, and then the collected tweets were labeled using a five-point scale system. A text is annotated as "−1" if its content is not related to stress and "−5" if it describes situations likely to cause high levels of stress. Thus, a number between −2 and −4 indicates that the text somehow describes stress-related situations or matters. During pre-processing, we have transformed the 1–5 annotation into a binary system to use with our binary classifier. To convert these levels into binary classes, we transformed only texts annotated as "−1" (i.e., those that did not contain any reference to stress or stressful situations) into "0". Thus, all tweets describing stressful situations to a greater or lesser extent and therefore originally annotated by a number between "−2" and "−5" have been annotated as "1" (stress). This transformation results in 41.30% of non-stress posts and 58.70% of stress posts.

The same pre-processing has been applied to all the data: normalization of Uniform Resource Locators (URLs), capital letters, numbering, and contractions (e.g., It's, we'll), resulting in posts consisting of lower case tokens with the punctuation removed. Table 1 presents the used datasets along with a summary of some statistics.

Furthermore, a word frequency analysis has been carried out in order to identify stress patterns. This analysis is shown in Fig. 1, a word frequency scatter plot which depicts a visualization of the most used words in the data accordingly to the stress and non-stress categories. The figure has been generated using Scattertext library (Kessler, 2017), a tool for finding specific terms in corpora and visualizing them.

Table 1
Statistics of the used datasets.

	Dreaddit	TensiStrength	NSE
No. of posts	2294	6142	2243
No. posts w/Stress	1246	3605	813
Avg. no. of words	88.32	15.56	16.65
Avg. no. of chars	243.48	47.01	45.62

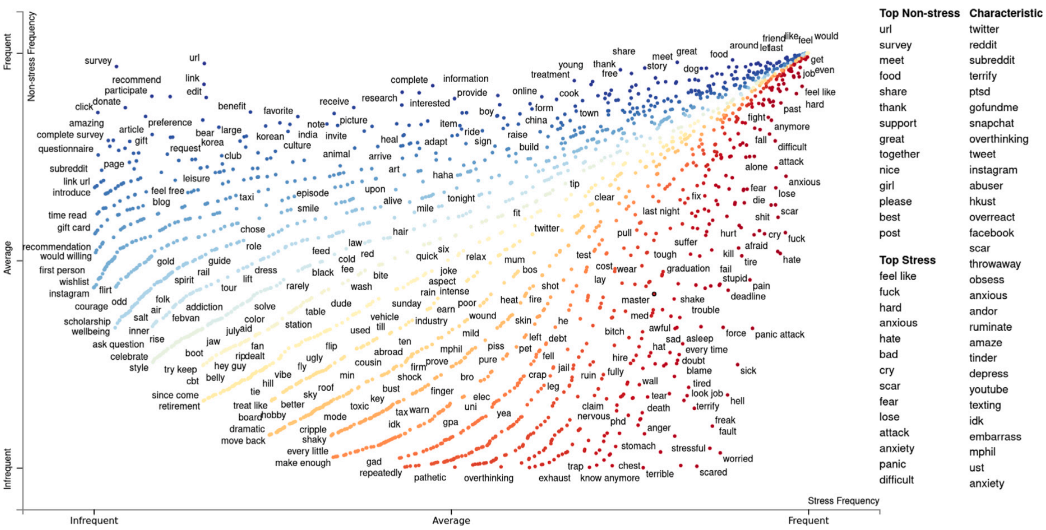


Fig. 1. Normalized word frequency for stress and non-stress categories for the data. On the right, a list containing most frequent words for non-stress (Top Non-stress), stress (Top Stress), and both (Characteristic) is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The different colors indicate the frequency of each class: blue for non-stress and red for stress. The frequency of each different word for each category is computed. Thereby, the frequency of the words in the stress category is represented on the x-axis, and the frequency of the words in the non-stress category is pictured on the y-axis. In this manner, a word that frequently appears in non-stress texts will be placed in the top area, whereas a word that frequently appears in stress-annotated texts will be placed in the right area. Consequently, the most frequent words appear in three particularly interesting areas of the figure according to the category: bottom right (common in stress texts), top left (common in non-stress texts), and top right (common in both non-stress and stress texts). Those areas present the most characteristic words for the non-stress, stress, and both categories, offering a view of which words are commonly used in each category. For instance, common stress words are "hate", "fear", and "anxious"; whereas non-stress texts frequently contain words such as "thank", "food", and "support".

Besides, an analysis has been carried out to identify topics. This allows us to visualize the most common topics present in our data depending on the stress level. We have performed topic identification using Scattertext (Kessler, 2017) with Empath (Fast, Chen, & Bernstein, 2016). This tool enables the on-demand generation and validation of new lexical categories and the analysis of text across 200 categories generated from common topics.

The results of this analysis can be seen in Fig. 2. The figure consists of a frequency scatter plot of the most frequent topics accordingly to the stress and non-stress categories. The color indicates the frequency of the topic with regard to each class. Those topics associated with non-stress are blue, and those more associated with stress are red. The position along the axis indicates the frequency in the classes. In the far upper right-hand corner, we can see topics highly associated with both classes, while in the bottom left corner, we see topics with low frequency in our data. The most frequent topics for non-stress and stress are listed on the right. For example, common stress topics are "timidity", "weakness" and "anger"; and non-stress common topics are "internet", "shopping" and "tourism". The analysis points out that most topics do not present substantial frequency differences between stress and non-stress posts. The frequency of topics such as "religion", "wealthy", and "home" is similar in both cases. However, there are a few key topics where this frequency difference is significant. For example, "swearing_terms", "horror", and "aggression" are topics very frequent in stress posts but uncommon in non-stress posts. This fact indicates that some topic-related features may provide relevant information for the task of stress classification.

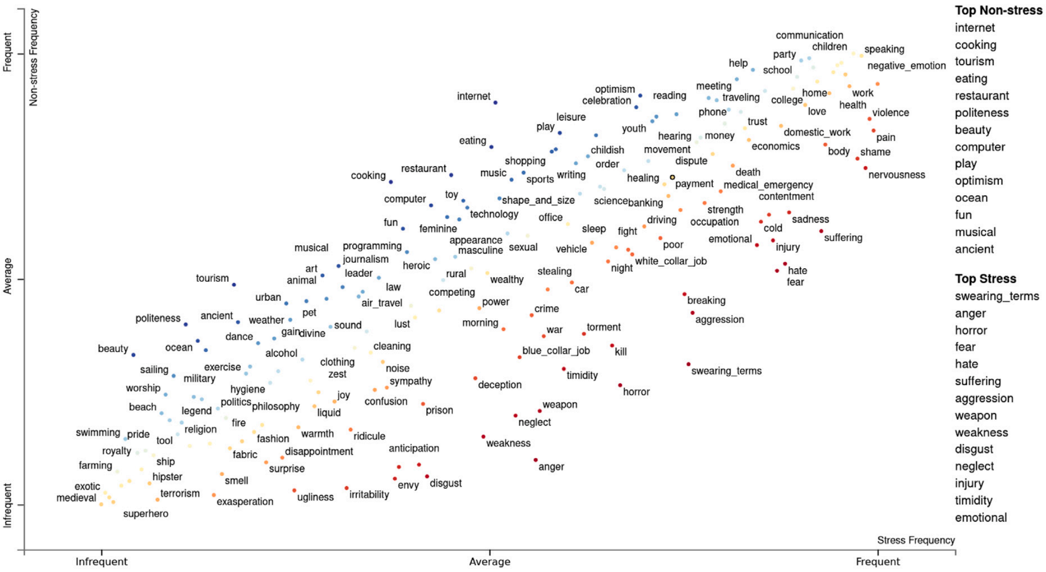


Fig. 2. Comparison of Empath topics for both stress and non-stress categories for the data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. A feature framework for stress detection from text

In this section we introduce a feature framework aiming at helping in the organization and characterization of features for stress detection in text. These features are based on existing lexicons and sentiment analysis methods that enable the extraction of information from textual data. Different lexicons have been used to identify which of them provide features more relevant for the task of stress detection: Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Francis, & Booth, 2001), General Inquirer (GI) (Stone, Dunphy, & Smith, 1966), Lasswell (Lasswell & Namenwirth, 1969), Geneva Affect Label Coder (GALC) (Martin & Pu, 2014), Affective Norms for English Words (ANEW) (Bradley & Lang, 1999), EmoLex (Mohammad & Turney, 2010), SenticNet (Cambria, Liu, Decherchi, Xing, & Kwok, 2022), Valence Aware Dictionary for Sentiment Reasoning (VADER) (Gilbert & Hutto, 2014), Hu–Liu polarity (Liu, Zeng, Li, & Hu, 2004), and Empath (Fast et al., 2016):

- *GI* (Stone et al., 1966) includes 119 features regarding institutions, roles, semantic, lexical, and syntactic dimensions, pleasure, places, communication, or social categories.
- *Lasswell* (Lasswell & Namenwirth, 1969) includes 69 features related to affection, wealth, well-being, respect, or power.
- *GALC* (Martin & Pu, 2014) consists of word lists concerning to 36 specific emotions (such as anger, guilt, joy, or hope) and two general emotional states.
- *ANEW* (Bradley & Lang, 1999) includes a total of 6 affective norms for valence, arousal, dominance, and pleasure.
- *EmoLex* (Mohammad & Turney, 2010) contains 10 lists of words and bigrams evoking particular emotions (such as joy, sadness, anger, fear, or disgust).
- *SenticNet* (Cambria et al., 2022) is a database expansion of WordNet that contains norms for around 13 000 words related to sensitivity, aptitude, attention, and pleasantness.
- *VADER* (Gilbert & Hutto, 2014) consists of a rule-based sentiment analysis system which was particularly developed for shorter texts, making it very useful in social media contexts.
- *Hu–Liu* (Liu et al., 2004) includes two large polarity lists for the purposes of sentiment analysis.

All these lexicons have been used utilizing SEANCE (Crossley, Kyle, & McNamara, 2017), an automatic tool that enables the analysis of sentiment, social order, and social cognition in text. Besides, some stylistic features have been considered, such as the Automated Readability Index (ARI) and the Flesch-Kincaid Grade Level (FKG). Both scales are intended to measure the understandability and readability of a text. Finally, also LIWC (Pennebaker et al., 2001) and Empath (Fast et al., 2016) have been used. LIWC is a lexicon-based tool that provides scores for psychologically relevant categories such as joy, sadness, or certain cognitive processes. In contrast, Empath contains 200 pre-validated categories generated from common topics. Furthermore, we

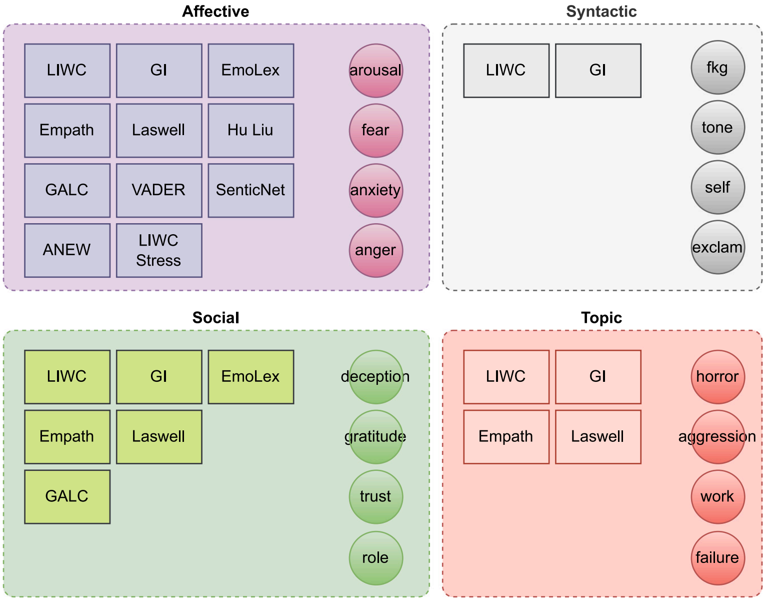


Fig. 3. Overview of the proposed feature framework. Blocks represent the lexicons which contribute to each set, whereas circles represent example features of each set.

have also included the LIWC stress dictionary (Wang, Hernandez, Newman, He, & Bian, 2016), a dictionary specifically intended to measure psychological stress.

This results in a total of more than 500 features with different types and natures. To analyze the influence of the features according to their nature, we have split them into four sets: affective, social, syntactic, and topic-related features. Affective features are those related to sentiment (positive, negative, or neutral), emotion (e.g., fear, sadness, joy), and mood (e.g., anxiety, attention, nervousness). All used lexicons contain categories of this nature, resulting in 173 features. Some examples of features belonging to this set are *fear*, *anxiety*, or *arousal*. Features related to social relations belong to the social set, which comprises categories such as *trust*, *deception* or *gratitude*. This set comprises 65 features extracted from the lexicons Empath, LIWC, GALC, GI, Laswell, and EmoLex. Only LIWC and GI contribute to the syntactic set, which also contains the previously mentioned stylistic features, ARI and FKG. Seventy-five features related to how the text has been written compound this set. Some examples are *tone*, *self* and *fkg*. Finally, the topic set comprises 272 categories to identify the different topics present in the text. Empath, LIWC, Laswell and GI dictionaries contain features of this nature, such as *horror*, *aggression* or *work*. Fig. 3 shows these feature sets, indicating which lexicon contributes to each set, along with some examples.

Finally, an analysis has been conducted to analyze how well each feature splits the posts from each corpus. This analysis allows us to gain insights into the differences between the different sources of the data and how each feature set performs on each source. The information gain criterion of each feature in each class has been computed for this purpose. This metric measures the entropy's reduction within each class once the best split induced by the feature has been conducted. Given a feature F and a class C, the information gain can be calculated as:

$$I_{gain}(F, C) = H(C) - H(C \mid F) \tag{1}$$

where $H(C)$ represents the entropy of the class and $H(C \mid F)$ is the conditional entropy of the class given the feature F. The minimum value of the information gain is achieved when $H(C \mid F) = 1$, that is, the feature F and the class C are unrelated. In contrast, a feature F that only appears in a specific class C would yield the maximum information gain value.

Fig. 4 shows the distribution of the achieved information gain values for each feature set. We can see that a significant part of the information gain values is close to 0. This indicates that a significant number of features do not provide enough information gain. However, we can also appreciate a long tail in the distribution for all feature sets. This elongated distribution, which is especially long for affective features, suggests that there are features that could potentially provide good results when being exploited by the classification model. For affective features, the elongated tail also has wider portions, indicating the existence of multiple features with high information gain. On the other hand, the tail is narrower for the social features set, indicating fewer social features with high information gain. We can observe that the information gain distribution shows a shorter tail for the syntactic features but presents wider sections for high information values. If we analyze the information gain ranges, we can see that they are greater for

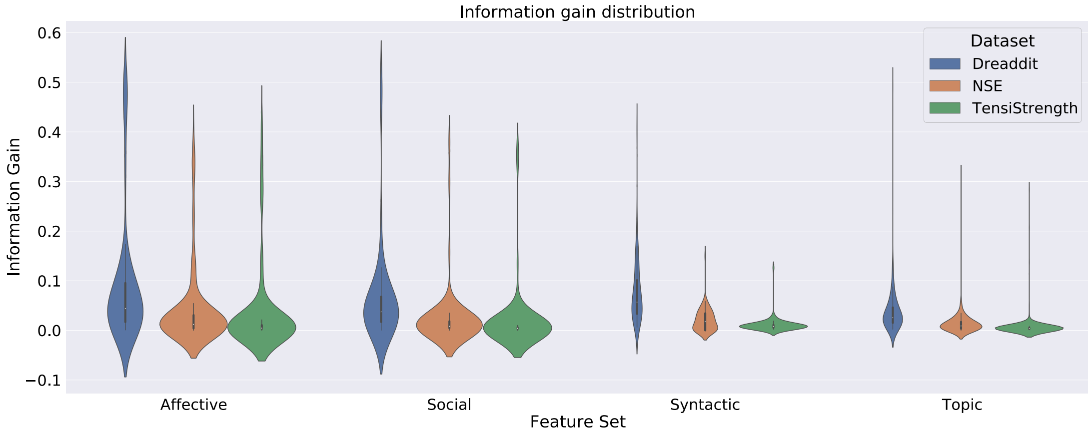


Fig. 4. Information gain distribution for features belonging to each set for the three datasets.

the affective set. This indicates that this set may contain a more significant number of relevant features for the stress classification task.

Moreover, when comparing among datasets, we can observe that the information gain achieved by all feature sets decreases with the length of the posts. Even so, analyzing the information gain ranges for NSE and TensiStrength, we can appreciate that affective features still obtain significant information values for corpora with a shorter length. Furthermore, we can observe how the information gain distribution form of affective features remains very similar for the three datasets, and the information gain range remains higher in all corpora.

By analyzing the scope, it can be noticed that the information provided by the lexicon-based features may be beneficial for the stress classification task. Specifically, the analysis points out the relevance of affective features (**RQ1**), given the promising information gain obtained by features from this set.

5. Stress classification models

Once we have analyzed the main features of the lexicons for stress characterization, we aim to use them to develop a stress classifier. Besides, we aim to explore the combination of lexicon-based features with distributional representations. With this purpose, three different models are proposed: a lexicon-based features model (M_{LF}), a distributional representation model (M_{DR}), and an ensemble model (Araque, Corcuera-Platas, Sánchez-Rada, & Iglesias, 2017) combining lexicon-based features with distributional representations (M_E). These models are described in the following subsections.

5.1. Lexicon-based features model (M_{LF})

The first model, M_{LF}, which is shown in Fig. 5, makes use of affective, social, syntactic, and topic features extracted from the text. These features have been extracted using the framework described in Section 4.

This model aims at investigating whether lexicon-based features are relevant for stress detection and to which extent. We propose a lexicon-based representation to encode the text into a fixed-length vector. Consider a set of c lexicons $L = \{l_1, \dots, l_i, \dots, l_c\}$. Each lexicon l_i is composed by a vocabulary of n words $W(l_i) = \{w_1, \dots, w_j, \dots, w_n\}$ and a set of m features $F(l) = \{f_1, \dots, f_k, \dots, f_m\}$. For each word w_j in the lexicon there is a feature vector $P(w_j) = [p^{w_j}_{f_1}, \dots, p^{w_j}_{f_k}, \dots, p^{w_j}_{f_m}]$ of numeric annotations that express the intensity of each feature f_k for this word. Thus, the lexicon contains n feature vectors of dimension m and the lexicon annotation matrix has dimension $n \times m$. Following, let $W(s) = \{w_1, \dots, w_h, \dots, w_S\}$ represent each post to analyze, with a length S and containing each input word w_h. For each word $w_h \in W(s)$, the associated feature vector P is extracted. In case the word w_h is not contained in the lexicon ($w_h \notin W(l_i)$), the resulting vector will have value zero in all positions. This process results in a matrix M containing the feature annotation for all the input words:

$$M(l_i) = \begin{pmatrix} p^{w_1}_{f_1} & \cdots & p^{w_1}_{f_k} & \cdots & p^{w_1}_{f_m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p^{w_h}_{f_1} & \cdots & p^{w_h}_{f_k} & \cdots & p^{w_h}_{f_m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p^{w_S}_{f_1} & \cdots & p^{w_S}_{f_k} & \cdots & p^{w_S}_{f_m} \end{pmatrix} \tag{2}$$

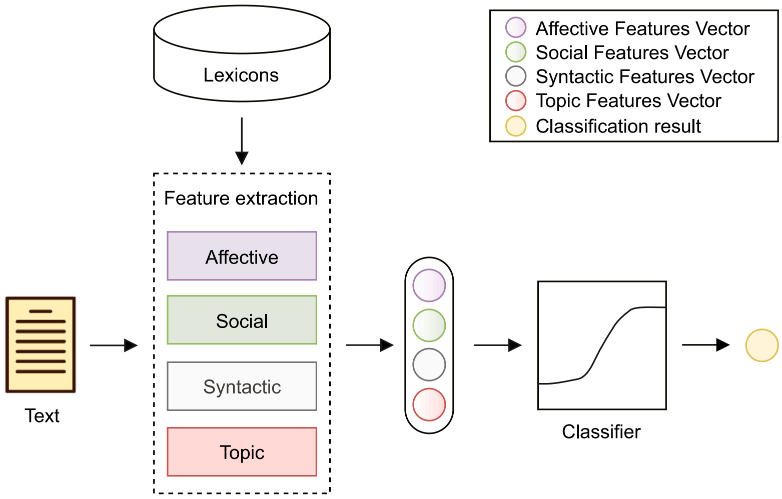


Fig. 5. General architecture representation of the M_{LF} model.

In order to compute the annotation of the post for this lexicon, the average of the feature annotation of all words is considered, resulting in a lexicon annotation vector $V(l_i) = [v_{f_1}, \dots, v_{f_k}, \dots, v_{f_m}]$, with length m, where:

$$v_{f_k} = \frac{1}{S}\sum_{h=1}^{S} p_{f_k}^{w_h} \tag{3}$$

In case no word in the post is contained in the lexicon, that is, $W(l_i) \cap W(s) = \emptyset$, the resulting matrix will have value zero in all positions, and consequently the lexicon annotation vector $V(l_i)$ will be a zero vector. Finally, the annotation vectors of all lexicons are concatenated, resulting in the annotation vector:

$$A = \bigoplus_{i=1}^{c} V(l_i) \tag{4}$$

Algorithm 1 shows the proposed feature extraction method. The function *annotation* extracts the feature vector corresponding to each input word w_h. If the word is not contained in the lexicon l_i, the resulting vector will be a null vector. This operation results in the matrix M containing the feature vector extracted from the lexicon for all the words in the post. The function *average* computes the average of the feature annotation of all words, resulting in the annotation vector V of the entire post for the lexicon l_i. Finally, the function *concat* concatenates the post's feature vectors of all the lexicons in L. The resulting vector A, containing all relevant information extracted from the lexicons, is fed to a machine learning classifier.

Algorithm 1 Lexicon-based feature extraction algorithm

Require: Set of lexicons L, each of them composed by a vocabulary $W(l_i)$ and the set of features $F(l)$; and an input post $W(s)$
Ensure: $A \in \mathbb{R}^m$
  for all $l_i \in L$ **do**
    for all $w_h \in W(s)$ **do**
      $M_{h,:} \leftarrow annotation(w_h, l_i)$
    end for
    $V(l_i) \leftarrow average(M)$
  end for
  for
    $i \leftarrow 1, c$ **do**
    $A \leftarrow concat(V(l_i))$
  end for

5.2. Distributional representation model (M_{DR})

The second model (M_{DR}) uses word embedding techniques to exploit distributional representations. An overview of the model is given in Fig. 6. Word embeddings are a type of computing distributed text representation that gives words with similar meanings

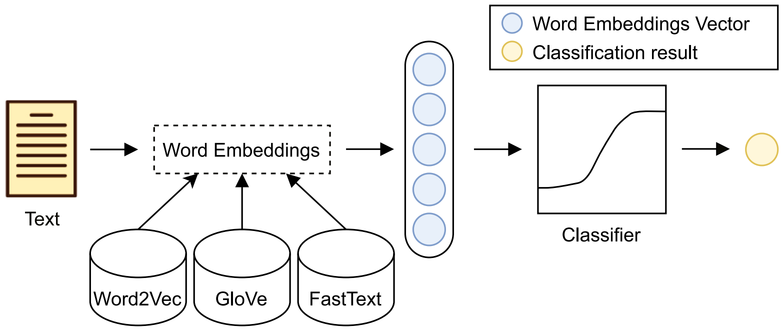


Fig. 6. General architecture representation of the M_{DR} model.

a similar representation (Mikolov et al., 2013). This enables the conversion of text into vector representations where semantic and syntactic information is encoded. However, these vectors do not enclose any information related to sentiment or cognition.

Consider a vocabulary of n words $W(e) = \{w_1, \ldots, w_j, \ldots, w_n\}$. Word embeddings are encoded by column vectors in an embedding matrix M with dimension $d \times n$. In this case, d is the dimension of the word vectors and n is the size of the vocabulary. Each column of the matrix M represents the embedding vector of a word existing in the vocabulary. The matrix components are parameters to be learned according to the word embedding technique used. Consider now a post to analyze $W(s) = \{w_1, \ldots, w_h, \ldots, w_S\}$ with a length S and containing each input word w_h. For each word w_h of the intersection $W(e) \cap W(i)$, its word embeddings vector Q_{w_h} results from the matrix–vector product $Q_{w_h} = M \cdot v_{w_j}$. In this case, v_{w_j} is the one-hot vector of word w_h. It has value one at j and zero in the rest. Finally, the embedding vector of each word in the post can be combined into a unique vector E representing the entire text:

$$E = \frac{1}{S} \sum_{h=1}^{S} Q_{w_h} \quad (5)$$

Thus, this vector E will be fed to the machine learning classifier. Some popular word embedding techniques are Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014) and FastText (Joulin et al., 2016). Each of them presents a different manner of learning the embedding matrix. Word2Vec learns it by connecting target words to their context, regardless of the frequency in which the words appear. A frequent co-occurrence of words in Word2Vec results in more training instances but does not give new information. GloVe, on the other hand, emphasizes the importance of taking into account the frequency of co-occurrence. In this way, GloVe learns embeddings so that a set of word vectors corresponds to the likelihood of these words co-occurring in the corpus. Also, whereas Word2Vec has a predictive nature, GloVe is count-based. Finally, FastText is intended to improve Word2Vec. It is based on the same principles, but instead of using words to build word embeddings, FastText uses a combination of lower-level embeddings of parts of words and characters. This reduces the amount of training data needed since each piece of text contains more information and enables generalization, as new words may contain the same characters as previously learned words. This fact allows FastText to obtain theoretically better vector representations than GloVe or Word2Vec in corpora with specific domain rare words. It constructs a word vector from its character n-grams even when the word is not contained in the training corpus.

To evaluate how each technique impacts the performance of the stress detection, the three described techniques have been used for extracting word vectors. However, to make GloVe more adequate to our data, the pre-trained model has been fine-tuned using our corpus. In this manner, the pre-trained model has been trained with the used datasets to learn the domain-specific vocabulary. This fine-tuning has been carried out with Mittens (Dingwall & Potts, 2018), an extension of GloVe that allows us to update general-purpose representations with data from a specialized domain. Thus, the unseen vocabularies are also added to the model.

5.3. Ensemble model (M_E)

Finally, the third model (M_E) takes into account both distributional representations and lexicon-based features. The general architecture representation of the model is found in Fig. 7. This model combines each text instance's word embedding vector representation with the lexicon-based feature vector drawn from it. Let $W(s) = \{w_1, \ldots, w_h, \ldots, w_S\}$ represent each post to analyze, with a length S and containing each input word w_h. The feature vector A and the embeddings vector E are computed using the above-mentioned methods. Then, these vectors are concatenated into a unified vector C containing information related to lexicon-based features and word embeddings:

$$C = A \oplus E \quad (6)$$

This vector is fed to the machine learning classifier. Thus, the information given by word embeddings is combined with the affective and lexical information given by the lexicon-based features. This information combination may improve the performance

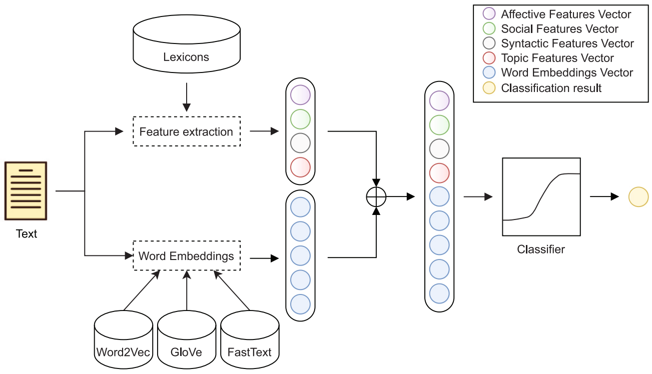


Fig. 7. General architecture representation of the M_E model.

of a classifier algorithm learning from this unified set compared to the one learning only from word embedding or lexicon-based features data.

The proposed models have been validated in three datasets from different sources (Reddit, Twitter, and personal interviews) using the data described in Section 3. Furthermore, the experiments allow us to analyze the performance of different lexicons or word embedding techniques, as described in the following section.

6. Evaluation

6.1. Methodology

To evaluate the effectiveness of the proposed models in stress classification, an experimental study has been designed according to the research questions described in Section 1. The main goal of the experiment is to provide insight into which models and techniques perform better for stress detection in different corpora. With this aim, we postulate the problem as a binary classification task that aims to detect stress evidence in text. This is accomplished by learning from the provided lexicon-based features and the computed word embeddings. The stress detection is conducted at the post level, classifying posts as either stress or non-stress. In order to analyze which lexicon-based features perform better for stress detection (**RQ1**), several experiments have been carried out, splitting them by nature or lexicon according to the feature framework proposed in Section 4. Also, the three proposed models are evaluated separately to investigate the performance of the different methods, and they have been compared with state-of-the-art solutions (**RQ2**). This evaluation has been conducted using three English public datasets to determine whether the proposed approach can achieve good results in different corpora (**RQ3**).

The proposed models have been evaluated using three different machine learning classifiers implemented with Python scikit-learn library (Pedregosa et al., 2011): Support Vector Machines (SVMs), logistic regression, and Stochastic gradient descend (SGD) classifier. SGD classifier implements regularized linear models (SVM, logistic regression, etc.) with Stochastic gradient descend (SGD) learning. In our experiments, SGD has been used with SVM as the model to fit. In addition, our work has been compared with the previous works in the field of stress detection from text that use any of the public datasets considered in this article. These works are: the lexicon approach (Thelwall, 2017); a combination of Word2Vec embeddings with LIWC features using a logistic regression classifier (Turcan & McKeown, 2019); and a Bidirectional Long-Short Term Memory (LSTM) with Attention and Word2Vec Embeddings (Winata et al., 2018). Since each method was validated only with one of the considered datasets in the original works, we evaluated them with the datasets considered in this article. In addition, we have applied Bidirectional Encoder Representations from Transformers (BERT) to the task of stress detection, using the pre-trained BERT-base (Kenton & Toutanova, 2019). We have used 10-fold validation and the weighted average of the F1-Score as the performance metric in all experiments. F1-Score is defined as the harmonic mean of the model precision and recall and enables the evaluation of a model accuracy:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{7}$$

Finally, a Friedman statistical test (Demšar, 2006) and a cross-dataset experiment have been carried out. The former allows us to further study the performance and impact of the proposed models. In contrast, the latter allows us to analyze the generalization performance. To sum up, the methodology followed for evaluating our models consists of the following steps:

1. Evaluation of the M_{LF} model, analyzing the impact of the proposed feature sets and the different lexicons.
2. Evaluation of the M_{DR} model, identifying the best word embedding technique.
3. Evaluation of the M_E model, comparing the performance of our best-performing method with existing works in stress detection from text.

Table 2
Best performance achieved with the M_{LF} model with each feature set for each dataset.

Model	Features	Dreaddit	NSE	TensiStrength
M_{LF}	Affective	0.7908	0.7289	**0.6314**
	Social	0.6762	0.6532	0.5288
	Syntactic	0.7869	0.7188	0.5341
	Topic	0.7541	0.7223	0.5415
	All	**0.8210**	**0.7750**	0.6201

Table 3
Best performance achieved with the M_{LF} model with each feature set for each dataset using RFE.

Model	Features	Dreaddit	NSE	TensiStrength
M_{LF} (RFE)	Affective	0.7931	0.7295	0.6322
	Social	0.6851	0.6645	0.5395
	Syntactic	0.7911	0.7269	0.5897
	Topic	0.7622	0.7289	0.5443
	All	**0.8257**	**0.7777**	**0.6322**

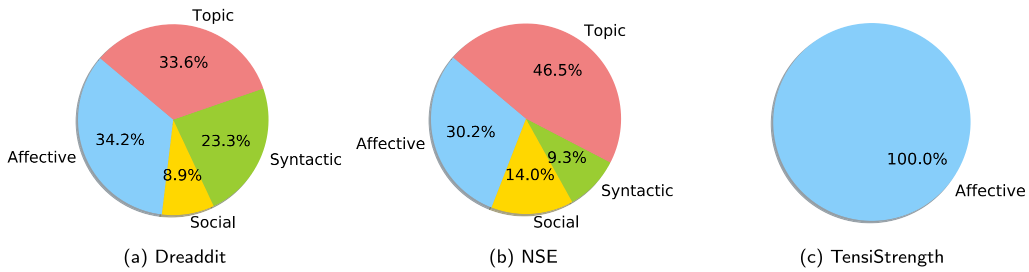


Fig. 8. Feature set representation between the selected features after running the RFE method for each dataset.

4. Comparison of the performance of all considered methods through a statistical test.
5. Analysis of the generalization performance in our best-performing method by a cross-dataset experiment.

6.2. Results

First of all, we proceed to evaluate the M_{LF} model in order to analyze whether the use of lexicon-based features can provide good results for stress classification. A first experiment has been carried out using all features of each set, and the results are shown in Table 2. The best results are obtained using Linear SVM for the NSE dataset and Logistic Regression for Dreaddit and TensiStrength datasets. As we can see, the best results are achieved when combining all the features, except for TensiStrength, where the best results are obtained using only affective features. This can be explained by observing the low results obtained in this dataset for the other sets and indicates that only affective features perform well on the microblogging source.

It is also logical to think that not all extracted features are useful in stress detection, and some might even decrease the classification accuracy. We have performed a feature selection to further study this fact and enhance our model. First, we have removed those features that presented a high correlation (greater than 0.95) with other features, as they would have almost the same effect on stress prediction. Then, we used RFE with cross-validation to obtain the best features. Table 3 shows the results of this experiment. The best results are obtained using Linear SVM for the NSE dataset and Logistic Regression for Dreaddit and TensiStrength datasets. As we can see, the results always improve after performing the feature selection.

One of the first points to highlight is that the best results are always achieved when performing a feature selection between all the feature sets combined. This shows that combining different kinds of features is a good approach to classifying stress. By analyzing which kind of features perform better, we find that the affective and syntactic features are those with better results in almost every case. Furthermore, we see that topic-related features also achieve good results for the NSE dataset, whereas social features present the lowest performance in all cases (**RQ1**). We can analyze this in more detail in Fig. 8, which shows the distribution of feature sets between the selected features for each dataset.

The figure shows that affective features present the highest rates for all datasets except for NSE. This exception can be explained by observing Table 3, which shows that the topic features presented promising results for this corpus. It is also interesting to see how for the TensiStrength dataset, all the selected features are from the affective set. The bad results obtained for this dataset using the other kinds of features explain this fact. Finally, the figure also shows the low representation of social-related features among the selected ones, as we could predict from Table 3. These results are coherent with the conclusions drawn from Fig. 4 in Section 3.

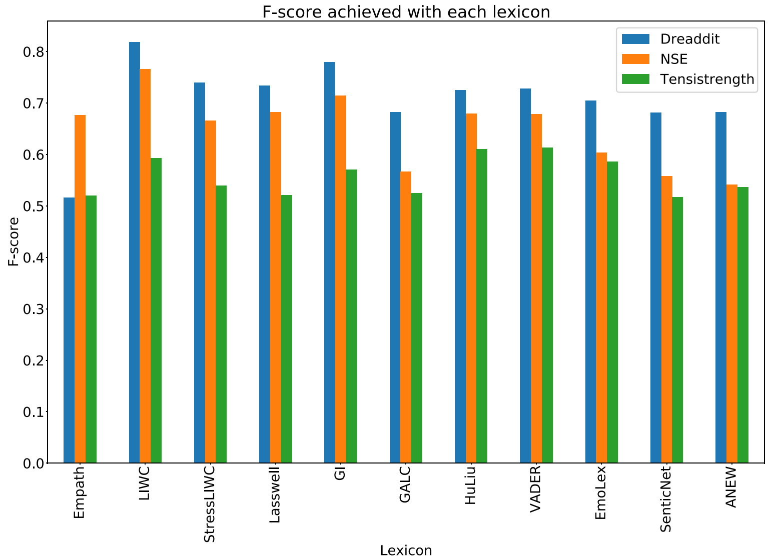


Fig. 9. Performance of the different lexicons used.

Table 4
Best performance achieved with the M_{DR} model for each word embedding technique and for each dataset.

Model	Embeddings	Dreaddit	NSE	TensiStrength
	Word2Vec	0.7901	0.6487	0.5822
M_{DR}	GloVe	0.8334	0.7813	0.7674
	FastText	**0.8487**	**0.8072**	**0.7705**

Besides, we have experimented with the features divided by the lexicon to which they belong to obtain insights into which lexicons are more interesting for stress classification. Fig. 9 shows the best performances of the M_{LF} model using only selected features from each lexicon. We can observe that LIWC and GI seem to be the lexicons that perform better for the Dreaddit and NSE datasets. If we analyze Fig. 9 along with Fig. 3, we can see that these lexicons are the only ones that provide features of all kinds. Also, we see that these lexicons are the only ones that contribute with syntactic features, one of the most relevant sets according to the previously commented results. However, in the TensiStrength dataset, better results are obtained with Hu–Liu and VADER lexicons. As previously mentioned, affective features perform much better than other types of features in this dataset. In addition, it contains shorter posts. This explains why the best results are obtained with lexicons more focused on affective features. Moreover, VADER is specifically designed for short texts and is particularly effective on texts coming from social networks.

The results of the M_{LF} model indicate that stress classification using affective, syntactic, social, and topic-related features can be successful for datasets coming from Reddit and interviews (achieving an F-score of 0.8257 for the Dreaddit dataset). However, these features do not perform so well for short text data coming from microblogging posts (as we can see from the results for the TensiStrength dataset). Still, the performance obtained using only lexicon-based features (M_{LF}) is comparable to more complex methods, which take into account the visual attributes existing in social media data (Winata et al., 2018). Besides, the results suggest that affective features are those which perform better for the stress classification; and that the best results come from their combination (**RQ1**).

Once obtained the results for the M_{LF} model, we proceed to evaluate the distributional representation model (M_{DR}). For this purpose, an experiment has been carried out using three different word embedding techniques: Word2Vec, GloVe, and FastText.

Results shown in Table 4 indicate that the use of distributional representations (M_{DR}) improves the performance over the M_{LF} model. Concerning the word embedding technique comparison, we can observe that the FastText method provides the best performance. Nevertheless, the differences between FastText and GloVe performances are pretty slight except for the NSE dataset. As explained in Section 5, FastText is supposed to achieve better results in corpora with rare domain-specific words. The fine-tuning of the GloVe embeddings to each dataset should enhance the performance, but we think that a greater amount of data would be needed to make it overtake the performance of FastText. We can support this hypothesis by analyzing the performance of Word2Vec. As can be observed, this method is more sensitive to corpora with domain-specific words. Therefore, it performs considerably worse than the other two, given that it has not been fine-tuned.

Analyzing the results from the M_{DR} model, we can observe that the use of distributional representations supposes an interesting approach. The improvement achieved by this method compared to the M_{LF} model is especially significant in short texts. The F-score

Table 5
Best performances achieved with the M_E model (lexicon-based features + distributional representation) and comparisons of these results with the state-of-the-art baselines in stress classification from text and with the other models proposed in this article.

Method	Dreaddit	NSE	TensiStrength
Lexicon approach (Thelwall, 2017)	0.7040	0.5921	0.7130
LogReg + LIWC + Word2Vec (Turcan & McKeown, 2019)	0.7980	0.6987	0.6270
BiLSTM + Word2Vec (Winata et al., 2018)	0.7460	0.7430	0.7410
BERT base	0.8479	0.7809	0.7562
M_{LF} (Features)	0.8257	0.7777	0.6322
M_{DR} (FastText)	0.8487	0.8072	0.7705
M_E (Features + FastText)	**0.8604**	**0.8372**	**0.7750**

of 77.05% achieved for the TensiStrength dataset supposes an improvement of more than 10 points compared to the previous model. For the NSE dataset, the improvement is also relatively significant. This model also improves the results for corpora with longer texts, as shown by the F-score of 84.87% achieved for Dreaddit. Regarding the performance comparison between classifiers, the best results are obtained using SGD classifier for the Dreaddit dataset, Logistic Regression for the NSE dataset, and Linear SVM for the TensiStrength dataset.

Once the results for the distributional representation and lexicon-based models have been analyzed separately, we proceed to analyze their combination (M_E model). For this experiment, we took those features that obtained the best results for each dataset and combined them with word embeddings. Some examples of the selected features are: "arousal", "disgust" and "deception".

Table 5 shows the best results obtained for the M_E model and compares them with other models proposed in this paper. The results show that the proposed method achieves considerably good performance for all datasets. Let us compare the proposed M_E model with the other models proposed in our work. We see that the combination of lexicon-based features with distributional representations enhances the performance in all datasets. This enhancement is especially high for data coming from personal interviews (NSE dataset), where the improvement in the performance reaches 3 points compared to the distributional representation model (M_{DR}) and 6 points compared to the lexicon-based features model (M_{LF}). Concerning the data coming from Reddit (Dreaddit dataset), the results show an improvement of 1 point compared to M_{DR}; and almost 3 points compared to M_{LF}. Finally, the improvement in the TensiStrength data is smaller compared to the M_{DR} model but very high compared to the M_{LF} model. This can be easily explained if we consider the lower performance of the lexicon-based features for this dataset.

The M_E model surpasses the 80% F-score in the Dreaddit (86.04%) and NSE datasets (83.72%) and yields to 77.50% in the TensiStrength dataset. Regarding classifiers, the best results are obtained using Logistic Regression for Dreaddit and Tensistrength datasets; and Linear SVM for the NSE dataset. The results obtained for each model with each classifier are publicly available online for the interested reader.[1] In addition, our work has been compared with previous works in the field of stress detection from text which uses any of the public datasets considered in this article. These works are: the lexicon approach (Thelwall, 2017); a combination of Word2Vec embeddings with LIWC features (Turcan & McKeown, 2019); and a Bidirectional LSTM with Attention and Word2Vec Embeddings (Winata et al., 2018). Furthermore, we have applied BERT to the task of stress detection, using the pre-trained BERT-base (Kenton & Toutanova, 2019). As a result, we can appreciate that our model outperforms the current state-of-the-art methods for all datasets. Whereas a greater amount of data could be needed to benefit from the advantages of deep neural networks, the proposed solution yields good results on small datasets.

To further study the performance and impact of the proposed models, a Friedman statistical test (Demšar, 2006) has been carried out. This test aims to determine if we may conclude from the sample of results that there is a difference between the classification methods. As a result, the Friedman test outputs a ranking of methods regarding their effectiveness in different datasets. A lower ranking indicates a better performance of the specific method than the rest.

The first step in calculating the Friedman test is to convert the original results to ranks. Let r_i^j be the rank of the jth algorithm on the ith dataset, and k and n the number of methods and datasets respectively. Friedman's test compares the average ranks of the methods $R_j = \frac{1}{n}\sum_i r_i^j$ and states that the Friedman statistic under the null hypothesis (that is, all the algorithms are equal, so their ranks are also equal) with $k-1$ degrees of freedom is:

$$X_F^2 = \frac{12n}{k(k+1)}\left(\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right) \tag{8}$$

However, Iman and Davenport (1980) proposed a better static distributed based on the F-distribution with $k-1$ and $(k-1)(n-1)$ degrees of freedom:

$$F_F = \frac{(n-1)X_F^2}{n(k-1)-X_F^2} \tag{9}$$

We perform the test with an α value of 0.1, $k = 19$ (the number of methods included in the analysis), and $n = 3$ (number of datasets). With these data, $X_F^2 = 39.72$, $F_F = 5.56$, and the critical value $F(k-1,(k-1)(n-1)) = 1.64$. As $F_F > F(18,36)$, the null hypothesis of the Friedman test is rejected, and the results are statistically relevant. For simplicity, Table 6 shows the best five approaches according to their ranks, as computed by the Friedman test.

[1] https://gsi.upm.es/~smunoz/stress-text/additional_material_stress_text.pdf.

Table 6
Friedman rank for the top-5 methods.

Method	Rank
M_E (Log. Reg. + Features + FastText embeddings)	2
M_E (Lin. SVM + Features + FastText embeddings)	3
M_{DR} (Log. Reg. + FastText embeddings)	5.5
M_E (Lin. SVM + Features + GloVe embeddings)	5.67
BERT base	7

Table 7
Averaged F1-Scores for the best-performing classification method in a cross-dataset evaluation.

	Dreaddit	TensiStrength	NSE
Dreaddit	0.8604	0.6883	0.6837
TensiStrength	0.7764	0.7750	0.6615
NSE	0.7370	0.6176	0.8373

As can be noted, Friedman's test confirms the combination of lexicon-based features with distributional representations as the best-performing approach for detecting stress in texts (**RQ2**). The two lower ranks are obtained when combining the extracted features with the FastText word embeddings using Logistic Regression and Linear SVM (ranks of 2 and 3, respectively). These two methods present results significantly better than the others. The results demonstrate the effectiveness of combining lexicon-based features with distributional representations for detecting stress from text, especially using FastText embeddings. Regarding the comparison with state-of-the-art works, only the use of BERT is ranked among the top-5 analyzed methods.

Once the best classification method has been identified, a cross-dataset experiment is conducted to analyze the generalization performance. This experiment has been performed using our best-performing method: the M_E model combining lexicon-based features and FastText embeddings with a logistic regression classifier. The classifier has been trained with data from a specific dataset and evaluated with other datasets. Results are shown in Table 7, where the rows indicate the training dataset and the columns indicate the test dataset.

As expected, the differences between the datasets led to a relatively high performance drop. This drop is exceptionally high when training with data from the Reddit corpus and evaluating with data from interviews or Twitter. However, reasonably good results are obtained when training with data from interviews or Twitter and evaluating with data from Reddit. The greater length in data from Reddit compared to the rest of the data can explain this difference: longer posts could contain a richer vocabulary and more relevant features which are not present in posts from interviews or micro-blogging platforms. The classifier can draw conclusions from these features when training with Reddit posts that are irrelevant in the other datasets. However, the relevant features existing in short posts are also present in longer posts. Besides, among the NSE and TensiStrength datasets, even if the length of the posts is similar for both corpora, the differences in the nature of the data also led to a performance drop. This can be due to the expressiveness difference between data coming from Twitter and data coming from interviews.

The results obtained from the evaluation confirm that an approach for stress classification from text combining lexicon-based features with distributional representations can achieve good performance on different datasets (**RQ3**). The yielded F-score with this method varies between 77% and 86%, demonstrating its effectiveness.

7. Discussion

Previously, we presented three research questions that drove this work (Section 1). The first question (**RQ1**) was concerned with comparing different kinds of lexicon-based features in terms of stress detection performance. In this respect, we propose a feature framework based on multiple existing lexicons for identifying and categorizing features among four sets: affective, topic, social, and syntactic related. The obtained results suggest that affective features perform better in text classification of stress. However, including syntactic and topic-related features can help enhance classification performance. Regarding the different lexicons, LIWC and GI are those which led to better performance. These lexicons provide features of all kinds and are the only ones that provide syntactic features.

In the second question (**RQ2**), we focused on which is the best-performing method for detecting stress of the individuals in different corpora. In this regard, we presented and evaluated three different stress classification models: a lexicon-based features model, a distributional representation model, and an ensemble model combining lexicon-based features with distributional representations. The evaluation of these models in terms of F-score points out the combination of lexicon-based features with distributional representations as the best-performing approach. This approach yields significant performance, as shown by the experiments. The obtained F-scores surpass the 80% in the Dreaddit (86.04%) and NSE datasets (83.72%) and yield 77.50% in the TensiStrength dataset. The Friedman test confirms these results and demonstrates that they are statistically relevant. Besides, to analyze how different word embedding techniques impact the text classification performance, three popular word embedding techniques have been exploited: FastText, GloVe, and Word2Vec. The statistical results also identify FastText embeddings as the best-performing distributional representation technique, given its better performance in corpora with rare domain-specific words. The fine-tuning of the GloVe embeddings achieves comparable performances, but a greater amount of data would be needed to make it surpass the outcomes of FastText. In this line, Word2Vec without fine-tuning performs considerably worse than the other two. These results point out the importance of domain-specific word embeddings.

Finally, the third question (**RQ3**) pertained to studying the capability of a machine learning approach to classify stress in texts with good performance on multiple datasets. Previous studies had demonstrated the effectiveness of this method on specific datasets (Turcan & McKeown, 2019; Winata et al., 2018), but none of them had performed a cross-dataset evaluation using multiple public English corpora. Our approach has been evaluated on three different public English datasets obtaining good performance metrics on all of them. These results confirm the potential of textual information for detecting stress in several scenarios.

Also, the experiments have shown that the proposed model outperforms previous works in the field of stress detection from text. Our method presents two significant advances compared with previous works: the evaluation and validation with data from multiple corpora and the exhaustive analysis of feature extraction techniques. Some of the previous works have reached good results with the use of deep neural networks for predicting stress from Sina Weibo posts (Cao et al., 2021; Lin et al., 2016; Wang, Zhang, Cao and Feng, 2020). However, these models require large quantities of labeled data and often suffer from higher inference times (Yang, Shou, Gong, Lin, & Jiang, 2020). To the aim of our knowledge, this is the first study that proposes a method for stress detection from text and validates it on different public English datasets. Besides, our proposal exploits a wide variety of feature extraction methods, including multiple lexicons and three word embeddings techniques. In this manner, apart from carrying out an exhaustive analysis of which techniques achieve better performance, the proposed method improves the performance of previous work in all the existing datasets considered (Thelwall, 2017; Turcan & McKeown, 2019; Winata et al., 2018).

When comparing with other stress detection methods or approaches (Alberdi, Aztiria, & Basarab, 2016; Can, Arnrich et al., 2019; Greene et al., 2016), we can see that stress classification from text achieves comparable performances while entailing less complex or expensive scenarios. Furthermore, the proposed techniques can be integrated into existing software solutions, enabling stress detection in real-time and reducing complex hardware implementations' costs. For example, these methods could be integrated into existing software solutions such as a company communication platform or even social messaging apps. This can allow managers to easily monitor workers' stress levels and users to have an insight into their state.

8. Conclusion

This work proposes a method for detecting psychological stress from texts. The presented approach uses two different kinds of information sources: (i) affective, syntactic, topic, and social-related features and (ii) distributional representations. The first method proposes to benefit from existing lexicons for generating features that can be used for stress detection. A feature framework is presented to provide insight into which kinds of features and lexicons perform better. According to this, the features have been split into four different sets depending on their nature. A total of eleven different lexicons have been used for extracting these features. As for the second method, word vectors have been generated for extracting semantic information from the text. We have analyzed the use of three different word embedding techniques: Word2Vec, GloVe, and FastText. Finally, the work proposes a combination of these approaches aiming to improve the performance scores achieved by a classifier algorithm that learns from this unified set instead of only distributional representations or lexicon-based feature data. In order to conduct a comparative experimental study that enables the analysis and evaluation of all the models, three English public datasets from different sources and nature have been used. Also, an exhaustive analysis of the data contained in these datasets has been carried out.

The experiments show that the affective, syntactic, topic, and social features can obtain considerably high scores, achieving an F-Score of 82.57% in the Dreaddit dataset. Therefore, it can be reasonable to assume the convenience of these features when performing stress detection. Regarding word embeddings, it has been stated that they can achieve accurate representations of the analyzed text, allowing simple classifiers to reach elevated classification scores: 84.87% for the Dreaddit dataset. Hence, given the promising results obtained, we conclude that the application of this method can be further studied. Besides, the combination of distributional representations with lexicon-based features significantly improves the performance in all the cases reaching an F-score of 86.04% for the Dreaddit dataset. A statistical analysis has been conducted to empirically verify that combining information from varied lexicon-based features with distributional representations is suitable for improving stress classification performance. This test verifies the conclusions drawn from the experiments and points out the enhancement compared with previous works. Finally, a cross-dataset study has been carried out in order to analyze how the proposed method performs when using data from different corpora.

In conclusion, this paper presents a machine learning method for classifying stress text from different sources, such as social media, microblogging sites, and interviews. The proposed approach exploits several lexicons and three word embedding techniques. Besides, the evaluation presented provides a baseline for other researchers. According to the results of this work, we think that it is worth advancing in stress classification from text. The presented advancement in automatic detection of stress from text can be beneficial for detecting stress early, which is crucial for fastening its diagnostic and reducing its high growing impact. Our approach provides an excellent compromise between performance and required computational resources since it is based on simple features and outperforms more complex architectures such as BERT. Moreover, our approach can be easily interpreted to understand better how social, personal, and environmental factors influence stress and find suitable regulation methods. Nevertheless, there exist more complex pre-trained models specially designed for mental health, such as MentalBERT (Ji et al., 2021), which outperform our performance by 4% of F-score in the Dreaddit and TensiStrength dataset and yield a slightly worse performance (2% less of F-score) in the NSE dataset. This result encourages us to explore two research directions. Firstly, our current solution can be easily deployed in real environments with good performance without requiring high demanding computational resources. Additionally, we aim to explore how pre-trained models such as MentalBERT can be trained taking into account our feature framework. An additional possible line of future work would be to extend the domain of the proposed method to other languages or even to different paradigms, like depression detection. Besides, to deepen the use of word embeddings for studying semantic similarity between texts could be another possible future line to enhance the classification performance.

CRediT authorship contribution statement

Sergio Muñoz: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Visualization. **Carlos A. Iglesias:** Conceptualization, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work is supported by the Spanish Ministry of Science and Innovation under the R&D project COGNOS (PID2019-105484RB-I00). We would also to thank the authors of the used datasets for their work and for providing free access to them, Turcan and McKeown (2019), Zuo et al. (2012), Winata et al. (2018) and Thelwall (2017).

References

Affanni, A., Bernardini, R., Piras, A., Rinaldo, R., & Zontone, P. (2018). Driver's stress detection using Skin potential response signals. *Measurement*, *122*, 264–274.
Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, *59*, 49–75.
Alberdi, A., Aztiria, A., Basarab, A., & Cook, D. J. (2018). Using smart offices to predict occupational stress. *International Journal of Industrial Ergonomics*, *67*, 13–26.
Almatarneh, S., & Gamallo, P. (2018). A lexicon based method to search for extreme opinions. *PLoS One*, *13*(5), Article e0197816.
Andreou, E., Alexopoulos, E. C., Lionis, C., Varvogli, L., Gnardellis, C., Chrousos, G. P., et al. (2011). Perceived stress scale: reliability and validity study in Greece. *International Journal of Environmental Research and Public Health*, *8*(8), 3287–3298.
Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, *77*, 236–246.
Araque, O., & Iglesias, C. A. (2020). An approach for radicalization detection based on emotion signals and semantic similarity. *IEEE Access*, *8*, 17877–17891.
Bandhakavi, A., Wiratunga, N., Padmanabhan, D., & Massie, S. (2017). Lexicon based feature extraction for emotion text classification. *Pattern Recognition Letters*, *93*, 133–142.
Banerjee, S., & Shaikh, N. F. (2021). A survey on mental health monitoring system via social media data using deep learning framework. In *Techno-societal 2020* (pp. 879–887). Cham: Springer International Publishing.
Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*: *Technical report, Technical report C-1*, the center for research in psychophysiology.
Burman, R., & Goswami, T. G. (2018). A systematic literature review of work stress. *International Journal of Management Studies*, *5*(3–9), 112–132.
Cambria, E., Liu, Q., Decherchi, S., Xing, F., & Kwok, K. (2022). SenticNet 7: a commonsense-based neurosymbolic AI framework for explainable sentiment analysis. *Proceedings of LREC 2022*.
Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of Biomedical Informatics*, *92*, Article 103139.
Can, Y. S., Chalabianloo, N., Ekiz, D., & Ersoy, C. (2019). Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors*, *19*(8), 1849.
Cao, L., Zhang, H., Li, N., Wang, X., Ri, W., & Feng, L. (2021). Category-aware chronic stress detection on microblogs. *IEEE Journal of Biomedical and Health Informatics*, *26*(2), 852–864.
Chan, S. F., & La Greca, A. M. (2020). Perceived stress scale (PSS). In *Encyclopedia of behavioral medicine* (pp. 1646–1648). Cham: Springer International Publishing.
Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digital Medicine*, *3*(1), 1–11.
Cox, T., & Griffiths, A. (1995a). The nature and measurement of work stress: theory and practice. In *The evaluation of human work: a practical ergonomics methodology*. London: Taylor & Francis.
Cox, T., & Griffiths, A. (1995b). Work-related stress: nature and assessment. In *IEE colloquium on stress and mistake-making in the operational workplace* (pp. 1/1–1/4).
Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, *49*(3), 803–821.
de Santos Sierra, A., Ávila, C. S., Casanova, J. G., & del Pozo, G. B. (2011). A stress-detection system based on physiological signals and fuzzy logic. *IEEE Transactions on Industrial Electronics*, *58*(10), 4857–4865.
Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(Jan), 1–30.
Dewe, P. J., O'Driscoll, M. P., & Cooper, C. L. (2012). Theories of psychological stress at work. In *Handbook of occupational health and wellness* (pp. 23–38). Boston, MA: Springer US.
Dingwall, N., & Potts, C. (2018). Mittens: an extension of glove for learning domain-specialized representations. In *NAACL-HLT (2)* (pp. 212–217).
Fast, E., Chen, B., & Bernstein, M. S. (2016). Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 4647–4657).
Ferdous, R., Osmani, V., & Mayora, O. (2015). Smartphone app usage as a predictor of perceived stress levels at workplace. In *2015 9th international conference on pervasive computing technologies for healthcare (PervasiveHealth)* (pp. 225–228).
French, J. R., Caplan, R. D., & Van Harrison, R. (1982). *The mechanisms of job stress and strain, Vol. 7*. New York: Chichester, J. Wiley.
Fu, X., Yang, J., Li, J., Fang, M., & Wang, H. (2018). Lexicon-enhanced LSTM with attention for general sentiment analysis. *IEEE Access*, *6*, 71884–71891.
Gao, H., Yüce, A., & Thiran, J.-P. (2014). Detecting emotional stress from facial expressions for driving safety. In *2014 IEEE international conference on image processing (ICIP)* (pp. 5961–5965). IEEE.
Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P. G., et al. (2017). Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, *31*, 89–101.
Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, *69*, 214–224.
Gilbert, E., & Hutto, C. J. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *81*, In *Eighth international conference on weblogs and social media (ICWSM-14)* (p. 82).
Giuntini, F. T., Cazzolato, M. T., dos Reis, M. d. J. D., Campbell, A. T., Traina, A. J., & Ueyama, J. (2020). A review on recognizing depression in social networks: challenges and opportunities. *Journal of Ambient Intelligence and Humanized Computing*, *11*(11), 4713–4729.
Glanz, K., Rimer, B. K., & Viswanath, K. (2008). *Health behavior and health education: theory, research, and practice*. San Francisco, CA: John Wiley & Sons.

Greene, S., Thapliyal, H., & Caban-Holt, A. (2016). A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine*, *5*(4), 44–56.
Hansen, J. H. L., & Patil, S. (2007). Speech under stress: Analysis, modeling and recognition. In *Speaker classification i: fundamentals, features, and methods* (pp. 108–137). Springer Berlin Heidelberg.
Hassard, J., Teoh, K., Cox, T., Cosmar, M., Gründler, R., Flemming, D., et al. (2014). *Calculating the cost of work-related stress and psychosocial risks*: *Technical report*, European Agency for Safety and Health at Work.
Holroyd, K. A., & Lazarus, R. S. (1982). Stress, coping and somatic adaptation. *Handbook of stress: theoretical and clinical aspects*, 21–35.
Iman, R. L., & Davenport, J. M. (1980). Approximations of the critical region of the Friedman statistic. *Communications in Statistics. Theory and Methods*, *9*(6), 571–595.
Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., & Cambria, E. (2021). MentalBERT: Publicly available pretrained language models for mental healthcare. arXiv preprint arXiv:2110.15621.
Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). Fasttext. zip: Compressing text classification models. arXiv preprint arXiv:1612.03651.
Kassymova, G., Tokar, O., Tashcheva, A., Gridneva, S., Bazhenova, N., Shpakovskaya, E., et al. (2019). Impact of stress on creative human resources and psychological counseling in crises. *International Journal of Education and Information Technologies*, *13*(1), 26–32.
Kenton, J. D. M.-W. C., & Toutanova, L. K. (2019). BERT: PRe-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171–4186).
Kessler, J. S. (2017). Scattertext: a browser-based tool for visualizing how corpora differ. In *Proceedings of ACL-2017 system demonstrations* (pp. 85–90). Vancouver, Canada: Association for Computational Linguistics.
Khoo, C. S., & Johnkhan, S. B. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, *44*(4), 491–511.
Kim, J. J., & Diamond, D. M. (2002). The stressed hippocampus, synaptic plasticity and lost memories. *Nature Reviews Neuroscience*, *3*(6), 453–462.
Lasswell, H. D., & Namenwirth, J. Z. (1969). *The Lasswell value dictionary*. New Haven.
Lewin, K. (1936). A dynamic theory of personality. *Journal of Heredity*, *27*(11), 441–442.
Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., et al. (2014). User-level psychological stress detection from social media using deep neural network. In *Proceedings of the 22nd ACM international conference on multimedia* (pp. 507–516).
Lin, H., Jia, J., Nie, L., Shen, G., & Chua, T.-S. (2016). What does social media say about your stress?. In *IJCAI* (pp. 3775–3781).
Liu, Y.-d., Zeng, H.-q., Li, R.-l., & Hu, Y.-f. (2004). Polarity text filtering based on semantic analysis. *Journal-China Institute of Communications*, *25*(7), 78–85.
Martin, L., & Pu, P. (2014). Prediction of helpful reviews using emotions extraction. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 28* (pp. 1551–1557).
Maxhuni, A., Hernandez-Leal, P., Morales, E. F., Sucar, L. E., Osmani, V., & Mayora, O. (2021). Unobtrusive stress assessment using smartphones. *IEEE Transactions on Mobile Computing*, *20*(6), 2313–2325.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).
Mohammad, S., & Turney, P. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26–34).
Murray, H. A. (1938). *Explorations in personality*. Oxford: Oxford Univ. Press.
Novais, P., & Carneiro, D. (2016). The role of non-intrusive approaches in the development of people-aware systems. *Progress in Artificial Intelligence*, *5*(3), 215–220.
Osman, A., Wong, J. L., Bagge, C. L., Freedenthal, S., Gutierrez, P. M., & Lozano, G. (2012). The depression anxiety stress Scales—21 (DASS-21): further examination of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology*, *68*(12), 1322–1338.
Pang, J., Rao, Y., Xie, H., Wang, X., Wang, F. L., Wong, T.-L., et al. (2019). Fast supervised topic models for short text emotion detection. *IEEE Transactions on Cybernetics*, *51*(2), 815–828.
Panicker, S. S., & Gayathri, P. (2019). A survey of machine learning techniques in physiology based mental stress detection systems. *Biocybernetics and Biomedical Engineering*, *39*(2), 444–469.
Parent-Thirion, A., Vermeylen, G., van Houten, G., Lyly-Yrjänäinen, M., Biletta, I., & Cabrita, J. (2012). *Fifth European working conditions survey*: *Technical report*, European Foundation for the Improvement of Living and Working Conditions.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.
Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001, Vol. 71* (p. 2001). Mahway: Lawrence Erlbaum Associates.
Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543).
for Safety, E. A., & at Work, H. (2013). *European opinion poll on occupational safety and health*: *Technical report*, European Agency for Safety and Health at Work.
Saute, S., Murphy, L., Colligan, M., Swanson, N., Hurrell, J., Scharf, F., et al. (1999). *Stress at work*: *Technical Report*, National Institute for Occupational Safety and Health.
Selye, H. (1956). *The stress of life*. New York: McGraw-Hill.
Skaik, R., & Inkpen, D. (2020). Using social media for mental health surveillance: A review. *ACM Computing Surveys*, *53*(6), 1–31.
Society, A. P. (2015). Stress and wellbeing. How Australians are coping with life.
Statista, I. (2018). Number of social media users worldwide from 2010 to 2021 (in billions). *Statista*.
Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: a computer approach to content analysis*. MIT Press.
Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: a scoping review. *Translational Psychiatry*, *10*(1), 1–26.
The American Institute of Stress (2013). Definition of stress. https://www.stress.org/what-is-stress/ Online; accessed 17 September 2020.
Thelwall, M. (2017). Tensistrength: Stress and relaxation magnitude detection for social media texts. *Information Processing & Management*, *53*(1), 106–121.
Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine learning in mental health: A systematic review of the hci literature to support the development of effective and implementable ml systems. *ACM Transactions on Computer-Human Interaction*, *27*(5), 1–53.
Tomba, K., Dumoulin, J., Mugellini, E., Abou Khaled, O., & Hawila, S. (2018). Stress detection through speech analysis. In *ICETE (1)* (pp. 560–564).
Turcan, E., & McKeown, K. (2019). Dreaddit: A Reddit Dataset for Stress Analysis in Social Media. In *Proceedings of the tenth international workshop on health text mining and information analysis (LOUHI 2019)* (pp. 97–107).
Vanitha, V., & Krishnan, P. (2017). Real time stress detection system based on EEG signals. *Biomedical Research — Tokyo*, 271–275.
Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, *67*(10), 870–886.
Wang, W., Hernandez, I., Newman, D. A., He, J., & Bian, J. (2016). Twitter analysis: Studying US weekly trends in work stress and emotion. *Applied Psychology*, *65*(2), 355–378.
Wang, X., Zhang, H., Cao, L., & Feng, L. (2020). Leverage social media for personalized stress detection. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 2710–2718).

Wang, S., Zhou, W., & Jiang, C. (2020). A survey of word embeddings based on deep learning. *Computing*, *102*(3), 717–740.
Widanti, N., Sumanto, B., Rosa, P., & Miftahudin, M. F. (2015). Stress level detection using heart rate, blood pressure, and GSR and stress therapy by utilizing infrared. In *2015 international conference on industrial instrumentation and control (ICIC)* (pp. 275–279). Ieee.
Winata, G. I., Kampman, O. P., & Fung, P. (2018). Attention-based LSTM for psychological stress detection from spoken language using distant supervision. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 6204–6208).
Xiong, B., Skitmore, M., & Xia, B. (2015). Exploring and validating the internal dimensions of occupational stress: evidence from construction cost estimators in China. *Construction Management and Economics*, *33*(5–6), 495–507.
Yang, Z., Shou, L., Gong, M., Lin, W., & Jiang, D. (2020). Model compression with two-stage multi-teacher knowledge distillation for web question answering system. In *Proceedings of the 13th international conference on web search and data mining* (pp. 690–698).
Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, *60*(2), 617–663.
Zuo, X., Lin, L., & Fung, P. (2012). A multilingual database of natural stress emotion. In *Proceedings of the eight international conference on language resources and evaluation (LREC'12)* (pp. 1174–1178).

Sergio Muñoz received the graduate and master's degrees in Telecommunications Engineering from the Universidad Politécnica de Madrid (UPM), Spain, in 2016 and 2017, respectively, where he is currently pursuing the Ph.D. degree. He is currently a Teaching Assistant with the Technical University of Madrid. His research interests include ambient intelligence and agent-based simulation. The main topic of his thesis is the adaptation of smart environments to users' emotions, to enhance well-being and performance.

Carlos A. Iglesias received the telecommunications engineering degree and the Ph.D. degree in telecommunications from the Universidad Politécnica de Madrid (UPM), Spain, in 1993 and 1998, respectively. He is currently a University Professor with the Telecommunications Engineering School, UPM, where he has been leading the Intelligent Systems Group, since 2014. He has been the Principal Investigator on numerous research grants and contracts in the field of advanced social and the IoT systems, funded by the regional, national, and European bodies. His primary research interests include social computing, multiagent systems, information retrieval, sentiment and emotion analysis, linked data, and web engineering.

3.2.2 Prediction of stress levels in the workplace using surrounding stress

Title	Prediction of stress levels in the workplace using surrounding stress
Authors	Muñoz, Sergio and Iglesias, Carlos Á. and Mayora, Oscar and Osmani, Venet
Journal	Information Processing & Management
Impact factor	JCR 2021 7.466
ISSN	03064573
Volume	59
Year	2022
Keywords	Stress prediction, Surrounding stress, Ambient stress, Machine learning, Workplace stress
Pages	103064
Online	`https://linkinghub.elsevier.com/retrieve/pii/S0306457322001650`
Abstract	Occupational stress has a significant adverse effect on workers' well-being, productivity, and performance and is becoming a major concern for both individual companies and the overall economy. To reduce negative consequences, early detection of stress is a key factor. In response several stress prediction methods have been proposed, whose primary aim is to analyse physiological and behavioural data. However, evidence suggests that solutions based on physiological and behavioural data alone might be challenging when implemented in real-world settings. These solutions are sensitive to data problems arising from losses in signal quality or alterations in body responses, which are common in everyday activities. The contagious nature of stress and its sensitivity to the surroundings can be used to improve these methods. In this study, we sought to investigate automatic stress prediction using both surrounding stress data, which we define as close colleagues' stress levels and the stress level history of the individuals. We introduce a real-life, unconstrained study conducted with 30 workers monitored over 8 weeks. Furthermore, we propose a method to investigate the effect of stress levels of close colleagues on the prediction of an individual's stress levels. Our method is also validated on an external, independent dataset. Our results show that surrounding stress can be used to improve stress prediction in the workplace, where we achieve 80% of F-score in predicting individuals' stress levels from the surrounding stress data in a multiclass stress classification.

Information Processing and Management 59 (2022) 103064

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Prediction of stress levels in the workplace using surrounding stress

Sergio Muñoz [a,*], Carlos Á. Iglesias [a], Oscar Mayora [b], Venet Osmani [b]

[a] Intelligent Systems Group, Universidad Politécnica de Madrid, Spain
[b] Fondazione Bruno Kessler Research Institute, Trento, Italy

ARTICLE INFO

Keywords:
Stress prediction
Surrounding stress
Ambient stress
Machine learning
Workplace stress

ABSTRACT

Occupational stress has a significant adverse effect on workers' well-being, productivity, and performance and is becoming a major concern for both individual companies and the overall economy. To reduce negative consequences, early detection of stress is a key factor. In response several stress prediction methods have been proposed, whose primary aim is to analyse physiological and behavioural data. However, evidence suggests that solutions based on physiological and behavioural data alone might be challenging when implemented in real-world settings. These solutions are sensitive to data problems arising from losses in signal quality or alterations in body responses, which are common in everyday activities. The contagious nature of stress and its sensitivity to the surroundings can be used to improve these methods. In this study, we sought to investigate automatic stress prediction using both surrounding stress data, which we define as close colleagues' stress levels and the stress level history of the individuals. We introduce a real-life, unconstrained study conducted with 30 workers monitored over 8 weeks. Furthermore, we propose a method to investigate the effect of stress levels of close colleagues on the prediction of an individual's stress levels. Our method is also validated on an external, independent dataset. Our results show that surrounding stress can be used to improve stress prediction in the workplace, where we achieve 80% of F-score in predicting individuals' stress levels from the surrounding stress data in a multiclass stress classification.

1. Introduction

Workplace changes have brought about new challenges to organisations and employees, leading to increasingly competitive and stressful working environments (Rigó et al., 2021).

Particularly in Europe, a poll found that 51% of European employees consider stress a common problem in their workplaces (for Safety & at Work, 2013). Also, the European Foundation for the Improvement of Living and Working Conditions stated that 22% of Europeans suffer from stress and fatigue (Parent-Thirion et al., 2007), finding that 40% of workers think that stress is not effectively addressed in their workplace (Parent-Thirion et al., 2012). Similar results can be observed in the United States, where reports say that around the 40% of workers see their job as quite a bit or extremely stressful, and 29% of them recognise feeling very or highly stressed at work (Gallup, 2021). According to the American Psychological Association (Association, 2019), 75% of adults reported experiencing moderate to high levels of stress. Reports available from other countries such as China (Le et al., 2020) or Australia (Ribeiro Santiago et al., 2020) illustrate the impact of occupational stress around the world.

The high impact that workplace stress has on workers' motivation, job performance, well-being, and productivity has made it one of the main challenges for organisations. Research has demonstrated that employees with high levels of stress have lower

* Corresponding author.
E-mail address: sergio.munoz@upm.es (S. Muñoz).

https://doi.org/10.1016/j.ipm.2022.103064
Received 14 October 2021; Received in revised form 2 August 2022; Accepted 12 August 2022
Available online 26 August 2022

performance and are more likely to suffer from severe physical and mental health problems (Moreno Fortes et al., 2020). This translates into a significant strain for organisations through direct and indirect costs such as increased absences, lower productivity, high turnover rates, decreased work engagement, increased staffing, and health benefit costs (Foy et al., 2019).

Considering these high impacts and the associated costs of workplace stress to individuals and society, managing stress has become a high priority health concern for populations around the world (Greene et al., 2016). However, even if 79% of managers are concerned about stress in their organisations, less than 30% of organisations have procedures for dealing with workplace stress (Muñoz & Iglesias, 2021; Parent-Thirion et al., 2012). Early detection and monitoring of stress problems can significantly improve the efficiency of interventions, decrease associated costs, and prevent stress from becoming chronic (Can, Chalabianloo et al., 2019). Nevertheless, the social stigma associated with mental illness (Kim et al., 2021) magnifies the challenge of its early detection.

Traditional methods for early stress detection mainly consisted of self-reports in response to standardised questionnaires (Andreou et al., 2011), such as Perceived Stress Scale (Chan & La Greca, 2013) and Depression Anxiety and Stress Scale (Osman et al., 2012). In recent years, the huge strides in affective computing have opened many possibilities for early stress detection. Affective computing makes use of technological means to recognise the affective state of a person (Picard, 2000). A significant amount of research has recently been conducted on automatic stress measurement systems, which use smart devices and advanced affective computing algorithms to detect stress. The two main directions for stress detection are the analysis of physiological data (e.g., skin conductance, pupil diameter, heart rate) and behavioural data (e.g., mobile phone usage, physical activity, facial expressions, keystroke dynamics) (Alberdi et al., 2016) or a combination thereof.

Whereas these solutions have shown promising results in stress prediction, their implementation in a real-life scenario presents additional challenges. For instance, everyday activities can amend body responses or produce noise that affects the quality of measurements (Han et al., 2020). This could lead to sensor malfunction or data problems such as data occlusion or corruption, thus hampering the task of stress recognition. Also, to enable automatic stress detection, a large number of sensors of different kinds and sizes are required that must be conveniently positioned for the users (Maxhuni et al., 2021). This often entails high costs and complexity.

To benefit from the contagious nature of stress and its proneness to be influenced by surroundings (Dimitroff et al., 2017) can help to address this challenge. Our article aims at exploiting this approach by investigating the use of surrounding stress data to predict workers' stress levels. We seek to reduce the complexity of automatic stress detection by using past stress levels from individuals and their closest colleagues to predict current or future levels. Therefore, we focus on the following research questions (RQ):

- **RQ1**: Can the surrounding stress information be used to predict the stress levels of an individual?
- **RQ2**: How do the individual's stress history and the stress levels of their colleagues impact the predictive performance?
- **RQ3**: Which features yield the best performance when predicting an individual's stress level from surrounding stress data?

Motivated by these questions, we propose a machine learning method able to predict workers' stress from their previous stress levels and the levels of their closest colleagues. In order to validate our model, we gathered data containing psychological self-assessment related information (acquired from standardised, validated questionnaires) and unobtrusive sensor information collected from smartphones during a real-life experiment conducted with 30 employees over a monitoring period of 8 weeks. Additionally, we externally validate our method using an independent dataset, the StudentLife (Wang et al., 2014), which is a public dataset that contains real-life behavioural data from students during an entire university course.

The experiments show that it is possible to use surrounding stress data for predicting stress, and the proposed method achieves an F-score of 81% to classify stress into three levels: low, medium, and high. Furthermore, we carried out a statistical study on the results of our method to analyse the performance of different classifiers and models. This is the first study on supervised stress recognition using surrounding stress data acquired from previous data from workers and their close colleagues to the best of our knowledge. Our findings may have important implications for enhancing stress recognition systems. The proposed approach could be combined with those exploited in other works (such as solutions based on physiological and behavioural data). This could help to reduce the number of sensors and data required and to improve the effectiveness of stress detection methods in the presence of scarce data.

The rest of the paper is organised as follows. An overview of stress theories and stress detection methods is presented in Section 2. In Section 3, the proposed stress prediction models are described. Following, Section 4 provides information on the experiments carried out to collect our dataset and gives a detailed description of both datasets. Later, in Section 5 we depict the experimental setup that we use to evaluate the proposed models, and in Section 6 we present the results obtained. We discuss our findings and draw the main conclusions in Section 7. Finally, Section 8 concludes with the main findings of the investigation and the outline of possible future directions of this work.

2. Literature review

The importance of stress in personal and professional life has increased interest and research on its nature and prevention. However, due to the different contexts in which the notion of stress is used and its subjectivity, none of the stress definitions has been universally recognised (O'Connor et al., 2021). One of the first and more generic definitions of stress was proposed by Hans Selye, who argued that stress is the generic reaction of the body to any demand (Selye, 1956). In recent years, many extended and more specific stress definitions have been proposed (Burman & Goswami, 2018). Kim and Diamond (2002) propose a three-component definition of stress: it requires heightened excitability or arousal, an experience that must be perceived as aversive, and lack of control. In accordance with the work proposed by Cox and Griffiths (1995) and Cox and Griffiths (2015), the definition of stress can be approached from three different points of view: physiological, psychological, and engineering. Regarding the physiological approach, stress refers to the changes that occur in a human under pressure. The psychological one states that stress is the dynamic process arising from the interaction between an individual and the environment. Lastly, from an engineering point of view, stress can be considered a stimulus of the environment in the form of a level of demand. Therefore, stress can be non-formally defined as the reaction of the human body to any demanding or hazardous situation (Can, Arnrich et al., 2019). Focusing on the workplace environment, work-related stress can be considered as a specific form of stress that has been provoked or exacerbated by specific aspects of work, work environments, or workplaces (Mishra et al., 2011). These aspects can be related to work conditions, organisational role, career development, work relationships, or environment and organisational structure (Universari & Harsono, 2021).

The context-dependence and subjectivity of stress have also led to a wide variety of stress theories (Dewe et al., 2012). One of the earlier and more transcendental theories is the Person–Environment (P–E) fit theory, which has been the source for other approaches to stress and well-being (French et al., 1982). This theory, founded on the work of Lewin (1936) and Murray (1938), argues that stress arises from the fit or congruence between the person and the environment. According to this theory, stress is a lack of harmonisation between a person's abilities and the claims placed on them. Another important theory in this field is the Transactional Model of Stress, proposed by Lazarus and Holroyd in 1982 (Holroyd & Lazarus, 1982). According to this theory, stress arises as a relationship between the person and the environment. The theory argues that the person appraises the environment as tough or demanding, therefore threatening well-being (Glanz et al., 2008).

These models have been the basis for understanding stress and have helped develop prevention, detection, and regulation methods. Over the last few years, much progress has been made in the research of automatic stress measurement systems to enable its early detection and avoid its negative health and economic-related consequences (Alberdi et al., 2018). Traditional approaches to stress detection consist of psychological evaluation through self-report questionnaires or psychologist interviews (Hayashi et al., 2012). To date, research in the field has evolved in two main directions: the analysis of physiological data and the analysis of behavioural data (Alberdi et al., 2016). Physiological data can provide objective information on the stress levels of an individual, and a wide variety of physiological signals have been studied (Singh et al., 2013). Between them, Electro-Dermal Activity (EDA) (Pakarinen et al., 2019), Heart Rate Variability (Castaldo et al., 2019), and Electroencephalogram (EEG) (Jebelli, Khalili et al., 2019) have yielded the most successful results. EDA was used by Pakarinen et al. (2019), showing promising results for long-term assessment of self-perceived stress and arousal during work. The use of HRV was analysed by Castaldo et al. (2019), who demonstrated the reliability and accuracy of HRV features to automatically detect mental stress. Jebelli, Khalili et al. (2019) proposed an EEG-based stress recognition framework by applying different supervised learning algorithms to identify the pattern of workers' brain waves while exposed to different stressors. Some other examples of signals include blood pressure (BP) (Gordon & Mendes, 2021), respiration (Sadat-Mohammadi et al., 2021), blood volume pulse (BVP) (Ladakis & Chouvarda, 2021), eye gaze and blinking (Wang et al., 2019), or pupil diameter (PD) (Pedrotti et al., 2014). Finally, some works propose the use of wearable sensors to predict stress from a combination of several signals (Jebelli, Choi et al., 2019).

On the other hand, analysis of behavioural data exploits variations in individuals' behaviour to predict stress (Sharma & Gedeon, 2012). These methods comprise the analysis of computer patterns such as mouse or keystroke dynamics (Dacunhasilva et al., 2021), text (Muñoz & Iglesias, 2022), activity (Giakoumis et al., 2012), facial expressions (Zhang et al., 2019), speech (Tomba et al., 2018), or the use of mobile phones (Ferdous et al., 2015). In the current literature, there are extensive reviews on the usage of these techniques to detect stress (Alberdi et al., 2016; Can, Arnrich et al., 2019; Greene et al., 2016).

Whereas the reliability of these methods for detecting stress has been proven, their implementation in real scenarios poses additional challenges. For instance, everyday activities can deteriorate the signal quality or alter body responses (Han et al., 2020), thus hindering stress recognition. Therefore, errors due to incorrect placement, movements, or detached equipment are common in daily life and lead to corrupted data (Can, Arnrich et al., 2019). Also, some solutions require massive data or the installation of obtrusive and expensive sensors (Novais & Carneiro, 2016). This diminishes the feasibility of these solutions in real-world settings. Our approach aims at addressing this challenge by introducing a method that uses the surrounding stress data to predict stress. We will explore how the use of stress-related data coming from past measures and close colleagues can help enhance the performance of stress recognition in an unobtrusive and low-cost way. This may help decrease the number of sensors and data needed, given that the data measured in specific individuals can be used for predicting the stress level of other individuals. Furthermore, it can also be used to predict the stress level of an individual at a specific point in time where data corruption or loss has occurred.

3. Stress prediction models

This work introduces a model that exploits surrounding stress information, that is, the stress levels of physically close colleagues and the individual's stress history, to predict current stress levels. For this purpose, a machine learning system is presented. Fig. 1 shows a representation of the proposed model. As can be seen, the surrounding stress information is processed by two different processing modules. These modules are responsible for receiving as input the previous stress levels from the subject and the closest colleagues, generating a feature vector representing this information, and producing this vector as output. The feature vectors are then concatenated and fed to a machine learning classifier, which yields a prediction based on the given information. The proposed models have been validated using the two datasets described in Section 4.

In the literature, affective states such as stress (Dimitroff et al., 2017) or emotions (Petitta et al., 2021) emanated from the individual have been discussed to influence the mood of close individuals. Furthermore, data related to past emotions can play a

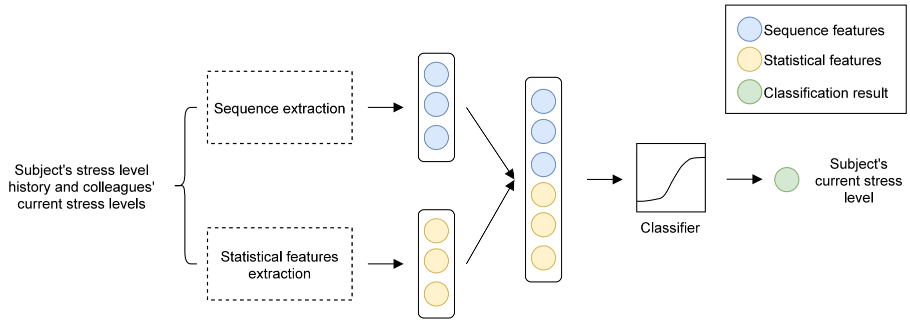


Fig. 1. General architecture representation of the proposed method.

role in future moods (Hollis et al., 2017). Nonetheless, to the best of our knowledge, the effect of surrounding stress has not been utterly analysed in the domain of stress level prediction. Therefore, this work proposes the use of close individuals' stress data to extract surrounding stress-related features. With this in mind, our objective is to analyse the extent to which this type of information is valuable for stress prediction.

We consider the surrounding stress data to be a compound of two factors: the personal and social components. The former refers to the past stress levels of the individual in question. The latter refers to the stress levels measured for the closest colleagues of this individual. We define closest colleagues as those in physical proximity of the individual, as inferred from smartphone sensor data, including WiFi, cell, and GPS location. Consider an employee whose personal stress levels during a certain period T are the following:

$$E(T) = [e_{t-w}, \dots, e_t] \tag{1}$$

where e_t is a number that expresses the stress level of the employee at the time t, on a scale of 1 to 3, thus $e_t \in [1,2,3]$; and w is the window size, that is, the number of previous stress measures considered. Considering that we want to predict the stress level of the individual at the moment t, that is, e_t, we can define the personal component as:

$$P(T) = [p_{t-w}, \dots, p_{t-1}] \tag{2}$$

In the personal component of the employee, we are predicting the level of stress p_t. Thus this measure is excluded from the input.

The social component is composed of the stress levels of colleagues in close physical proximity. Thus if we define the stress levels of a certain colleague i as $C^i(T) = [c^i_{t-w}, \dots, c^i_t]$, being $c^i_t \in [1,2,3]$, the social stress component is defined as the concatenation of all vectors:

$$S(T) = \bigoplus_{i=1}^{n} C^i \tag{3}$$

where n is the total number of close colleagues to consider ($n \in \mathbb{N}$). Based on these components, we propose different ways to combine the information in order to improve the performance of the stress prediction. In this way, three different models are proposed: (i) using only sequential data (M_{SEQ}); (ii) using only statistical features (M_{SF}); and (iii) using a feature ensemble of sequential and statistical data (M_{FE}).

3.1. Sequential data model (M_{SEQ})

The first model aims to predict stress using only sequential data of the surrounding stress. Sequential data consist of an individual's stress levels ordered in a timely manner for a certain period. Given the surrounding stress data components defined above, we can define the sequential feature vectors as:

$$P_{SEQ}(T) = [p_{t-w}, \dots, p_{t-1}] \tag{4}$$

$$S_{SEQ}(T) = [c^i_{t-w}, \dots, c^n_{t-w}, \dots, c^i_t, \dots, c^n_t] \tag{5}$$

We can see that these vectors contain the last w levels measured for a specific employee and the top-n closest colleagues. Note that while the measure at the moment t is not included in the personal component (since it is the value to be predicted), it is included in the social component of colleagues. This approach evaluates the prediction of an individual's current stress level given the current stress levels of colleagues.

Table 1
Statistical features extracted from the data.

Feature	Formula	Description
avg	$E_{avg} = \frac{1}{w}\sum_{i=t}^{t-w} e_i$	Average of all levels
std	$E_{std} = \sqrt{\frac{1}{w-1}\sum_{i=t}^{t-w}(e_i - \overline{e})^2}$	Standard deviation of all levels
max	$E_{max} = \max E(T)$	Maximum value between levels
min	$E_{min} = \min E(T)$	Minimum value between levels
last	$E_{last} = e_t$	Last value of all levels
Δ	$E_{\Delta} = e_t - e_{t-w}$	Total level increment
δ	$E_{\delta} = e_t - e_{t-1}$	Last level increment

3.2. Statistical features model (M_{SF})

This second model aims to exploit several statistical features contained in the surrounding stress data for the purpose of improving the classification performance. The extracted features are shown in Table 1.

Note that all formulas are generalised for a generic employee, so the last measure of the period (e_t) is included. However, when calculating these features for the personal component, the last measure of the period is not included, as it is the measure to be predicted. Given this, we can compute the statistical feature vector of an individual, that is, the personal component as:

$$P_{SF} = [P_{mean}, P_{std}, P_{max}, P_{min}, P_{last}, P_{\Delta}, P_{\delta}] \tag{6}$$

In the personal component, all these features are extracted at the subject level (taking into account only one individual). However, for the social component, we can compute additional features related to the group of close colleagues, that is, statistical features related to the measures of all the colleagues considered. In this way, we can define the social component of the statistical features model as the concatenation of the features related to the group with the statistical features vectors of all colleagues:

$$S_{SF} = G_{SF} \oplus \bigoplus_{i=0}^{n} C^i_{SF} \tag{7}$$

where C^i_{SF} is the statistical feature vector of the colleague i, n is the number of close colleagues considered, and G_{SF} is the feature vector computed from the data of all the colleagues considered. This vector contains the features described in Table 1 computed at the group level rather than at the individual level. Additionally, the average, standard deviation, maximum, and minimum values of all colleagues have been calculated for the increments and last values.

3.3. Features ensemble model (M_{FE})

The primary purpose of this model is to merge the two kinds of features into a unified feature set and hence to benefit from the combination of the different information types provided by these features. In this manner, a machine learning classifier may achieve better performance scores learning from the merged set than learning only from a feature subset. The surrounding stress components in this model are computed as the concatenation of the feature vectors of the previous models:

$$P_{FE} = P_{SEQ} \oplus P_{SF} \tag{8}$$

$$S_{FE} = S_{SEQ} \oplus S_{SF} \tag{9}$$

All proposed models have been validated using data from two datasets: StudentLife and our own collected dataset described in Section 4. Finally, to explore the influence of the personal and social stress components of each model, this work proposes three different ways to compute the surrounding stress data: (i) using only personal data; (ii) using only social data; and (iii) using both personal and social data. The evaluation of each model, along with the different components of the surrounding stress, is presented in Section 5.

4. Datasets

4.1. Our dataset

An experiment was conducted in order to collect data from a group of 30 employees from two different organisations for 8 weeks. Participants were voluntarily recruited, and the Institutional Ethics Review Board approved all experimental procedures. A presentation was made that described the objectives and methods of the study, with the participation of nearly double the final participants. Later, interested participants received a smartphone with our data collection app installed and configured. The data collection framework was based on a server-client architecture built around the Samsung Galaxy S3 Mini 32 GB smartphone. No additional clinical screening has been performed, except the annual health screening performed in each organisation. The main reason some employees refused to participate in the study was that they did not want to use another smartphone. The 30 participants selected for the study were workers from two different companies in Trento, Italy. Table 2 offers a summary of the demographic

Table 2
Demographics of the participants in the performed study.

Variable	Characteristics	No. (%)
Gender	Male	18 (60%)
	Female	12 (40%)
Education	High-school	9 (30%)
	Bachelor degree	11 (36.67%)
	Graduate degree	10 (33.33%)
Age	26–30	5 (16.67%)
	31–40	18 (60%)
	>40	7 (23.33%)
	Mean	37.46
Marital status	Married	15 (50%)
	Never married	15 (50%)
No. of children	None	17 (56.67%)
	1–2	10 (33.33%)
	3–4	3 (10%)

characteristics of the employees. It can be observed that there is a pretty balanced mix of age, gender, marital status, education level, and number of children among the participants.

During the study, the workers used the provided smartphone daily as their phone. No restrictions on the handling of their smartphone were given to the participants to guarantee the most realistic conditions possible for our analysis. The application responsible for collecting data was automatically started at 9 am on business days (Monday–Friday) without any interaction from the user and then continued to run uninterruptedly in the background. Two types of variables were extracted from this experiment: objective variables (employees' behaviour captured by sensors during work hours); and subjective variables (responses obtained from questionnaires). With the purpose of collecting users' mood and stress levels, the app automatically prompted users to complete a questionnaire at three different times of the day: at 9 am (at the beginning of the working day), at 2 pm (after lunch break) and 5 pm (at the end of the working day). The user had the option of answering the questions at these times or postponing the questionnaire to another moment. The questionnaires contained 14 questions related to stress, sleep quality, work abandonment, energy levels, and the affect of mood states. The questions had an estimated response time of one minute and were founded on two validated questionnaires: Profile of Mood States (POMS) (Shacham, 1983) and Oldenburg Burnout Inventory (Demerouti & Bakker, 2008).

Each question had five possible answers corresponding to five stress-related aspects on a scale ranging from 1 (definitely agree) to 5 (definitely disagree). The first part of the questionnaire is intended to collect information about the occupational health outcomes of the participants: (i) job-induced stress, (ii) job control, (iii) job demand, and (iv) energy perceived during working days. The second part consists of different questions to measure mood: feelings of sadness, friendliness, anxiety, anger, cheerfulness, time pressure, job-related tension, and sleep quality.

The number of completed questionnaires was 1455, which denotes a response rate of 79.97%. Some of the most relevant insights drawn from the questionnaire responses are: throughout the entire monitoring period, the employees perceived a moderate (35.15%) to high (22.18%) stress level; at some time, almost all of them (29 out of 30) reported that their job tasks and responsibilities were highly demanding (50.58%); at some point, 19 workers felt High-Tense, 18 employees felt High-Anxious, and 11 of the respondents reported High-Angry (5.67%); and finally, 24 respondents reported Poor Sleep Quality as a reaction to stress.

Furthermore, the locations of the subjects were analysed, focusing on understanding the frequent changes in location throughout the workday. In this way, the app retrieved: (i) the record of WiFi networks available with their corresponding Basic Service Set Identifier (BSSID) address, (ii) cell tower locations, and (iii) Google Maps locations information (latitude, longitude). Google Maps locations where the subjects stayed for more than 15 min were clustered with a maximum diameter of 300 m (using the Haversine distance equation (Robusto, 1957)). Also, the number of locations on each day was computed. For cell tower information and WiFi networks, location information was clustered on an hourly basis. Thus, the locations were compared every hour, increasing the count when different clusters appeared with respect to the previous hour. In this work, we used only location-related data along with self-reported questionnaire items related to stress, as our objective is to explore the influence of surrounding stress among colleagues who work close to each other.

4.2. StudentLife dataset

In addition to the dataset presented in Section 4.1, another dataset has been used to explore the performance of the presented method in a different population. The StudentLife dataset is a broad longitudinal dataset containing passive and automatic sensing data from smartphones of 48 Dartmouth College students for 10 weeks, with the purpose of assessing their mental health (Wang et al., 2014). It contains more than 53 GB of continuous data, 32 000 self-reports, and pre- and post-surveys. Amid the 48 students who completed the study, 30 were undergraduates, and 18 were graduate students. Regarding gender, 38 participants were male, and 10 participants were female.

Participants were asked to answer different questions related to stress, mood, or current events during the collection phase as they used their smartphones. On average, 3–13 questions were administered per day. The students responded to several scheduled questionnaires, including stress, mood, social interaction, duration of sleep, physical activity, and a short personality item. A total of 35 295 completed questionnaires were collected.

Besides, the locations of all access points on the network and WiFi scan logs were collected as part of the study. This information includes all encountered BSSIDs and their signal strength values that were used to determine the location of a student. Among all the data provided in the dataset, in this work, we used only location-related data along with self-reported questionnaire items related to stress. In this sense, we consider the physical proximity of students as surrounding stress and explore its effect on stress prediction.

4.3. Data preprocessing

The own collected and StudentLife datasets provide a large set of diverse passively detected data. However, for our work, only some of these data are required. Therefore, only the stress levels obtained from the questionnaires and location-related data have been selected among all data provided.

Once the required data are selected, a homogenisation of the stress levels extracted from the questionnaires is performed. Given the different nature of the surveys used in each experiment, the stress scales are different in the two datasets. For example, while in our dataset, a stress value of 5 refers to the highest stress level, in the StudentLife dataset, it refers to the lowest stress level. To solve this issue, the stress levels obtained from the StudentLife dataset have been converted to the same scale used in our dataset. Furthermore, as described above, the responses to the stress questionnaires in both datasets range between 1 and 5. Self-reported stress is inherently highly subjective and, as such, is prone to significant inter-subject differences. Therefore, we have categorised the stress levels to smooth out the differences between subjects. In this way, these levels have been split into three different regions: "low", score < 3; "moderate", score $= 3$; and "high", score > 3.

In both datasets, each stress or location measure is stored along with the timestamp of the moment it was registered. Moreover, the two datasets contain several daily measures over several weeks. This gives us two possible ways to pose the problem of stress classification over time: to perform the classification at the day level or the week level. The former refers to the prediction of the stress level at a specific interval of the day, given the data from the previous intervals on that day. The latter refers to predicting the stress level on a specific day, given the data from the previous days in that week. Given this, we performed a different transformation in the datasets for each case. For prediction at the day level, we have grouped the data by day and defined three different intervals for each day. Taking into account the different times of the questionnaires of the datasets, these intervals are as follows: from 7 am to 11 am (morning), from 11 am to 3 pm (noon), and from 3 pm to 7 pm (evening) for the own collected dataset; and from 9 am to 6 pm (morning), from 6 pm to 12 am (evening), and from 12 am to 9 am (night) for the StudentLife dataset. For the prediction at the week level, we have grouped the data by week, and then for each day of the week, we have calculated the average stress value between all the measures on that day. Also, only working days are considered in both cases, so the datasets have been filtered to include only these days.

While there are several ways to detect proximity between colleagues, based on our previous experience (Carreras et al., 2012; Osmani et al., 2014), we opted for the use of location data, including WiFi, Cell ID, and location service. The proximity between colleagues is computed on the basis of the time spent in the same room during the considered interval. Specifically, we have used the Jaccard similarity to calculate the intersection between the sets containing the locations of each employee during a certain day/week. Jaccard similarity is defined as the relation between the size of the intersection and the size of the union of two sets. The closest colleagues of an employee are those who present a higher Jaccard index during a certain period. In this way, we can obtain the stress level of each worker at each interval, along with the stress levels of the closest colleagues. The final data format used for the proposed method has one row for each interval/day of each day/week (depending on the scenario considered), along with the stress level of the individual and the n-closest colleagues.

Finally, it is worth mentioning that the questionnaires of both datasets are based on clinically validated, ecological momentary assessments (EMAs) (Shiffman et al., 2008). These assessments address the validity challenge of standard retrospective questionnaires regarding the influence of subjects' prior experience. Standard retrospective questionnaires can suffer from issues that affect recall when people attempt to recollect or summarise past experiences or feelings. EMAs tackle this challenge by repeatedly sampling subjects' experiences and feelings in real time and in natural environments. Thus, EMAs are intended to minimise recall bias and maximise ecological validity. This makes them suitable for the study of behaviour or mental state in real-world contexts, such as the one considered in our work. The experiment we conducted for data collection and the one conducted in the StudentLife study are based on these assessments and responses are taken multiple times a day making them a robust approach to potential recall bias.

5. Methods

The presented models have been evaluated through several stress prediction tasks, where the aim is to predict an individual's stress level. These models can be implemented using different resources and methods. An experimental study has been designed to thoroughly evaluate each model's effectiveness in multiclass stress prediction and identify the optimum resources and methods. The proposed methods have been validated using the datasets described in the previous section. The methodology followed in this validation is described in this section, and the obtained results of the experiments are shown in Section 6.

We postulate the task of predicting perceived stress as a classification problem. In this way, the class to predict is the self-reported stress level (low, moderate, high), while the attributes correspond to the surrounding stress data. With the purpose of analysing the performance of the models, three different machine learning classifiers have been used (logistic regression, decision tree, and Adaboost). Besides, it is interesting to evaluate the performance of a more complex neural network that takes into account sequential

Table 3
Hyperparameters that obtain the best performance for each kind of data.

Dataset	Level	Personal	Social	Personal+Social
Our dataset	Day	$w = 3$	$w = 2;\ n = 3$	$w = 3;\ n = 2$
	Week	$w = 4$	$w = 2;\ n = 3$	$w = 5; n = 4$
StudentLife	Day	$w = 2$	$w = 1;\ n = 3$	–
	Week	$w = 3$	$w = 2;\ n = 3$	$w = 4;\ n = 1$

data. To this end, we have developed a sequential model with a Bidirectional LSTM layer followed by a Dense layer with "sigmoid" activation. All experiments have been carried out using 10-fold cross-validation and the weighted average of the F1-Score as the performance metric. The F1-score is a measure of a model accuracy defined as the harmonic mean of the model precision and recall:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{10}$$

The research questions described in Section 1 have shaped the experimental study. The main objective of the experiment is to provide insight into whether surrounding stress information can be used to predict stress (**RQ1**). Furthermore, to analyse the influence of each component of the surrounding stress (**RQ2**), we propose three different scenarios for the experiment: (i) the use of only personal data; (ii) the use of only social data; and (iii) the combination of personal and social data. Finally, we test the three models separately in Section 5 to investigate the impact of features on the prediction performance of stress levels (**RQ3**).

Considering these scenarios, different model hyperparameters must be tuned to optimise the classification performance. The parameters to tune are the window size, w, which represents the number of measures considered for the prediction, and the number of close colleagues to consider, n. As commented in Section 4, the nature of the datasets used provides two different possibilities to predict stress level: prediction at the day level and prediction at the week level. When predicting at the day level, each day has a maximum of three intervals in which stress has been measured, so $w \in [1, 2, 3]$; while when predicting at the week level, $w \in [1, 2, 3, 4, 5]$, since only working days are considered. Also, when predicting at the week level, we only take into account the data for the days of the week in which the prediction is made. That is, each week is used separately, and data from one week are not used to predict the stress levels of the following week. Thus, when the window size is 5, the Friday level is predicted from the data of the rest of the working days within that week ([Mon, Tue, Wed, Thu]). Similarly, when the window size is 4, we can predict the level of Thursday ([Mon, Tue, Wed]) or Friday ([Tue, Wed, Thu]). And so on.

Note that when using personal data, w must always be greater than 1, as one of the individual measures is the one level to be predicted. Regarding the number of close colleagues to consider, we have performed the experiment with up to four colleagues, so $n \in [1, 2, 3, 4]$. To select the optimal value of w and n for each case and each dataset, an exhaustive experiment has been carried out in which the model has been tested using all possible values. The results of this experiment are publicly available online for the interested reader.[1] In summary, Table 3 shows the hyperparameters chosen for each kind of data.

Finally, it is worth mentioning some limitations of the experiment concerning the available data. As expected, our dataset contains missing values as a result of a lack of response to questionnaire prompts. Missing data are common in these types of experiments in real-world settings involving participants monitored longitudinally. Therefore, to ensure the reliability and robustness of the system, only samples where all data are available are used to make predictions. In each scenario, we use the maximum amount of data available for that particular scenario. Thus, for the scenario using only personal data, we need at least two measures of a user: the stress level to be predicted and the previous level. These data would enable the M_{SEQ} model, as it requires only one interval. However, to extract statistical features (M_{SF} and M_{FE} models), at least two measures from previous levels are needed. Similarly, for the scenario using only social data, at least the user's predicted stress level and the current level of one of his or her close colleagues are required. Again, these data would enable the sequential data model (M_{SEQ}), but to enable M_{SF} and M_{FE} at least one measure from two colleagues or two measures from one colleague are needed. Finally, for the scenario combining the two types of data, at least the predicted stress level of the user, his or her previous level, and the current level of one of his or her close colleagues are needed. In this third case, analysing at the daily level for the StudentLife dataset, we find that there are not enough samples containing at least two measurements of a user's stress level and one of his or her close colleagues for the same day. However, there are enough samples with at least two measurements of the user's stress level and enough samples with at least one measurement of his or her stress level, and one from a close colleague. For this reason, when we use the StudentLife dataset at the day level, we have sufficient data for the scenario that uses only the personal component and for the scenario that uses only the social component, but not for the scenario that combines both.

6. Results

Once the hyperparameters have been tuned for the personal and social data, we proceed to evaluate the proposed models. First, we evaluate the models in the first scenario, that is, using only personal data. This experiment allows us to analyse whether the use of the individual's previous data can provide good results for stress prediction. We have completed the experiment using the

[1] https://gsi.upm.es/~smunoz/stress-ambient/.

Table 4
Experiment results for the three defined scenarios. The results show the weighted F-score measure obtained with each model in each dataset with the corresponding classifier.

Component	Classifier	Our dataset			StudentLife		
		M_{SEQ}	M_{SF}	M_{FE}	M_{SEQ}	M_{SF}	M_{FE}
		Day level					
Personal	LogR	55.69	55.01	55.69	34.58		
	DT	58.38	58.38	58.84	**56.22**		
	ADA	58.68	**58.98**	58.01	**56.22**		
Social	LogR	45.77	44.3	47.52	60.48	40.08	60.52
	DT	47.79	48.63	51.31	59.09	49.18	**60.97**
	ADA	37.87	51.67	**53.27**	59.51	46.25	57.15
Personal+Social	LogR	71.49	66.74	70.43			
	DT	52.5	55.62	**73.17**			
	ADA	57.89	59.17	67.11			
		Week level					
Personal	LogR	62.78	66.62	69.85	53.58	52.92	55.25
	DT	72.01	67.09	**72.35**	60.9	**61.54**	61.5
	ADA	63.16	64.35	69.04	57.64	57.06	59.35
Social	LogR	43.71	52	55.8	67.59	71.37	**72.89**
	DT	57.45	55.32	**59.92**	65.01	67.37	70.55
	ADA	49.33	53.71	57.47	61.01	68.16	66.6
Personal+Social	LogR	69.22	64.42	74.6	64	73.46	78.75
	DT	76.48	71.83	**79.16**	78.31	80.57	**81.79**
	ADA	68.13	63.97	67.01	76.3	69.58	72.27

hyperparameters shown in Table 3. The results of the experiment are shown in Table 4. We can see that the classifier is able to predict a future stress level using only previous personal data. The results show that when predicting at the day level, we can predict the stress level with a nearly 59% and 56% F-score for the own collected dataset and the StudentLife dataset correspondingly. When predicting at the week level, the F-score increases up to 72% for our dataset and 61% for the StudentLife dataset. If we analyse the performance of the different proposed models, we can appreciate that, in general, the use of statistical features improves over the use of only sequential data, and the best average results are obtained when combining sequential data with statistical features, that is, with the M_{FE} model.

After analysing the scenario of the personal data, we proceed to analyse the second scenario: the use of only social information. Again, we have completed the experiment using the hyperparameters obtained from the tuning, shown in Table 3. Table 4 shows the results of the experiment. We can see that, in general, the classifiers perform a little worse on this kind of data for our dataset. However, it still achieves results close to 60% of F-score. Although these results may seem low or poor, it is worth remembering that they are obtained using only data external to the individual. Furthermore, for the StudentLife dataset, we can see a significant enhancement, with a performance that surpasses 70% of F-score. When analysing the performance of the different models proposed, we can appreciate that the best results are obtained when combining sequential data with statistical features (M_{FE} model). In this case, the improvement of the M_{FE} model is more evident.

Finally, we conclude our experiments with the third scenario: the combination of personal and social information to predict individual stress. We can see the results of the experiment in Table 4. There were not enough data for predicting at the day level in the StudentLife dataset for this experiment. However, by analysing the results for the own collected dataset, we can appreciate a significant improvement in the classifier's performance that yields an F-score of 73.17%. This entails an increase of almost 15 and 20 percentage points with respect to the use of only personal data and only social data. This enhancement is confirmed when analysing the experiment results at the week level. We can see that the combination of personal and social information results in a significant performance improvement. In terms of F-score, the classifier achieves peaks of 79.16% and 81.79% using our dataset and the StudentLife dataset, respectively. This entails an increase of almost 10 percentage points. These results confirm that it is possible to predict stress using surrounding stress-related data (**RQ1**).

When comparing classifiers, we see that, in general, decision tree is the one that performs better. Whereas these classifiers have shown good performance, they do not take into account sequential data. To study whether the proposed approach could benefit from a more complex neural network that considers this kind of data, we have implemented our approach using LSTM. We have developed a sequential model with a Bidirectional LSTM layer followed by a Dense layer with "sigmoid" activation. The results, along with their comparison with the best performing classifier, are shown in Table 5.

The results indicate that the limited amount of data available hinders the performance of complex neural networks such as LSTM. Among all the experiments, LSTM only yields the best result when using the social component at the week level in our dataset. Analysing how the results vary according to the hyperparameters (window size and the number of colleagues), we can observe that LSTM requires a larger amount of information. To improve performance, this method requires larger window sizes or a higher number of close colleagues compared to other classifiers. This can be easily observed by looking at the hyperparameters chosen for the scenarios using only personal or only social data. However, as the size of the window or the number of colleagues increases, the

Table 5
Performance comparison between the best performing classifier and an LSTM model.

	Dataset	Classifier	Hyperparameters	F-score
Day level				
Personal	Our dataset	LSTM	$w = 3$	54.34
		ADA	$w = 3$	**58.98**
	StudentLife	LSTM	$w = 2$	31.17
		DT	$w = 2$	**56.22**
Social	Our dataset	LSTM	$w = 3; n = 3$	43.97
		ADA	$w = 2; n = 3$	**53.27**
	StudentLife	LSTM	$w = 1; n = 3$	55.32
		DT	$w = 1; n = 3$	**60.97**
Personal+Social	Our dataset	LSTM	$w = 3; n = 1$	52.69
		DT	$w = 3; n = 2$	**73.17**
Week level				
Personal	Our dataset	LSTM	$w = 5$	65.68
		DT	$w = 4$	**72.35**
	StudentLife	LSTM	$w = 4$	43.83
		DT	$w = 3$	**61.54**
Social	Our dataset	LSTM	$w = 2; n = 5$	**63.56**
		DT	$w = 2; n = 3$	59.92
	StudentLife	LSTM	$w = 3; n = 2$	66.33
		DT	$w = 2; n = 3$	**72.89**
Personal+Social	Our dataset	LSTM	$w = 4; n = 4$	59.13
		DT	$w = 5; n = 1$	**79.16**
	StudentLife	LSTM	$w = 3; n = 1$	49.82
		DT	$w = 4; n = 1$	**81.79**

number of samples available in the used datasets decreases. This causes the performance to drop as there are insufficient samples to successfully train the neural network. Thus, in the scenario combining the personal component with the social component, the maximum performance is obtained with a slightly smaller window size compared to other classifiers.

From this analysis, we can observe that LSTM could be an interesting proposal to consider. It would allow the system to take advantage of the temporal information inherent in sequential stress data. Nevertheless, it requires a larger amount of information.

Once the results of the experiment have been analysed, we have performed the Friedman statistical test (Demšar, 2006) in order to further study the impact of the presented models and their performance. This test aims to determine whether we may conclude from the sample of results that there is a difference between the classification methods. As a result, the Friedman test ranks methods according to their performance on different datasets. The lower the ranking of a specific method, the better its performance in comparison to the rest.

The first step in calculating the Friedman test is to convert the actual results into ranks. Let r_i^j be the rank of the jth algorithm in the ith dataset, and k and n the number of methods and datasets, respectively. Friedman's test compares the mean ranks of the methods $R_j = \frac{1}{n}\sum_i r_i^j$ and establishes that the Friedman statistic under the null hypothesis (all algorithms are equal, so their ranks are also equal) is:

$$X_F^2 = \frac{12n}{k(k+1)}\left(\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right) \tag{11}$$

and with $k - 1$ degrees of freedom. However, Iman and Davenport (1980) proposed a better static distributed according to the F-distribution, with $k - 1$ and $(k - 1)(n - 1)$ degrees of freedom:

$$F_F = \frac{(n-1)X_F^2}{n(k-1) - X_F^2} \tag{12}$$

We perform the test with an α value of 0.1. On those averages, $X_F^2 = 24.13$, $F_F = 9.20$, and the critical value $F(k-1,(k-1)(n-1)) = 3.36$. Given that $F_F > F(8, 24)$, the null hypothesis of the Friedman test is rejected. For simplicity, Table 6 shows the five best approaches according to their ranks, as computed by the Friedman test.

As can be seen, Friedman's test points out that the method of combining personal with social information is the best classification model. The three lower ranks are obtained when combining data from both sources, whereas the following two are obtained using only personal data. This confirms that when using only one source of information, personal data outperforms social data (**RQ2**) and that the best results are obtained when combining personal and social information. Also, we see that the combination of statistical features and sequential data surpasses the use of only sequential data or only statistical features (**RQ3**). Best results are obtained when using the feature ensemble model along with the combination of personal and social data. These results demonstrate the effectiveness of using surrounding data for predicting stress (**RQ1**), especially when combining personal and social data.

Table 6
Friedman rank for the top-5 models.

Model	Components	Rank
M_{FE}	Personal+Social	1
M_{SEQ}	Personal+Social	2.5
M_{SF}	Personal+Social	2.75
M_{FE}	Personal	4.75
M_{SF}	Personal	6.25

7. Discussion

Three research questions drove this work as presented in Section 1. First, in RQ1, we investigated whether surrounding stress-related information can be used to predict the stress levels of individuals. In this regard, a novel approach that exploits this information yielded substantial performance, as shown by the experiments, reaching an F-Score of 79% in our dataset and 81% in the external dataset. Therefore, it is reasonable to assume that surrounding stress-related information is highly relevant to stress prediction.

Our second question (**RQ2**) pertained to how personal and social stress information compare in terms of predictive performance. In this sense, the statistical results identify the combination of personal and social data as the best-performing approach. This confirms the effectiveness of the combination of different sources of stress information that these methods perform, resulting in a notable improvement compared to the other methods. If we compare with the use of only one kind of data, analysing the results, we can conclude that the use of personal data achieves better performance than the use of social data.

The last question (**RQ3**) was concerned with investigating methods of feature extraction that can improve the stress prediction performance. In this respect, the results have shown that the combination of sequential and statistical features is the best-performing method. This model yields the best results for all the scenarios.

Based on our results, we can conclude that the use of surrounding stress-related information can yield substantial performance in stress prediction. Our findings are in line with current research on affective state contagion occurring during interpersonal social interactions, which states the significant influence of closest people on the individual's affective state (Dimitroff et al., 2017; Engert et al., 2019; Jia & Cheng, 2021; Petitta et al., 2021). Furthermore, our results are also consistent with current literature in terms of how historical affective-related data can play a role in future moods (Goodday & Friend, 2019; Hollis et al., 2017). However, as far as we know, this is the first study investigating the use of surrounding stress-related data to predict future stress levels.

Our findings have important implications in stress prediction approaches. Taking into account personal and social-related stress when attempting to predict an individual's stress level could enhance the system's prediction performance. This kind of information can be combined with approaches used in other works (such as physiological and behavioural data) to reduce the quantity of data needed (and consequently the number of sensors) and improve the effectiveness of stress detection methods in the presence of scarce data. Finally, one of the main findings of this work is that stress can be contagious and stress of an individual can affect close colleagues. Therefore, this is an indication that in addressing well-being of the workforce a holistic approach should be taken in tackling occupational stress rather than focusing on individuals only.

8. Conclusions

Workplace stress is a concern since it negatively affects employees' health and organisational performance, reducing workers' well-being and decreasing productivity. However, appraising stress is a complicated issue that entails high costs and complexity, particularly when relying on non-obtrusive approaches. In this work, we presented a method for predicting stress using surrounding stress-related data, that is, using previous levels from the individuals and their close colleagues. We considered two components of surrounding stress data: the personal and social components.

Besides, an extensive analysis based on two different datasets with real-life behavioural data of office workers and students has been carried out, achieving an F-score near 80% for stress level prediction. To the best of our knowledge, this is the first study into supervised stress recognition using surrounding stress data acquired from previous data from workers and their closest colleagues. Our results confirm that it is possible to predict perceived stress at work using surrounding stress information.

We believe our findings on the effect of surrounding stress on workers open new research lines to improve future monitoring systems that may enable a better understanding of work-related stress, the impact on occupational health, and the management of human resources. Besides, we consider that this work presents some challenges that can be addressed by future work. One of these challenges is to extend the dataset with new experiments in real environments. A larger amount of data may allow the system to benefit from the advantages of more complex neural network models, which are better suited for dealing with sequential data. Also, it may enable long-term prediction using longer sequences, allowing evaluation of the impact of sequence length on prediction performance. In this line, weekend stress levels may also be an interesting challenge to address. This would not only allow predictions to be made between different weeks but also help to understand how certain parameters relating to weekends (e.g., rest, activities, leisure) can help to reduce work-related stress. Finally, remote working practices are becoming common nowadays, and they are a new interesting area of research for studying occupational stress.

CRediT authorship contribution statement

Sergio Muñoz: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Carlos Á. Iglesias:** Conceptualization, Methodology, Funding acquisition, Supervision, Resources, Writing – review & editing. **Oscar Mayora:** Investigation, Data curation, Resources, Writing – review & editing. **Venet Osmani:** Investigation, Data curation, Supervision, Resources, Writing – review & editing.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work is supported by the Spanish Ministry of Science and Innovation under the R&D project COGNOS (PID2019-105484RB-I00). In addition, Universidad Politécnica de Madrid has granted researcher Sergio Muñoz a mobility grant for performing his research stay at Fondazione Bruno Kessler (Trento, Italy), making this work possible. The authors would like to thank the participants of our dataset and those of the StudentLife study, and for making the StudentLife dataset available to the research community.

References

Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, *59*, 49–75.
Alberdi, A., Aztiria, A., Basarab, A., & Cook, D. J. (2018). Using smart offices to predict occupational stress. *International Journal of Industrial Ergonomics*, *67*, 13–26.
Andreou, E., Alexopoulos, E. C., Lionis, C., Varvogli, L., Gnardellis, C., Chrousos, G. P., & Darviri, C. (2011). Perceived stress scale: reliability and validity study in Greece. *International Journal of Environmental Research and Public Health*, *8*(8), 3287–3298.
Association, A. P. (2019). Stress in america: Stress and current events. Stress in America survey.
Burman, R., & Goswami, T. G. (2018). A systematic literature review of work stress. *International Journal of Management Studies*, *5*(3–9), 112–132.
Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of Biomedical Informatics*, *92*, Article 103139.
Can, Y. S., Chalabianloo, N., Ekiz, D., & Ersoy, C. (2019). Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors*, *19*(8), 1849.
Carreras, I., Matic, A., Saar, P., & Osmani, V. (2012). Comm2sense: Detecting proximity through smartphones. In *2012 IEEE international conference on pervasive computing and communications workshops* (pp. 253–258). IEEE.
Castaldo, R., Montesinos, L., Melillo, P., James, C., & Pecchia, L. (2019). Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life. *BMC Medical Informatics and Decision Making*, *19*(1), 1–13.
Chan, S. F., & La Greca, A. M. (2013). Perceived stress scale (PSS). In *Encyclopedia of behavioral medicine* (pp. 1454–1455). New York, NY: Springer New York, (Chapter 1).
Cox, T., & Griffiths, A. (1995). Work-related stress: nature and assessment. In *IEE colloquium on stress and mistake-making in the operational workplace* (pp. 1/1–1/4).
Cox, T., & Griffiths, A. (2015). Assessment of potentially stressful work and organisational environments: Theory and method. In *Evaluation of human work* (fourth ed.). (pp. 565–582). CRC Press.
Dacunhasilva, D. R., Wang, Z., & Gutierrez-Osuna, R. (2021). Towards participant-independent stress detection using instrumented peripherals. *IEEE Transactions on Affective Computing*.
Demerouti, E., & Bakker, A. B. (2008). The Oldenburg Burnout inventory: A good alternative to measure burnout and engagement. *Handbook of Stress and Burnout in Health Care*, 65–78.
Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(Jan), 1–30.
Dewe, P. J., O'Driscoll, M. P., & Cooper, C. L. (2012). Theories of psychological stress at work. In *Handbook of Occupational Health and Wellness* (pp. 23–38). Boston, MA: Springer US, (Chapter 1).
Dimitroff, S. J., Kardan, O., Necka, E. A., Decety, J., Berman, M. G., & Norman, G. J. (2017). Physiological dynamics of stress contagion. *Scientific Reports*, *7*(1), 1–8.
Engert, V., Linz, R., & Grant, J. A. (2019). Embodied stress: The physiological resonance of psychosocial stress. *Psychoneuroendocrinology*, *105*, 138–146, Festschrift for Dirk Hellhammer.
Ferdous, R., Osmani, V., & Mayora, O. (2015). Smartphone app usage as a predictor of perceived stress levels at workplace. In *2015 9th international conference on pervasive computing technologies for healthcare (PervasiveHealth)* (pp. 225–228). IEEE.
Foy, T., Dwyer, R. J., Nafarrete, R., Hammoud, M. S. S., & Rockett, P. (2019). Managing job performance, social support and work-life conflict to reduce workplace stress. *International Journal of Productivity and Performance Management*, *68*(6), 1018–1041.
French, J. R., Caplan, R. D., & Van Harrison, R. (1982). *The mechanisms of job stress and strain, Vol. 7*. Chichester: J. Wiley.
Gallup (2021). *State of the global workplace 2021 report*. Gallup, URL: https://www.gallup.com/workplace/349484/state-of-the-global-workplace.aspx.
Giakoumis, D., Drosou, A., Cipresso, P., Tzovaras, D., Hassapis, G., Gaggioli, A., & Riva, G. (2012). Using activity-related behavioural features towards more effective automatic stress detection. *PLoS One*, *7*(9), 1–16.
Glanz, K., Rimer, B. K., & Viswanath, K. (2008). *Health behavior and health education: theory, research, and practice*. John Wiley & Sons.
Goodday, S. M., & Friend, S. (2019). Unlocking stress and forecasting its consequences with digital technology. *NPJ Digital Medicine*, *2*(1), 1–5.
Gordon, A. M., & Mendes, W. B. (2021). A large-scale study of stress, emotions, and blood pressure in daily life using a digital platform. *Proceedings of the National Academy of Sciences*, *118*(31).
Greene, S., Thapliyal, H., & Caban-Holt, A. (2016). A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine*, *5*(4), 44–56.
Han, H. J., Labbaf, S., Borelli, J. L., Dutt, N., & Rahmani, A. M. (2020). Objective stress monitoring based on wearable sensors in everyday settings. *Journal of Medical Engineering & Technology*, *44*(4), 177–189.
Hayashi, T., Mizuno-Matsumoto, Y., Okamoto, E., Kato, M., & Murata, T. (2012). An fMRI study of brain processing related to stress states. In *World automation congress 2012* (pp. 1–6). IEEE.

Hollis, V., Konrad, A., Springer, A., Antoun, M., Antoun, C., Martin, R., & Whittaker, S. (2017). What does all this data mean for my future mood? Actionable analytics and targeted reflection for emotional well-being. *Human–Computer Interaction*, *32*(5–6), 208–267.
Holroyd, K. A., & Lazarus, R. S. (1982). Stress, coping and somatic adaptation. *Handbook of Stress: Theoretical and Clinical Aspects*, 21–35.
Iman, R. L., & Davenport, J. M. (1980). Approximations of the critical region of the Friedman statistic. *Communications in Statistics. Theory and Methods*, *9*(6), 571–595.
Jebelli, H., Choi, B., & Lee, S. (2019). Application of wearable biosensors to construction sites. I: Assessing workers' stress. *Journal of Construction Engineering and Management*, *145*(12), Article 04019079.
Jebelli, H., Khalili, M. M., & Lee, S. (2019). Mobile EEG-based workers' stress recognition by applying deep neural network. In *Advances in informatics and computing in civil and construction engineering* (pp. 173–180). Springer.
Jia, M., & Cheng, J. (2021). Emotional experiences in the workplace: Biological sex, supervisor nonverbal behaviors, and subordinate susceptibility to emotional contagion. *Psychological Reports*, *124*(4), 1687–1714.
Kim, J. J., & Diamond, D. M. (2002). The stressed hippocampus, synaptic plasticity and lost memories. *Nature Reviews Neuroscience*, *3*(6), 453–462.
Kim, S.-W., Kim, J.-K., Jhon, M., Lee, H.-J., Kim, H., Kim, J.-W., Lee, J.-Y., Kim, J.-M., & Shin, I.-S. (2021). Mindlink: A stigma-free youth-friendly community-based early-intervention centre in Korea. *Early Intervention in Psychiatry*, *15*(5), 1389–1394.
Ladakis, I., & Chouvarda, I. (2021). Overview of biosignal analysis methods for the assessment of stress. *Emerging Science Journal*, *5*(2), 233–244.
Le, H., Newman, A., Menzies, J., Zheng, C., & Fermelis, J. (2020). Work–life balance in Asia: A systematic review. *Human Resource Management Review*, *30*(4), Article 100766.
Lewin, K. (1936). A dynamic theory of personality. *Journal of Heredity*, *27*(11), 441–442.
Maxhuni, A., Hernandez-Leal, P., Morales, E. F., Sucar, L. E., Osmani, V., & Mayora, O. (2021). Unobtrusive stress assessment using smartphones. *IEEE Transactions on Mobile Computing*, *20*(6), 2313–2325.
Mishra, B., Mehta, S., Sinha, N. D., Shukla, S. K., Ahmed, N., & Kawatra, A. (2011). Evaluation of work place stress in health university workers: a study from rural India. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine*, *36*(1), 39.
Moreno Fortes, A., Tian, L., & Huebner, E. S. (2020). Occupational stress and employees complete mental health: A cross-cultural empirical study. *International Journal of Environmental Research and Public Health*, *17*(10).
Muñoz, S., & Iglesias, C. A. (2022). A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations. *Information Processing & Management*, *59*(5), Article 103011.
Muñoz, S., & Iglesias, C. A. (2021). An agent based simulation system for analyzing stress regulation policies at the workplace. *Journal of Computer Science*, *51*, Article 101326.
Murray, H. A. (1938). *Explorations in personality*. Oxford Univ. Press.
Novais, P., & Carneiro, D. (2016). The role of non-intrusive approaches in the development of people-aware systems. *Progress in Artificial Intelligence*, *5*(3), 215–220.
O'Connor, D. B., Thayer, J. F., & Vedhara, K. (2021). Stress and health: A review of psychobiological processes. *Annual Review of Psychology*, *72*(1), 663–688.
Osman, A., Wong, J. L., Bagge, C. L., Freedenthal, S., Gutierrez, P. M., & Lozano, G. (2012). The depression anxiety stress scales—21 (DASS-21): further examination of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology*, *68*(12), 1322–1338.
Osmani, V., Carreras, I., Matic, A., & Saar, P. (2014). An analysis of distance estimation to detect proximity in social interactions. *Journal of Ambient Intelligence and Humanized Computing*, *5*(3), 297–306.
Pakarinen, T., Pietilä, J., & Nieminen, H. (2019). Prediction of self-perceived stress and arousal based on electrodermal activity. In *2019 41st annual international conference of the ieee engineering in medicine and biology society (EMBC)* (pp. 2191–2195). IEEE.
Parent-Thirion, A., Fernández-Macías, E., Hurley, J., & Vermeylen, G. (2007). *Fourth European working conditions survey*: *Technical report*, European Foundation for the Improvement of Living and Working Conditions.
Parent-Thirion, A., Vermeylen, G., van Houten, G., Lyly-Yrjänäinen, M., Biletta, I., & Cabrita, J. (2012). *Fifth European working conditions survey*: *Technical report*, European Foundation for the Improvement of Living and Working Conditions.
Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J.-R., Mérienne, F., Benedetto, S., & Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction*, *30*(3), 220–236.
Petitta, L., Probst, T. M., Ghezzi, V., & Barbaranelli, C. (2021). The impact of emotional contagion on workplace safety: Investigating the roles of sleep, health, and production pressure. *Current Psychology*, 1–15.
Picard, R. W. (2000). *Affective computing*. MIT Press.
Ribeiro Santiago, P., Nielsen, T., Smithers, L., Roberts, R., & Jamieson, L. (2020). Measuring stress in Australia: validation of the perceived stress scale (PSS-14) in a national sample. *Health and Quality of Life Outcomes*, *18*.
Rigó, M., Dragano, N., Wahrendorf, M., Siegrist, J., & Lunau, T. (2021). Work stress on rise? Comparative analysis of trends in work stressors using the European working conditions survey. *International Archives of Occupational and Environmental Health*, *94*(3), 459–474.
Robusto, C. C. (1957). The cosine-haversine formula. *American Mathematical Monthly*, *64*(1), 38–40.
Sadat-Mohammadi, M., Shakerian, S., Liu, Y., Asadi, S., & Jebelli, H. (2021). Non-invasive physical demand assessment using wearable respiration sensor and random forest classifier. *Journal of Building Engineering*, *44*, Article 103279.
for Safety, E. A., & at Work, H. (2013). *European opinion poll on occupational safety and health*: *Technical report*, European Agency for Safety and Health at Work.
Selye, H. (1956). *The stress of life*. McGraw-Hill.
Shacham, S. (1983). A shortened version of the profile of mood states. *Journal of Personality Assessment*, *47*(3), 305–306.
Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine*, *108*(3), 1287–1301.
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32.
Singh, R. R., Conjeti, S., & Banerjee, R. (2013). A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control*, *8*(6), 740–754.
Tomba, K., Dumoulin, J., Mugellini, E., Abou Khaled, O., & Hawila, S. (2018). Stress detection through speech analysis. In *ICETE (1)* (pp. 560–564).
Universari, N., & Harsono, M. (2021). Antecedents and consequences of work stress behavior. *Economics and Business Solutions Journal*, *5*(1), 46–63.
Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (pp. 3–14).
Wang, J., Fu, E. Y., Ngai, G., Leong, H. V., & Huang, M. X. (2019). Detecting stress from mouse-gaze attraction. In *Proceedings of the 34th ACM/SIGAPP symposium on applied computing* (pp. 692–700).
Zhang, J., Mei, X., Liu, H., Yuan, S., & Qian, T. (2019). Detecting negative emotional stress based on facial expression in real time. In *2019 IEEE 4th international conference on signal and image processing (ICSIP)* (pp. 430–434). IEEE.

3.2.3 Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices

Title	Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices
Authors	Muñoz, Sergio and Iglesias, Carlos A. and Scheianu, Andrei and Suciu, George
Journal	Electronics
Impact factor	JCR 2021 2.690
ISSN	2079-9292
Volume	11
Year	2022
Keywords	automation
Pages	326
Online	https://www.mdpi.com/2079-9292/11/3/326
Abstract	The evolution of ambient intelligence has introduced a range of new opportunities to improve people's well-being. One of these opportunities is the use of these technologies to enhance workplaces and improve employees' comfort and productivity. However, these technologies often entail two major challenges: the requirement for fast and reliable data transmission between the vast number of devices connected simultaneously, and the interoperability between these devices. Conventional communication technologies present some drawbacks in these kinds of systems, such as lower data rates and electromagnetic interference, which have prompted research into new wireless communication technologies. One of these technologies is visible light communication (VLC), which uses existing light in an environment to transmit data. Its characteristics make it an up-and-coming technology for IoT services but also aggravate the interoperability challenge. To facilitate the continuous communication of the enormous amount of heterogeneous data generated, highly agile data models are required. The semantic approach tackles this problem by switching from ad hoc application-centric representation models and formats to a formal definition of concepts and relationships. This paper aims to advance the state of the art by proposing a semantic vocabulary for an intelligent automation platform with VLC enabled, which benefits from the advantages of VLC while ensuring the scalability and interoperability of all system components. Thus, the main contributions of this work are threefold: (i) the design and definition of a semantic model for an automation platform; (ii) the development of a prototype automation platform based on a VLC-based communication system; and (iii) the integration and validation of the proposed semantic model in the VLC-based automation platform.

electronics

Article

Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices

Sergio Muñoz [1,*], Carlos A. Iglesias [1], Andrei Scheianu [2] and George Suciu [2]

[1] Intelligent Systems Group, Universidad Politécnica de Madrid, Avenida Complutense 30, 28040 Madrid, Spain; carlosangel.iglesias@upm.es
[2] R&D Department, BEIA Consult International, 041386 Bucharest, Romania; andrei.scheianu@beia.ro (A.S.); george@beia.ro (G.S.)
* Correspondence: sergio.munoz@upm.es

Abstract: The evolution of ambient intelligence has introduced a range of new opportunities to improve people's well-being. One of these opportunities is the use of these technologies to enhance workplaces and improve employees' comfort and productivity. However, these technologies often entail two major challenges: the requirement for fast and reliable data transmission between the vast number of devices connected simultaneously, and the interoperability between these devices. Conventional communication technologies present some drawbacks in these kinds of systems, such as lower data rates and electromagnetic interference, which have prompted research into new wireless communication technologies. One of these technologies is visible light communication (VLC), which uses existing light in an environment to transmit data. Its characteristics make it an up-and-coming technology for IoT services but also aggravate the interoperability challenge. To facilitate the continuous communication of the enormous amount of heterogeneous data generated, highly agile data models are required. The semantic approach tackles this problem by switching from ad hoc application-centric representation models and formats to a formal definition of concepts and relationships. This paper aims to advance the state of the art by proposing a semantic vocabulary for an intelligent automation platform with VLC enabled, which benefits from the advantages of VLC while ensuring the scalability and interoperability of all system components. Thus, the main contributions of this work are threefold: (i) the design and definition of a semantic model for an automation platform; (ii) the development of a prototype automation platform based on a VLC-based communication system; and (iii) the integration and validation of the proposed semantic model in the VLC-based automation platform.

Keywords: semantic modeling; task automation; visible light communication; smart office; environment adaptation; ambient intelligence

Citation: Muñoz, S.; Iglesias, C.A.; Scheianu, A.; Suciu, G. Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices. *Electronics* **2022**, *11*, 326. https://doi.org/10.3390/electronics11030326

Academic Editor: José Santa

Received: 23 December 2021
Accepted: 19 January 2022
Published: 20 January 2022

1. Introduction

The extension of Internet connectivity to physical devices and everyday objects has become the foundation of the concept of "smart homes" [1]. The main goal of smart technology is to improve people's well-being by reacting and adapting according to information obtained from the environment [2]. Enabling new services and applications that take advantage of the interconnection of the physical and virtual domains leads to new opportunities and opens up a range of possibilities [3]. One of these possibilities is the application of these technologies in offices and workspaces, allowing companies to improve the comfort, performance, and productivity of their employees by supporting and meeting their needs [4].

However, the huge (and increasing) number of connected devices, along with the requirement for fast, reliable, and cost-effective data transmission, pose a challenge for these kinds of systems [5]. Several Radio Frequency (RF)-based wireless technologies have been used to provide Internet of Things (IoT) services: WiFi [6], Bluetooth [7], Near-Field Communication (NFC) [8], or Zigbee [9], and also Low Power Wide Area Network (LPWAN) technologies such as SigFox [10], NB-IoT [11], or LoRa/LoRaWAN [12]. Nevertheless, these technologies present some drawbacks, such as Electromagnetic Interference (EMI), a limited frequency spectrum, and, in some cases, a lower data rate [5]. These problems have pushed the investigation of new wireless communication technologies, such as Visible Light Communication (VLC) [13]. VLC uses the existing light in an environment (e.g., Light Emitting Diode (LED) lights) for transmitting data, exploiting the bandwidth of visible light, which is almost ten times wider and completely different from that of conventional wireless access technologies [14]. Furthermore, the use of existing LED-based lighting infrastructure in buildings for lighting and data communication simultaneously can notably reduce the operating costs and carbon footprint [5]. These features make VLC a very promising technology for IoT services [15].

The success of the technology largely depends on its user acceptance and ease of use [16]. In this regard, one of the major challenges of intelligent environments is the interoperability between different platforms and components, which is further compounded by the integration of new communication technologies [17]. The information in these systems is transmitted to and from various devices and services through heterogeneous network protocols and communication technologies [18]. In order to facilitate the continuous communication of the huge amount of heterogeneous data generated, very agile data models are required [19]. The semantic approach addresses this problem by replacing ad hoc application-centric representation models and formats with a formal definition of concepts and relationships [20]. These definitions are known as ontologies or vocabularies and enable the interoperable representation of inferred data, solving the problems of isolation between heterogeneous information and providing a better understanding of the environment [21]. In addition, semantic representation significantly improves system scalability by providing a rich, machine-readable format that can be understood, reasoned about, and reused. These capabilities have boosted the tendency of integrating semantic technologies into the Internet of Things [22]. Nevertheless, there is still a lack of a semantic model that covers the representation of the automation domain. Moreover, whereas smart office automation is already being implemented by using conventional wireless technologies, VLC is hardly used [23].

This paper aims to fill this gap by proposing a semantic vocabulary for an intelligent automation platform based on VLC that benefits from the advantages of VLC while guaranteeing the interoperability and scalability of all system components. Thus, the main contributions of this paper are threefold: (i) the design and definition of a semantic model for an automation platform; (ii) the development of a prototype automation platform based on a VLC-based communication system; (iii) the integration and validation of the proposed semantic model in the VLC-enabled automation platform.

The remainder of the article is structured as follows: Section 2 introduces the background and related work of the technologies used in this work. Once the technologies involved in this work have been reviewed, a scenario for introducing the proposed system is described in Section 3. The design and definition of the semantic model are explained in Section 4, while a prototype of the VLC-enabled semantic task automation platform is given in Section 5. Lastly, an evaluation of the proposed semantic model using competency questions is presented in Section 6, and Section 7 describes the conclusions drawn from this paper and an outline of possible lines of future work is presented.

2. Related Work

This section presents the background and related work of the technologies involved in the paper. Given the different technologies used in this work, this section has been split into three subsections. First, Section 2.1 gives an overview of related work in ambient intelligence and specifically in smart offices. Then, an introduction and revision of the state of the art in VLC are described in Section 2.2. Lastly, Section 2.3 presents the state of art regarding semantic technologies.

2.1. Smart Offices

The rise in the popularity of IoT has spread widely to simple household applications and everyday tasks, intending to improve the well-being and comfort of people. For example, a system can be configured to monitor specific parameters of an environment, such as temperature, air quality, or luminosity, helping users to keep them within an acceptable range and improving human comfort [24].

Applying these technologies to the work environment has resulted in the concept of the "smart office". This new paradigm for industrial activities has been applied to various official and industrial tasks [25], including manufacturing, healthcare, communications, and agriculture [26]. A company can leverage these technologies to improve the comfort and well-being of the workers, which may result in higher performance and productivity [4]. With this aim, smart offices must efficiently and proactively manage the infrastructure deployed in the workplace and the enterprise systems [27]. The scientific literature shows different approaches addressing smart offices and identifies the main challenges to tackle in the design, implementation, and optimization of these systems [28–31]. Shigeta et al. [32] propose a smart office system able to recognize the workers' mental and physiological states through several biometric sensors and to adapt the environment to the mood of the users for enhancing their comfort. The adaptation uses an aroma generator, a speaker, and color lights. The use of lights is also exploited by Andrei Mocanu [33], who provided a solution aimed at identifying a specific task and adapting the light accordingly. The system can predict future activities and provide reliable recommendations. These technologies have also been used for improving energy efficiency, as proposed by Xin Wang et al. [34]. They describe an automatic office lighting control system based on the occupancy status. In addition, the presence detection system can also be used as a security system to discover the presence of intruders or burglars.

The reviewed works show specific examples of smart office solutions to improve well-being. Several systems capable of performing environment adaptation according to the users' needs are presented in them. However, as far as we know, no work in the literature adequately addresses the use of VLC as a communication system in a smart office to take advantage of the benefits of this type of technology compared to RF-based communication. This work aims to fill this gap by proposing a smart office automation platform that uses a VLC system to communicate with sensors.

2.2. Visible Light Communication

The growing demand for high data rates and the immense number of connected appliances leads to the overloading of conventional radiofrequency technologies, saturating the RF spectrum. This makes it necessary to resort to new communication technologies. One of the technologies that has gained much popularity in recent years is VLC [13]. VLC has emerged as an alternative solution, in which a light source is used simultaneously for lighting and data transmission. The possibility of enabling simultaneous illumination and communication has led VLC to become a novel alternative to RF. Compared to RF links, VLC links feature a very large bandwidth that enables much faster transmission rates. Specifically, VLC uses the visible light region of the electromagnetic spectrum, whose wavelength ranges from 380 nm to 780 nm, equivalent to a frequency between 430 THz and 790 THz approximately [35]. Compared to other wireless technologies, VLC technology presents better stability, higher speed, immunity to electromagnetic interference, a higher level of security, the availability of unlicensed channels, and lower costs [36]. The superior stability of VLC systems refers to the specific low degree of failure. As long as a VLC communication system does not require a permanent Internet connection, it will continue to operate (data transmission) at full capacity as long as the lighting infrastructure is operational. Moreover, VLC can achieve data rates of over 10 Gbit/s and can be used in electromagnetically sensitive areas without electromagnetic interference. From the point of view of the higher degree of security compared to WiFi technology, it must be taken into account that a closed VLC communication system can only be penetrated from inside the areas where it is implemented. Outside penetration, as can be encountered in the case of systems based on the usage of TCP/IP communication protocols, is impossible. Furthermore, even if an intruder succeeds in traversing the VLC perimeters, the penetration of the optical communication channel is only possible by using specially developed VLC devices for demodulating the VLC data packets. Hence, a high degree of reliability and security in communication can be guaranteed [35]. Finally, VLC is license-free and it is considered cheaper than other licensed technologies (for example, it is expected to be ten times cheaper than WiFi [36]). This communication system can be based on low-cost LED devices, taking advantage of their capability to switch between different light intensities at high speed, so that it is imperceptible to the human eye [37]. In this way, they facilitate high data transfer. Nowadays, LED illumination is the most broadly used light source [38], providing a robust infrastructure for establishing fast VLC communications.

The beginnings of visible light communication date back to the 2000s in Japan, when researchers at Keio University used white LED for illumination and communication [39]. Following this achievement, several studies have been published on the use of LEDs in communication systems, and many initiatives have been observed to foster and standardize VLC technology. One of these initiatives is the Visible Light Communication Association (VLCA) (available at https://j-photonics.org/vlca/en/ (accessed on 10 June 2021), which is the successor of Visible Light Communication Consortium (VLCC), a consortium including the main Japanese companies that was established in 2003 for the standardization of VLC technology. In addition, in 2011, the IEEE 802.15 working group for wireless personal area networks published the first IEEE standard related to VLC (802.15.7-2011). The details of this standard regarding data rates, modulation schemes, coding, and attenuation mechanisms are discussed in [40].

However, regardless of the great advantages of VLC technology and the variety of its applications, its dependence on a direct line of sight has hindered its use. In recent years, hybrid RF/VLC systems have been proposed to benefit from the high capacity of VLC links and the improved connectivity of RF links. Therewith, RF/VLC hybrid systems are a pivotal element in optimizing capacity, interference, and overall network power consumption. Numerous studies are investigating how RF and VLC systems can complement each other, leveraging the advantages of each of the technologies to overcome their limitations. One of the first studies to present this idea was proposed by Haijun Zhang et al. [41], presenting an RF/VLC asymmetric combination in which VLC is used for downlink and RF for uplink in a duplex system. A hybrid RF/VLC system was also proposed by Jiaxuan Chen et al. [42] for a vehicular communication system. In addition, the use of hybrid technologies for energy efficiency improvement represents an important application. Justin Kong et al. [43] explored this idea to improve the energy efficiency of a heterogeneous network of VLC and RF systems. The obtained results of the conducted experiments confirm the improvement in the energy consumption of the proposed system. Moreover, many studies have been conducted on VLC to improve its functionality [13,44]. However, the use of VLC for providing IoT services has not been extensively studied [14].

2.3. Semantic Technologies in Ambient Intelligence

The evolution and growth of ambient intelligence have boosted the number of connected devices and services in our daily lives. The result is that an enormous amount of data is produced. The great heterogeneity of the IoT in terms of devices, communication protocols and technologies, and data formats makes semantic interoperability one of the most critical challenges to resolve in these systems [22]. Therefore, it is necessary to establish a standard set of rules for exchanging information between independent systems. These rules should include expected formats, schemas, and behavior. Semantic technologies help to address this problem with a standard definition of concepts and relationships that replaces application-centric representation models and formats. These definitions establish a mechanism for sharing knowledge, and are known as ontologies or vocabularies [45]. Ontologies help to exchange semantically annotated information between heterogeneous applications, resulting in a significant enhancement in interoperability and scalability.

Each ontology typically represents a domain in detail, and they take some concepts from others when necessary. Later, systems can use parts or modules of different ontologies jointly to represent the full breadth of their knowledge. Semantic modeling has become a ubiquitous approach to data semantization in healthcare and smart cities. One of the first ontologies for IoT systems was Brick, proposed by Balaji et al. [46]. Brick is an ontology for energy efficiency in smart buildings that uses tags to identify components and devices inside buildings. Another relevant ontology for IoT systems is the Smart Appliance Reference Ontology (SAREF) [47], intended to describe, at a high level, smart appliances along with their services. The SAREF model includes concepts such as sensor, service, state, and function, being generic enough to be used in different smart domains. However, it applies most naturally to limited spaces such as homes or offices. Finally, the W3C Semantic Sensor Network Incubator Group proposed the Semantic Sensor Network (SSN) ontology [48], one of the most prominent ontologies within the IoT domain. SSN, published in 2012, enables the representation of concepts such as sensor, actuator, and observation. A few years later, in 2017, the SSN ontology was updated, giving rise to the Sensor, Observation, Sample, and Actuator (SOSA) ontology [49].

The previously described ontologies allow the modeling of the devices, platforms, and communication systems present in an intelligent environment, but do not cover the modeling of the automation process. The Evented Web (EWE) ontology [50] was designed with this purpose: to enable the modeling of the rule-based automation process in an intelligent environment. It describes the most significant aspects of task automation services in an illustrative way, and provides a standard model for its representation. The main elements of the ontology have been defined and formalized on the basis of several identified perspectives (configurability, privacy, communication, and integration). Moreover, EWE is based on Web Ontology Language (OWL) 2 classes, and there are implementations of EWE enabling inferences using a SPARQL Inferencing Notation (SPIN) Engine (TopBraid (available at https://www.w3.org/2001/sw/wiki/TopBraid (accessed on 20 June 2021)) and Notation 3 (N3) Logic (Euler Yet another proof Engine (EYE)) (available at http://eulersharp.sourceforge.net/ (accessed on 5 June 2021)).

With the rise of the Semantic Web, technologies for defining and managing ontologies have grown considerably. For example, numerous repositories have emerged for storing data in a semantic format, such as Virtuoso, Apache Jena TDB, and GraphDB. In addition, the formats, protocols, and conventions used to model and exchange semantic information are standardized and well known (SPARQL Protocol and RDF Query Language (SPARQL), Resource Description Framework (RDF), JSON-based Serialization for Linked Data (JSON-LD)). However, no semantic model has been developed for an automation platform based on VLC communication, to the best of our knowledge.

3. Scenario

To better understand the scope of the proposed system, we have defined the following scenario. Consider a highly skilled Information Technology (IT) Small and Medium Enterprise (SME) where company leaders want to improve the comfort and well-being of their employees. Company leaders expect that an increase in their well-being would be translated into better performance and productivity [51]. For this purpose, they want to implement an intelligent automation platform in their offices that adapts the environment to the workers' needs. The employee's expectations for a workplace include proper lighting, a comfortable temperature, clean air, and reduced noise levels [4,52,53]. In order to address these expectations, the company proposes the use of different sensors and actuators that help to maintain these conditions at their optimum levels:

1. A temperature sensor is used to detect whether the office temperature is outside of the range configured as a reference for the employees' comfort;
2. A motion sensor is used to detect whether there are or are not people in the office;
3. A dust sensor is used to measure the air quality and detect when this quality is below the required value;
4. A luminosity sensor is used to continuously monitor the intensity of the light in order to guarantee optimum light conditions;
5. A noise sensor is intended to detect whether the environment's noise level is too high;
6. Intelligent LED lights enable the configuration of the intensity, color, and temperature;
7. An air purifier is able to filter the air to improve its quality;
8. A smart Heating, Ventilation, Air Conditioning (HVAC) system allows users to control the air conditioning and heating in order to achieve the optimum temperature;
9. A voice assistant is intended to provide voice alerts to users when any of the conditions is outside of the comfort ranges.

In order to not interfere with the existing RF systems in the office, the company considers the use of VLC for communicating with the sensors. In this way, the company can use the LED lights simultaneously for illumination and communication. Furthermore, intending to guarantee the interoperability and scalability of the system, the company proposes the use of semantic modeling for all the components involved in the intelligent environment. This allows the company to ensure interoperability between all the elements in the system and facilitate the integration of new components. An overview of the proposed smart environment is given in Figure 1.

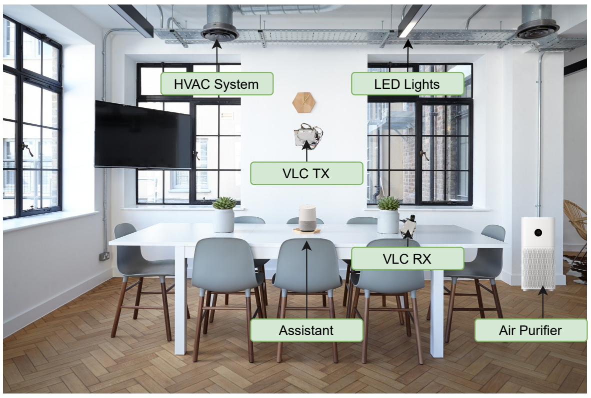


Figure 1. Smart office scenario.

Once the semantic task automation platform is implemented, the company can define different automation rules that help to maintain the ambient conditions at the required levels. Some examples of these rules are:

1. *If the dust level is too high, increase the power of the air purifier*. This rule helps to maintain a good value for the air quality, increasing the power of the air purifier when the dust level rises above a certain limit.
2. *If the noise level is high, suggest silence by means of the voice assistant*. The voice assistant can be used for sending alerts to workers. In this rule, an alert is played by the voice assistant if the noise level is too high.
3. *If the temperature falls below 20 °C, adjust the temperature to 22 °C*. Working at too low or too high temperatures can significantly reduce employees' well-being, so this rule enables the automatic control of the ambient temperature to prevent too low values.

The platform allows the adaptation of the environment to the users' needs. Therefore, the company and workers can configure and customize their own automation rules to enhance the working conditions, comfort, and well-being. The following section describes the approach followed for the semantic modeling of the platform.

4. Semantic Modeling

In order to apply a semantic layer to the VLC-enabled automation platform, a vocabulary with the corresponding relationships between ontologies has been defined. This vocabulary enables the semantic modeling of all entities in the smart office environment. The defined vocabulary is based on two existing ontologies: SSN (and its core, SOSA) and EWE. Figure 2 shows the classes of these ontologies that we have used for defining our model, along with the relationships between them.

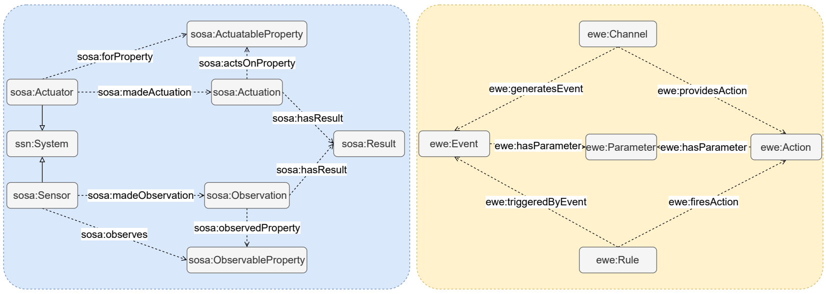


Figure 2. Main classes of SOSA and EWE ontologies.

The SSN ontology is organized into several modules [54]. Its core module is the so-called SOSA (Sensor, Observation, Sampling, Actuator), an independent and lightweight module that defines the elementary classes and properties. SOSA can be used as a standalone ontology to broaden the target audience and application, as it provides a flexible framework that enables the representation of all the entities involved in sensing and actuation, along with the relationships between them [49]. We can analyze SOSA from two different perspectives: observation and actuation.

From the observation perspective, we find classes that provide a pattern and terminology that can be used for sensing and measuring activities. The observation of a property is considered an event, completed when the result is available. In particular, SOSA defines the class *Sensor*, a physical device that performs an *Observation* about some *ObservableProperty*. From the actuation perspective, we find classes that allow us to model the changes and actuations. With this purpose, SOSA provides the class *Actuator*, which represents a physical device performing an *Actuation* on some *ActuatableProperty*. *ObservableProperty* and *ActuableProperty* are subclasses of *ssn:Property*, which represents a quality of an entity. Finally, SOSA provides two models to attach the value of a property to observation or actuation activities. First, the *sosa:hasSimpleResult* property is used to indicate a simple literal. Second, the object property *sosa:hasResult* is used to link an individual to the observation.

The SSN ontology allows us to model the physical devices existing in the system. In order to model the process of automation, we use the EWE ontology [50], a vocabulary that allows us to model the most relevant aspects of task automation services. The core of EWE is formed by four major classes: *Channel*, *Event*, *Action*, and *Rule*. The class *Channel* represents entities (devices or services) that either generate events, provide actions, or both. In the smart office context, sensors and actuators such as a motion sensor or an air purifier are described as channels, and their observations or actuations as events and actions. The class *Event* represents a particular occurrence of a process (such as an observation of a property) and allows users to define under which conditions the corresponding actions should be triggered. Event individuals are generated by a specific channel and usually provide additional details that can be modeled as parameters, using the class *Parameter*. These parameters can be used within rules to customize actions. The detection of motion generated by the motion detector sensor is an example of an entity that belongs to the *Event* class. The class *Action* represents an operation (or actuation) provided by a channel that is triggered under some conditions. Actions provide actuations on specific properties, and can be configured to react according to the data collected from a certain event using the class *Parameter*. Following the smart office context mentioned above, to switch on the air purifier is an example of *Action* generated by the air purifier channel. Lastly, the class *Rule* represents an *Event, Condition, Action* rule, i.e., an automation rule triggered by an event that fires the execution of an action. An example of a rule is: "*If a motion is detected, switch on the air purifier*".

Based on these two ontologies, we propose a semantic model for the VLC-enabled automation platform. A general overview of the model is given in Figure 3. In order to benefit from the features of the SSN and EWE ontologies, the proposed model divides some of the provided classes into subclasses. In this way, we define the classes *vlctas:SensorChannel* and *vlctas:ActuatorChannel*, which are subclasses of *ewe:Channel*. In addition, *vlctas:SensorChannel* is a subclass of *sosa:Sensor* and *vlctas:ActuatorChannel* is a subclass of *sosa:Actuator*. This allows us to treat each channel as a sensor or an actuator depending on its nature. Moreover, by means of the *ewe:generatesEvent* and *ewe:providesAction* properties, we can model the events and actions of a channel. We consider that each *sosa:Observation* can be seen as an *ewe:Event*, and each *sosa:Actuation* as a *ewe:Action*. Following this idea, we have defined the class *vlctas:ObservedEvent*, which is a subclass of *ewe:Event* and *sosa:Observation*; meanwhile, *vlctas:TriggeredAction*, is a subclass of *ewe:Action* and *sosa:Actuation*. Finally, the observations and actuations have a result, which we can consider a *ewe:Parameter*. In this way, we have defined the class *vlctas:Parameter* as a subclass of *ewe:Parameter* and *sosa:Result*.

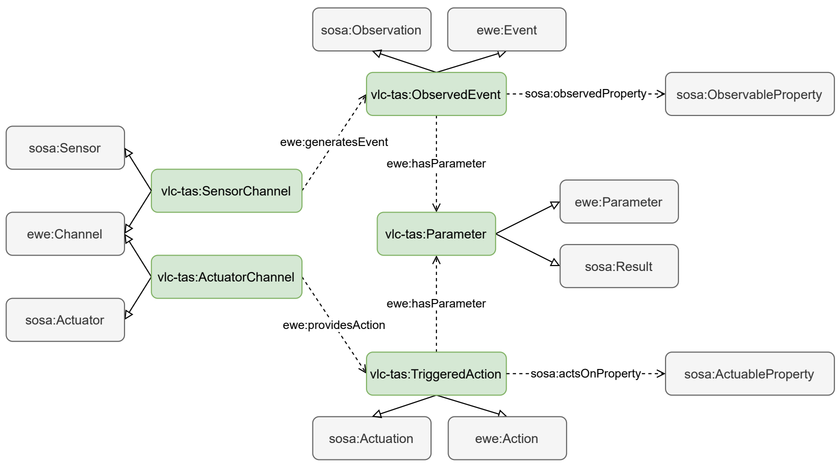


Figure 3. Semantic model proposed.

In the considered scenario, each sensor or actuator can be modeled as a channel: temperature sensor, air purifier, etc. In order to offer a better idea of how specific components of the system have been modeled, we show an example written in Notation3 with the modeling of one of the rules described in Section 3: "*If the temperature falls below 20 °C, increase the temperature to 22 °C*". The first step is to define the property to measure or change (in this case, the temperature). Then, we can define the channels representing the sensors and actuators, along with their events and actions. Finally, we could also define some custom properties useful for defining the parameters of events and actions. Listing 1 shows these definitions.

Listing 1. Semantic representation written in N3 of channels, events and actions related to temperature.

```
vlctas:Temperature a owl:Class ;
rdfs:label "Temperature"@en ;
rdfs:isDefinedBy vlctas: ;
rdfs:subClassOf sosa:ObservableProperty;
rdfs:subClassOf sosa:ActuableProperty.

vlctas:TemperatureSensor a vlctas:SensorChannel;
rdfs:label "Temperature sensor"@en;
sosa:observes vlctas:Temperature ;
ewe:generatesEvent vlctas:TemperatureMeasured .

vlctas:HVACSystem a vlctas:ActuatorChannel ;
sosa:forProperty vlctas:Temperature ;
ewe:providesAction vlctas:AdjustTemperature .

vlctas:TemperatureMeasured a vlctas:ObservedEvent ;
rdfs:label "Temperature Measured"@en;
sosa:observedProperty vlctas:Temperature.

vlctas:AdjustTemperature a vlctas:TriggeredAction ;
rdfs:label "Adjust Temperature"@en;
sosa:actsOnProperty vlctas:Temperature.

vlctas:hasTemperature rdfs:subPropertyOf ewe:hasParameter ;
rdfs:subPropertyOf sosa:hasSimpleResult ;
schema:domainIncludes vlctas:TemperatureMeasured ;
schema:domainIncludes vlctas:AdjustTemperature ;
schema:rangeIncludes xsd:double .

vlctas:hasTime rdfs:subPropertyOf ewe:hasParameter ;
rdfs:subPropertyOf sosa:resultTime ;
schema:domainIncludes vlctas:TemperatureMeasured ;
schema:rangeIncludes xsd:dateTime .
```

In this way, every observation or actuation of a device can be modeled semantically. For example, Listing 2 shows the definition of two events representing the observation of the temperature at two different times. The events have been modeled with grounded parameters on the basis of the concepts defined in the listing given above.

Listing 2. Instances of the event of temperature measurement.

```
:tempmeasure1 rdf:type vlctas:TemperatureMeasured ;
ewe:generatedBy vlctas:TemperatureSensor ;
vlctas:hasTemperature "15.7"^^xsd:double ;
vlctas:hasTime "2021-09-24T06:56:12Z"^^xsd:dateTime .

:tempmeasure2 rdf:type vlctas:TemperatureMeasured ;
ewe:generatedBy vlctas:TemperatureSensor ;
vlctas:hasTemperature "22.7"^^xsd:double ;
vlctas:hasTime "2021-09-24T12:51:52Z"^^xsd:dateTime .
```

Finally, Listing 3 shows the definition of rule instances. This example represents a rule that is triggered by the event of *TemperatureMeasured* when the observed temperature is lower than 20 °C and fires the action of *adjusting temperature to 22 °C* (both defined in Listing 1). Moreover, Notation3 supports rules and formulae, so we can express the rule using the formulae as shown in Listing 4. This allows us to use a semantic engine (or reasoner) to infer the action from the event received.

Listing 3. Rule instance.

```
vlctas:RegulateTemperature a ewe:Rule ;
dcterms:title "Temperature Regulation Rule" ;
ewe:triggeredByEvent [
rdf:type vlctas:TemperatureMeasured ;
ewe:generatedBy vlctas:TemperatureSensor ;
vlctas:hasTemperature [
math:lessThan "20.0" ^^xsd:double
]
] ;
ewe:firesAction [
rdf:type vlctas:AdjustTemperature ;
ewe:providedBy vlctas:HVACSystem ;
vlctas:hasTemperature "22.0" ^^xsd:double
] .
```

Listing 4. Automation rule formulae.

```
{
?event  rdf:type vlctas:TemperatureMeasured ;
ewe:generatedBy vlctas:TemperatureSensor ;
vlctas:hasTemperature [
math:lessThan "20.0"
] .
}
=>
{
?action  rdf:type vlctas:AdjustTemperature ;
ewe:providedBy vlctas:HVACSystem ;
vlctas:hasTemperature "22.0" ^^xsd:double.
}.
```

5. A Prototype of a Semantic Task Automation Platform with VLC

The proposed architecture has been designed on the basis of the reference architecture for task automation services [50], which has been extended to enable VLC. The system is composed of two main parts: a VLC system and a task automation server, as shown in Figure 4. The former implements the visible light communication system and contains the modules involved in the communication with the sensors. The automation system receives the data coming from the sensors, applies semantic modeling to these data, and evaluates the events along with the automation rules in order to trigger the respective actions. In addition, it includes several functions for managing and configuring automation rules.

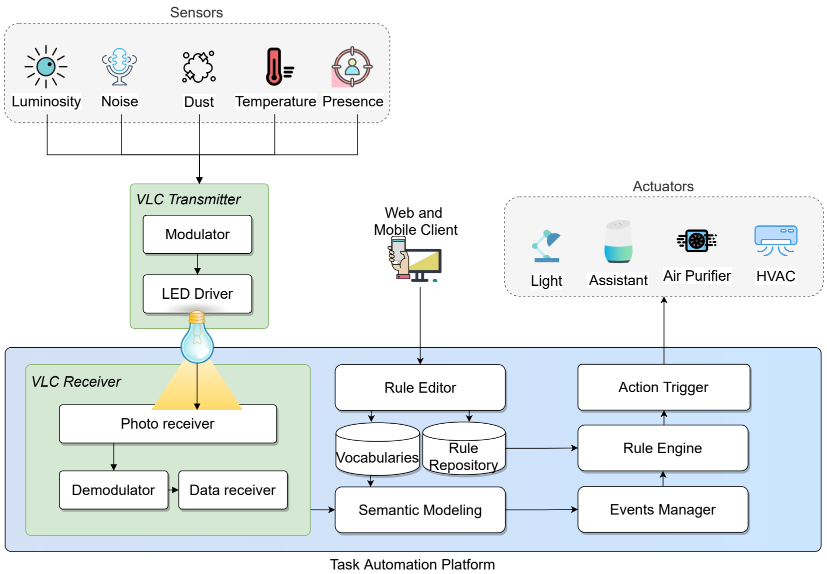
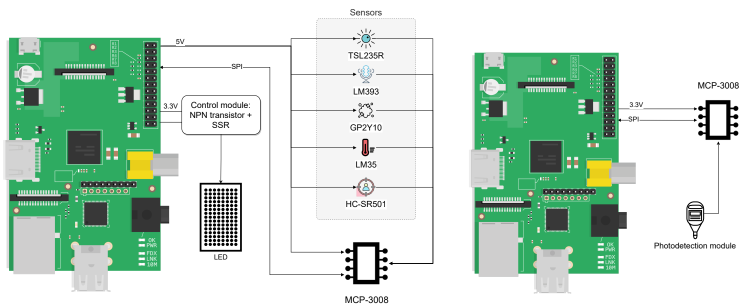


Figure 4. General architecture of the proposed system.

5.1. VLC System

This module implements the communication based on visible light. The system uses this kind of communication to receive data from sensors, so this module is responsible for the collection of these data and their dispatch to the automation platform. The module consists of two main parts: a transmitter and a receiver. The former obtains data from the sensors and sends the information to the receiver through visible light. An overview of the architecture of each part is given in Figure 5. As can be seen, both are based on the Raspberry Pi 3 Model B+ minicomputer.



(**a**) VLC transmitter. (**b**) VLC receiver.

Figure 5. Architecture of the VLC system.

The transmitter includes the processing unit, a control module, and an LED matrix. In addition, it is connected to various sensors in order to measure different ambient conditions: temperature, air quality, noise, luminosity, and motion:

1. The LM35 sensor is used to measure the environment's ambient temperature;
2. The TSL235R sensor monitors the luminosity;
3. The PIR HC-SR501 is a motion sensor designed to monitor the ambient environment and detect when there is someone in the workplace;

4. The GP2Y1014AU0F sensor is used to monitor the air quality, measuring the dust concentration within the air;
5. The LM393 module is intended to detect the environment's noise level.

The output data of each sensor are transformed into a binary number. These binary values collected for each measurement are concatenated to obtain a binary string. Using the 3.3 V logic, each 1 in the string corresponds to a HIGH signal, whereas 0 corresponds to a LOW value. The 3.3 V interface is connected to a signal conditioning circuit with a Negative-Positive-Negative (NPN) transistor and an electromechanical relay. In order to obtain data from the sensor and convert them into a digital signal, the transmitter uses an MCP-3008 analog-to-digital converter. Even if some of the sensors are digital and would not require a connection through the MCP-3008, it was preferred that all of them be connected to it. This enables easier integration and processing of the data. Moreover, in order to facilitate the integration of new sensors, the Serial Peripheral Interface (SPI) protocol is used between the Raspberry Pi and the converter. The data packets are collected by the microcontroller, and then are sent to the control circuit of the LED. In order to achieve the fast switching of the LED, this circuit is formed by two transistors in the Darlington connection. Finally, the data packets are modulated using the On-Off Keying (OOK) modulation and sent to the receptor using the LED lights.

The receptor includes the processing unit and a photodiode. The photodiode detects the optical signal emitted by LEDs and converts it into an electrical signal. Then, the microcontroller converts the analog signal into a digital one, which is dispatched to the automation platform. Finally, it is interesting to note that the photodiode can be calibrated using ambient light, in order to determine the threshold value used to identify the logic value of the received binary signal.

Regarding the placement of the transmitter and the receiver, the presence of a line-of-sight link between them is essential to provide a suitable data rate performance. For this reason, the employee behavioral patterns and building-specific electrical plans must be taken into account during the implementation planning. In this way, the possibility of unintentional blocking of optical transmission or reception sources is minimized. The proposed system has been thoroughly evaluated in previous works, which include numerous scenarios to test the system's performance, both in indoor and outdoor conditions [55,56].

5.2. Automation Platform

The main goal of this module is to enable the adaptation of the environment depending on the data received from the sensors. It is based on semantic Event, Condition, Action automation rules and provides integration with several devices and services, such as an air purifier or the HVAC system. Moreover, it allows users to configure custom automation rules or import rules created by other users easily.

The platform receives data coming from the sensors through the VLC system. It encodes these data using semantic technologies (as described in Section 4), and then the received events are captured and dispatched to the *rule engine*, where they are evaluated along with the stored automation rules. The *rule engine* module receives events from the *events manager* and loads the automation rules stored in the repository. It consists of a semantic engine reasoner [57] based on an ontology model. Once the event is received and the available rules are loaded, the reasoner runs the ontology model inferences in order to draw the corresponding actions depending on the incoming events and the automation rules. The drawn actions are then sent to the *action trigger*, which connects with the corresponding actuators channels that will perform the generated actions. This connection with the channels is carried out using their own communication interfaces. Therefore, the integration of such components and optical communication elements into the platform actually results in a hybrid system.

In addition, the platform provides the repositories where rules are stored, as well as functions for managing rules and channels. In order to enable the configuration and management of automation rules, the platform provides a mobile client and a web client where users can easily create, remove, or edit rules. The clients connect with the *rule editor* module, which handles the corresponding changes in the repositories. The platform includes two different repositories: the *rule repository* and the *vocabularies repository*. The former is used to store information regarding rules, whereas the latter contains information regarding channels. In the smart office context proposed, the rules are intended to regulate the ambient conditions in order to maximize comfort and productivity. With this aim, the rules enable the automation and regulation of temperature, air quality, noise, and luminosity conditions.

The implementation of this architecture has been performed using Python, and the semantic automation is based on N3 technology and the EYE reasoning engine (available at http://n3.restdesc.org/) (accessed on 27 June 2021). This results in a lightweight program that can be run also on devices with low resources. For the web and mobile clients, React and Kotlin are used, respectively.

Finally, the system enables rule adaptation on the basis of channel description. Rule adaptation is based on discovering whether the channels used by a specific rule are integrated into an intelligent environment. The system detects available channels of the same class used by the rule and asks for a confirmation from the user to import the *adapted rule*. In this way, rules can be adapted to different channel providers. The EWE ontology enables this adaptation by means of the OWL2 (available at: https://www.w3.org/TR/owl2-overview/) (accessed on 4 July 2021) punning mechanism for attaching properties to channels [50]. Moreover, the notion of adapters [58,59] is used for supporting the semantic integration of devices and services. These adapters interact with devices and Internet services in order to provide a semantic output.

6. Evaluation

In order to evaluate the proposed semantic model, we defined a set of Competency Questions (CQs) to check its selection capabilities. CQs represent the ontology requirements and consist of a set of queries related to the data that the ontology must be able to answer correctly [60]. A common user may complete these queries in order to retrieve information about the rules and channels available, the events and actions of a particular type, or the number of configurable parameters. The defined CQs are:

1. *Which sensor channels are integrated into the automation platform?* Task automation platforms offer a list of available channels. Using the proposed model, it is also straightforward to filter these channels depending on whether they are sensors or actuators.
2. *Which sensors can measure temperature?* Often, users in a task automation platform need to know which of the integrated sensors can measure a specific property, or which of the actuators can change this property. The proposed model enables filtering between all the provided channels by those able to measure or act on a certain property.
3. *Which actions provide the HVAC system actuator?* Knowing which events or actions are available for a specific channel allows users in the task automation platform to have a clear notion of the capabilities of each device.
4. *Which parameters and which type accept the action "Adjust Temperature"?* This query gives insight into the configurable parameters of each event or action, which enables the custom configuration of automation rules. Knowledge of the configurable parameters allows users to benefit from all the possibilities offered by the system.
5. *How many rules are triggered by a change in noise level?* The proposed model allows users to know the platform's capabilities concerning a specific property. In this way, they could extract information about the number of rules affecting the property and obtain the particular devices involved in these rules.
6. *How many rules trigger an action provided by the AirPurifier actuator channel?* Occasionally, users need insights into the use they are giving to each channel. This allows them to analyze whether some sensor or actuator is not being used as it should.

These queries have been coded using SPARQL and run on the semantic repository. Table 1 shows each executed query, along with the results retrieved from the repository.

Table 1. Example CQs using the proposed model written in SPARQL.

CQ	SPARQL Query	Results
Which sensor channels are integrated into the automation platform?	SELECT ?sensorLabel WHERE { ?sensor rdf:type vlctas:SensorChannel ; rdfs:label ?sensorLabel. }	Luminosity sensor, Motion sensor, Noise sensor, Dust sensor, Temperature sensor
Which sensors can measure temperature?	SELECT ?sensorLabel WHERE { ?sensor rdf:type vlctas:SensorChannel ; sosa:observes vlctas:Temperature ; rdfs:label ?sensorLabel. }	Temperature sensor
Which actions provide the HVAC system actuator?	SELECT ?action WHERE { vlctas:HVACSystem ewe:providesAction ?action. ?action rdfs:label ?actionLabel. }	Adjust Temperature, Turn On, Turn Off
Which parameters and which type accept the action "Adjust Temperature"?	SELECT ?param ?type WHERE { ?param rdfs:subPropertyOf ewe:hasParameter ; schema:domainIncludes vlctas:AdjustTemperature ; schema:rangeIncludes ?type. }	Temperature (double)
How many rules are triggered by a change in noise level?	SELECT (COUNT(?rule) as ?number) WHERE { ?rule rdf:type ewe:Rule; ewe:triggeredByEvent ?event . ?event ewe:generatedBy ?channel . ?channel ?observes vlctas:Noise . }	4
How many rules trigger an action provided by the AirPurifier actuator channel?	SELECT (COUNT(?rule) as ?number) WHERE { ?rule rdf:type ewe:Rule; ewe:firesAction ?action . ?action ewe:providedBy vlctas:AirPurifier. }	5

Compared with other automation systems, the semantic modeling of the automation platform enables access to information about components that are not usually available. Furthermore, it allows users to perform complex queries about these components and to use external information available in the Linked Data cloud to complement the information about them. Finally, information available in other Linked Data sources (such as DBPedia) can be integrated into queries to the repository of the automation platform, enabling the extension of the queries with data that are present in other systems. For example, DBPedia contains further information about the sensors integrated in the platform. This information can be obtained using the SPARQL Federated Query extension (available at: https://www.w3.org/TR/sparql11-federated-query/) (accessed on 17 September 2021), which enables the execution of queries distributed over different SPARQL endpoints. Listing 5 shows the SPARQL query for obtaining the DBPedia comments of the actuators existing in the platform.

Listing 5. SPARQL query for finding the DBPedia comments of the actuators of the platform.

```
SELECT ?sensor ?comment
WHERE {
?actuator rdf:type vlctas:ActuatorChannel ;
owl:sameAs ?dbpconcept .
SERVICE <http://dbpedia.org/sparql/> {
?dbpconcept rdf:type dbo:Device;
rdfs:comment ?comment
}
FILTER (lang(?comment) = 'en')
}
```

In this way, SPARQL can be used to express queries across diverse data sources, allowing users in the platform to expand and complement their knowledge about the existing components with the information present in other data sources.

7. Discussion

This paper proposes a semantic model for a task automation platform that enables the adaptation of the environment in a smart office to the needs of the employees. The proposed model guarantees the interoperability and the scalability of all the components of the system, as well as the portability of automations. In addition, this paper contributes to the research on the use of VLC communication technology for developing intelligent systems, and proposes VLC as a communication system in the platform. The paper introduces a practical use case scenario in order to better understand the scope of the paper and to illustrate the progress made beyond the state of the art. Moreover, it presents an implementation of a VLC-enabled task automation platform, describing the different components involved. Finally, the semantic model proposed is evaluated with a set of meaningful queries.

The proposed platform architecture benefits from the advantages of semantic technologies and visible light communication to facilitate the continuous communication of all the heterogeneous data generated. Users of the platform can configure their own automation rules based on the ambient conditions to regulate these conditions and enhance comfort and productivity. With this aim, several sensors and actuators have been included in the proposed prototype for measuring and regulating different ambient conditions: air quality, temperature, luminosity, and noise.

Compared with other automation or intelligent systems [61–63], the provided platform presents some advantages. First, thanks to the semantic model, it enables a straightforward and common representation of all the components in the smart environment, along with their relationships. Apart from the enhancement in the interoperability and scalability of the system, it also enables access to linked data related to the components. This offers the users the possibility to access complementary information about the different elements of the platform. Moreover, the use of visible light communication alleviates the load of the RF spectrum. This prevents the electromagnetic interference of sensors with other devices of the environment, which may deteriorate the data rate and the reliability of the communication. Furthermore, the use of the existing LED lighting infrastructure simultaneously for illumination and data communication may entail a reduction in operating costs and carbon footprint [5]. Given the interoperability challenge that the integration of new communication technologies could entail, the use of semantic technologies is particularly useful for the system. In this way, the data exchanged between the sensors and the automation platform follow a model that facilitates continuous communication among the different components, and the system can benefit from the advantages of VLC without compromising its extensibility.

As future work, one possible line is the integration of methods for recognizing the personal parameters of the user, such as emotions or activities. This may enhance the environmental adaptation and may benefit from these technologies for regulating emotions. Moreover, the proposed system could be applied to different scenarios apart from smart offices. The high scalability offered by the developed system facilitates the extension of both the architecture and the developed tools to give a more solid solution to a broader range of scenarios. Another planned line of future work is the enhancement of the VLC system using a hybrid VLC–WiFi approach for communication with the sensors, in order to improve its reliability and robustness. In this way, WiFi could be used if the VLC link was interrupted, guaranteeing seamless reliable coverage. Finally, the data generated by the platform could be used for applying machine learning techniques. This would allow the system to learn from the users in order to be able to improve the adaptation of the environment, or even the recommendation of automation rules between users.

Author Contributions: Conceptualization, S.M., C.A.I., A.S. and G.S.; methodology, S.M. and A.S.; software, S.M. and A.S.; validation, S.M. and C.A.I.; formal analysis, S.M.; investigation, S.M. and A.S.; resources, C.A.I. and G.S.; data curation, S.M. and A.S.; writing—original draft preparation, S.M.;

writing—review and editing, C.A.I., A.S. and G.S.; visualization, S.M. and C.A.I.; supervision, C.A.I. and G.S.; project administration, C.A.I. and G.S.; funding acquisition, C.A.I. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research has received funding from the European Union's Horizon 2020 Research and Innovation programme under the TETRAMAX grant agreement no. 761349.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CQ	Competency Question
EMI	Electromagnetic Interference
EWE	Evented Web
EYE	Euler Yet another proof Engine
HVAC	Heating, Ventilation, Air Conditioning
IoT	Internet of Things
IT	Information Technology
JSON-LD	JSON-based Serialization for Linked Data
LED	Light Emitting Diode
LPWAN	Low Power Wide Area Network
NPN	Negative-Positive-Negative
N3	Notation 3
OOK	On-Off Keying
OMEGA	HOME Gigabit Access
RDF	Resource Description Framework
RF	Radio Frequency
SAREF	Smart Appliance Reference Ontology
SME	Small and Medium Enterprise
SOSA	Sensor, Observation, Sample, and Actuator
SPIN	SPARQL Inferencing Notation
SSN	Semantic Sensor Network
SPARQL	SPARQL Protocol and RDF Query Language
SPI	Serial Peripheral Interface
VLC	Visible Light Communication
VLCA	Visible Light Communication Association
VLCC	Visible Light Communication Consortium

References

1. Jiang, L.; Liu, D.Y.; Yang, B. Smart home research. In Proceedings of the 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), Shanghai, China, 26–29 August 2004; Volume 2, pp. 659–663.
2. Sovacool, B.K.; Del Rio, D.D.F. Smart home technologies in Europe: A critical review of concepts, benefits, risks and policies. *Renew. Sustain. Energy Rev.* **2020**, *120*, 109663. [CrossRef]
3. Farhan, L.; Kharel, R. Internet of things: Vision, future directions and opportunities. In *Modern Sensing Technologies*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 331–347.
4. Prasetyo, T.; Zaliluddin, D.; Iqbal, M. Prototype of smart office system using based security system. *J. Phys. Conf. Ser.* **2018**, *1013*, 012189. [CrossRef]
5. Almadani, Y.; Plets, D.; Bastiaens, S.; Joseph, W.; Ijaz, M.; Ghassemlooy, Z.; Rajbhandari, S. Visible Light Communications for Industrial Applications—Challenges and Potentials. *Electronics* **2020**, *9*, 2157. [CrossRef]
6. Sheth, J.; Dezfouli, B. Enhancing the energy-efficiency and timeliness of IoT communication in WiFi networks. *IEEE Internet Things J.* **2019**, *6*, 9085–9097. [CrossRef]

7. Ryu, D.S.; Yeon, Y.M.; Kim, S.H. Multilink Internet-of-Things Sensor Communication Based on Bluetooth Low Energy Considering Scalability. *Electronics* **2021**, *10*, 2335. [CrossRef]
8. Okwori, M.; Behfarnia, A.; Eslami, A. Towards Microscale NFC-Enabled IoT Sensors: Physical and MAC Layer Design Analysis. *IEEE Access* **2020**, *8*, 88076–88084. [CrossRef]
9. Ali, A.I.; Partal, S.Z.; Kepke, S.; Partal, H.P. ZigBee and LoRa based wireless sensors for smart environment and IoT applications. In Proceedings of the 2019 1st Global Power, Energy and Communication Conference (GPECOM), Nevsehir, Turkey, 12–15 June 2019; pp. 19–23.
10. Lavric, A.; Petrariu, A.I.; Popa, V. Sigfox communication protocol: The new era of iot? In Proceedings of the 2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI), Lisbon, Portugal, 29–30 August 2019; pp. 1–4.
11. Feltrin, L.; Tsoukaneri, G.; Condoluci, M.; Buratti, C.; Mahmoodi, T.; Dohler, M.; Verdone, R. Narrowband IoT: A survey on downlink and uplink perspectives. *IEEE Wirel. Commun.* **2019**, *26*, 78–86. [CrossRef]
12. Haxhibeqiri, J.; De Poorter, E.; Moerman, I.; Hoebeke, J. A survey of LoRaWAN for IoT: From technology to application. *Sensors* **2018**, *18*, 3995. [CrossRef]
13. Abuella, H.; Elamassie, M.; Uysal, M.; Xu, Z.; Serpedin, E.; Qaraqe, K.A.; Ekin, S. Hybrid RF/VLC systems: A comprehensive survey on network topologies, performance analyses, applications, and future directions. *IEEE Access* **2021**, *9*, 160402–160436. [CrossRef]
14. Kim, S.Y.; Kim, C.M.; Koh, S.J. Framework of IoT Services over Unidirectional Visible Lights Communication Networks. *Electronics* **2020**, *9*, 1349. [CrossRef]
15. Delgado-Rajo, F.; Melian-Segura, A.; Guerra, V.; Perez-Jimenez, R.; Sanchez-Rodriguez, D. Hybrid RF/VLC Network Architecture for the Internet of Things. *Sensors* **2020**, *20*, 478. [CrossRef]
16. Kumar, A.; Mihovska, A.; Kyriazakos, S.; Prasad, R. Visible light communications (VLC) for ambient assisted living. *Wirel. Pers. Commun.* **2014**, *78*, 1699–1717. [CrossRef]
17. Mouromtsev, D. Semantic Reference Model for Individualization of Information Processes in IoT Heterogeneous Environment. *Electronics* **2021**, *10*, 2523. [CrossRef]
18. Rahman, H.; Hussain, M.I. A comprehensive survey on semantic interoperability for Internet of Things: State-of-the-art and research challenges. *Trans. Emerg. Telecommun. Technol.* **2020**, *31*, e3902. [CrossRef]
19. Palavalli, A.; Karri, D.; Pasupuleti, S. Semantic internet of things. In Proceedings of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 4–6 February 2016; pp. 91–95.
20. Muñoz, S.; Araque, O.; Sánchez-Rada, J.F.; Iglesias, C.A. An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices. *Sensors* **2018**, *18*, 1499. [CrossRef] [PubMed]
21. Muñoz, S.; Sánchez, E.; Iglesias, C.A. An Emotion-Aware Learning Analytics System Based on Semantic Task Automation. *Electronics* **2020**, *9*, 1194. [CrossRef]
22. Shi, F.; Li, Q.; Zhu, T.; Ning, H. A Survey of Data Semantization in Internet of Things. *Sensors* **2018**, *18*, 313. [CrossRef] [PubMed]
23. Swain, K.P.; Prasad, M.V.S.V.; Palai, G.; Sahoo, J.; Mohanty, M.N. Exploiting VLC Technique for Smart Home Automation Using Arduino. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*; Dash, S.S., Vijayakumar, K., Panigrahi, B.K., Das, S., Eds.; Springer: Singapore, 2017; pp. 211–220.
24. Al-Kuwari, M.; Ramadan, A.; Ismael, Y.; Al-Sughair, L.; Gastli, A.; Benammar, M. Smart-home automation using IoT-based sensing and monitoring platform. In Proceedings of the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), Doha, Qatar, 10–12 April 2018; pp. 1–6.
25. Younan, M.; Houssein, E.H.; Elhoseny, M.; Ali, A.A. Challenges and recommended technologies for the industrial internet of things: A comprehensive review. *Measurement* **2020**, *151*, 107198. [CrossRef]
26. Elijah, O.; Rahman, T.A.; Orikumhi, I.; Leow, C.Y.; Hindia, M.N. An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges. *IEEE Internet Things J.* **2018**, *5*, 3758–3773. [CrossRef]
27. Zhang, X.; Zheng, P.; Peng, T.; He, Q.; Lee, C.; Tang, R. Promoting employee health in smart office: A survey. *Adv. Eng. Inform.* **2022**, *51*, 101518. [CrossRef]
28. Bogdan, R.; Tatu, A.; Crisan-Vida, M.M.; Popa, M.; Stoicu-Tivadar, L. A Practical Experience on the Amazon Alexa Integration in Smart Offices. *Sensors* **2021**, *21*, 734. [CrossRef]
29. Barišić, A.; Amaral, V.; Challenger, M. Enhancing Occupants Comfort and Well-being through a Smart Office setup. In Proceedings of the 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 28 September–2 October 2020; pp. 1825–1830.
30. Zenonos, A.; Khan, A.; Kalogridis, G.; Vatsikas, S.; Lewis, T.; Sooriyabandara, M. HealthyOffice: Mood recognition at work using smartphones and wearable sensors. In Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), Sydney, NSW, Australia, 14–18 March 2016; pp. 1–6.
31. Fujimoto, Y.; Fujita, K. Depth-based human detection considering postural diversity and depth missing in office environment. *IEEE Access* **2019**, *7*, 12206–12219. [CrossRef]
32. Shigeta, H.; Nakase, J.; Tsunematsu, Y.; Kiyokawa, K.; Hatanaka, M.; Hosoda, K.; Okada, M.; Ishihara, Y.; Ooshita, F.; Kakugawa, H.; et al. Implementation of a smart office system in an ambient environment. In Proceedings of the 2012 IEEE Virtual Reality Workshops (VRW), Costa Mesa, CA, USA, 4–8 March 2012; pp. 1–2.

33. Mocanu, A. An argumentative approach to smart home office ambient lighting. In *International Symposium on Intelligent and Distributed Computing*; Springer: Cham, Switzerland, 2017; pp. 225–234.
34. Wang, X.; Tjalkens, T.; Linnartz, J.P. Smart office lighting control using occupancy sensors. In Proceedings of the 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), Calabria, Italy, 16–18 May 2017; pp. 453–458.
35. Vappangi, S.; Mani, V. Concurrent illumination and communication: A survey on Visible Light Communication. *Phys. Commun.* **2019**, *33*, 90–114. [CrossRef]
36. Singh, S.; Kakamanshadi, G.; Gupta, S. Visible light communication-an emerging wireless communication technology. In Proceedings of the 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), Chandigarh, India, 21–22 December 2015; pp. 1–3.
37. Yuvaraj, M. Li-Fi technology in libraries: An introduction and overview. *Library Hi Tech News* **2016**, *33*, 1–4. [CrossRef]
38. Moghavvemi, S.; Jaafar, N.I.; Sulaiman, A.; Parveen Tajudeen, F. Feelings of guilt and pride: Consumer intention to buy LED lights. *PLoS ONE* **2020**, *15*, e0234602. [CrossRef] [PubMed]
39. Tanaka, Y.; Haruyama, S.; Nakagawa, M. Wireless optical transmissions with white colored LED for wireless home links. In Proceedings of the 11th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2000), London, UK, 18–21 September 2000; Volume 2, pp. 1325–1329.
40. Vučić, J.; Kottke, C.; Nerreter, S.; Langer, K.D.; Walewski, J.W. 513 Mbit/s visible light communications link based on DMT-modulation of a white LED. *J. Lightwave Technol.* **2010**, *28*, 3512–3518. [CrossRef]
41. Zhang, H.; Liu, N.; Long, K.; Cheng, J.; Leung, V.C.; Hanzo, L. Energy efficient subchannel and power allocation for software-defined heterogeneous VLC and RF networks. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 658–670. [CrossRef]
42. Chen, J.; Wang, Z. Topology control in hybrid VLC/RF vehicular ad-hoc network. *IEEE Trans. Wirel. Commun.* **2019**, *19*, 1965–1976. [CrossRef]
43. Kong, J.; Ismail, M.; Serpedin, E.; Qaraqe, K.A. Energy efficient optimization of base station intensities for hybrid RF/VLC networks. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 4171–4183. [CrossRef]
44. Obeed, M.; Salhab, A.M.; Alouini, M.S.; Zummo, S.A. On optimizing VLC networks for downlink multi-user transmission: A survey. *IEEE Commun. Surv. Tutorials* **2019**, *21*, 2947–2976. [CrossRef]
45. Gruber, T.R. A translation approach to portable ontology specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [CrossRef]
46. Balaji, B.; Bhattacharya, A.; Fierro, G.; Gao, J.; Gluck, J.; Hong, D.; Johansen, A.; Koh, J.; Ploennigs, J.; Agarwal, Y.; et al. Brick: Towards a unified metadata schema for buildings. In Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, Palo Alto, CA, USA, 16–17 November 2016; pp. 41–50.
47. Daniele, L.; den Hartog, F.; Roes, J. Created in close interaction with the industry: The smart appliances reference (SAREF) ontology. In *International Workshop Formal Ontologies Meet Industries*; Springer: Cham, Switzerland, 2015; pp. 100–112.
48. Compton, M.; Barnaghi, P.; Bermudez, L.; Garcia-Castro, R.; Corcho, O.; Cox, S.; Graybeal, J.; Hauswirth, M.; Henson, C.; Herzog, A.; et al. The SSN ontology of the W3C semantic sensor network incubator group. *J. Web Semant.* **2012**, *17*, 25–32. [CrossRef]
49. Janowicz, K.; Haller, A.; Cox, S.J.; Le Phuoc, D.; Lefrançois, M. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *J. Web Semant.* **2019**, *56*, 1–10. [CrossRef]
50. Coronado, M.; Iglesias, C.A.; Serrano, E. Modelling rules for automating the Evented WEb by semantic technologies. *Expert Syst. Appl.* **2015**, *42*, 7979–7990. [CrossRef]
51. Khoreva, V.; Wechtler, H. HR practices and employee performance: The mediating role of well-being. *Empl. Relat.* **2018**, *40*, 227–243. [CrossRef]
52. Muñoz, S.; Iglesias, C.A. An agent based simulation system for analyzing stress regulation policies at the workplace. *J. Comput. Sci.* **2021**, *51*, 101326. [CrossRef]
53. Di Blasio, S.; Shtrepi, L.; Puglisi, G.E.; Astolfi, A. A cross-sectional survey on the impact of irrelevant speech noise on annoyance, mental health and well-being, performance and occupants' behavior in shared and open-plan offices. *Int. J. Environ. Res. Public Health* **2019**, *16*, 280. [CrossRef] [PubMed]
54. Haller, A.; Janowicz, K.; Cox, S.J.; Lefrançois, M.; Taylor, K.; Le Phuoc, D.; Lieberman, J.; García-Castro, R.; Atkinson, R.; Stadler, C. The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. *Semant. Web* **2019**, *10*, 9–32. [CrossRef]
55. Scheianu, A.; Drosu, A.; Petre, I. Hybrid VLC Communications System for E-Learning Based on the ARM CORTEX-A53 Processor. In Proceedings of the International Scientific Conference eLearning and Software for Education, Bucharest, Romania, 23–24 April 2020; "Carol I" National Defence University: Bucharest, Romania, 2020; Volume 2, pp. 238–245.
56. Scheianu, A.; Suciu, G.; Drosu, A.; Petre, I.; Miu, D. Hybrid VLC Communications System for Increased Security Based on Raspberry Pi Microcomputer. *Ann. Disaster Risk Sci. ADRS* **2020**, *3*. [CrossRef]
57. Verborgh, R.; De Roo, J. Drawing Conclusions from Linked Data on the Web: The EYE Reasoner. *IEEE Softw.* **2015**, *32*, 23–27. [CrossRef]
58. Sánchez-Rada, J.F.; Iglesias, C.A.; Coronado, M. A modular architecture for intelligent agents in the evented web. *Web Intell.* **2017**, *15*, 19–33. [CrossRef]
59. Coronado Barrios, M. A Personal Agent Architecture for Task Automation in the Web of Data. Bringing Intelligence to Everyday Tasks. Ph.D. Thesis, Universidad Politécnica de Madrid, Madrid, Spain, 2016.

60. Wiśniewski, D.; Potoniec, J.; Ławrynowicz, A.; Keet, C.M. Analysis of Ontology Competency Questions and their formalizations in SPARQL-OWL. *J. Web Semant.* **2019**, *59*, 100534. [CrossRef]
61. Gupta, S.; Gupta, S. Analysis and Design of Smart Office Automation System. In *Computer Networks and Inventive Communication Technologies*; Smys, S., Palanisamy, R., Rocha, Á., Beligiannis, G.N., Eds.; Springer: Singapore, 2021; pp. 963–973.
62. Bhatia, M.; Sood, S.K. Exploring temporal analytics in fog-cloud architecture for smart office healthcare. *Mob. Netw. Appl.* **2019**, *24*, 1392–1410. [CrossRef]
63. Ullo, S.L.; Sinha, G.R. Advances in Smart Environment Monitoring Systems Using IoT and Sensors. *Sensors* **2020**, *20*, 3113. [CrossRef]

3.2.4 An agent based simulation system for analyzing stress regulation policies at the workplace

Title	An agent based simulation system for analyzing stress regulation policies at the workplace
Authors	Muñoz, Sergio and Iglesias, Carlos A.
Journal	Journal of Computational Science
Impact factor	JCR 2021 3.817
ISSN	18777503
Volume	51
Year	2021
Keywords	agents
Pages	101326
Online	`https://linkinghub.elsevier.com/retrieve/pii/S1877750321000247`
Abstract	Workplace stress has a significant impact on productivity, since keeping workers' stress on an adequate level results a key factor for companies to increase their performance. While a high stress level may conduct to anxiety or absenteeism, a low level may also have undesirable consequences, such as lack of motivation. To identify and understand all the elements which interfere on workers' stress results a key factor in order to improve workers' performance. However, the complexity of human behavior increases the difficulty of recognizing the influence of these stressors and finding a way to regulate workers' stress. This paper proposes the use of agent-based simulation techniques for addressing the challenge of analyzing workers' behavior and stress regulation pol icies. The main contributions of the paper are: (i) the definition of a stress model that takes into account work and ambient conditions to calculate the stress and the productivity of workers; (ii) the implementation of this model in an agent-based simulation system, enabling the analysis of workplace stress and productivity for different stress regulation policies; (iii) the analysis of four different stress regulation policies; and (iv) the validation of the model with a sensitivity analysis and with its application to a living lab.

Journal of Computational Science 51 (2021) 101326

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

An agent based simulation system for analyzing stress regulation policies at the workplace

Sergio Muñoz *, Carlos A. Iglesias

Intelligent Systems Group, Universidad Politécnica de Madrid, Spain

ARTICLE INFO

Keywords:
Stress
Agent-based simulation
Productivity
Stress modeling
Workplace stress

ABSTRACT

Workplace stress has a significant impact on productivity, since keeping workers' stress on an adequate level results a key factor for companies to increase their performance. While a high stress level may conduct to anxiety or absenteeism, a low level may also have undesirable consequences, such as lack of motivation. To identify and understand all the elements which interfere on workers' stress results a key factor in order to improve workers' performance. However, the complexity of human behavior increases the difficulty of recognizing the influence of these stressors and finding a way to regulate workers' stress. This paper proposes the use of agent-based simulation techniques for addressing the challenge of analyzing workers' behavior and stress regulation policies. The main contributions of the paper are: (i) the definition of a stress model that takes into account work and ambient conditions to calculate the stress and the productivity of workers; (ii) the implementation of this model in an agent-based simulation system, enabling the analysis of workplace stress and productivity for different stress regulation policies; (iii) the analysis of four different stress regulation policies; and (iv) the validation of the model with a sensitivity analysis and with its application to a living lab.

1. Introduction

Recent changes in working life have led to new challenges for organizations and employees, which may result in increasingly stressful working environments [1]. The economic and health costs of stressful work environments may be much greater than expected. The Fourth European Working Conditions Survey [2] stated that 22% of Europeans were suffering from stress and fatigue; and in a recent opinion poll, 51% of European workers considered that stress is common in their workplaces [3]. Also, it is estimated that work-related stress is the cause of about the 50% of all lost working days in European enterprises [4]. Similar results can be found for United States, where 40% of workers reported their job was very or extremely stressful and 29% of workers felt quite a bit or extremely stressed at work [5].

The huge impact of workplace stress on business has enhanced the interest and research on stress prevention. Work related stress can be prevented if the main provoking factors (or stressors) are identified. This could enable the design and implementation of regulation policies that decrease their adverse effects.

However, evaluating the effectiveness of these policies in a real scenario is challenging, mainly due to two reasons: the complexity of human behavior, and the costs that entails the implementation of stress detection and regulation techniques in a real scenario. These challenges could be addressed with the use of agent-based simulation. Agent-based simulation techniques have an established place in the analysis of human behavior [6], as they provide an approach to evaluate a large number of behavioral dimensions or variables without the need of implementing expensive systems. The use of agent-based simulation is particularly useful when there are complex interactions between agents, heterogeneous populations, and complexes behaviors [7]. One of the open challenges in agent-based systems is the difficulty of designing reliable agent behavior models, due to the high number of variables involved in human behavior [8], which are often hard to quantify, calibrate, and even justify. Focusing on the modeling of workplace stress, there are some approaches [9,10] in the literature which propose a model based on the working conditions (workload or available time).

This paper aims to advance the state of the art in human stress models, proposing a model that integrates not only working conditions, but also ambient. The model has been implemented in an agent-based simulation system, and evaluated by mean of a sensitivity analysis. In addition, a real experiment in a living lab has been carried out in order to validate the simulation results. The article also provides an open

* Corresponding author.
E-mail addresses: sergio.munoz@upm.es (S. Muñoz), carlosangel.iglesias@upm.es (C.A. Iglesias).

https://doi.org/10.1016/j.jocs.2021.101326
Received 22 May 2020; Received in revised form 29 December 2020; Accepted 6 February 2021
Available online 11 February 2021

source tool for studying and analyzing different stress regulation policies, measuring their effectiveness for regulating stress and improving workers' productivity. This tool has been used for carrying out an analysis of four different stress regulation policies, whose results are also presented.

The remainder of the article is structured as follows: Section 2 gives an overview of stress theories and current work on stress simulation models. Once existing models have been reviewed, the proposed scenario taken into account for the simulations is presented in Section 3. The definition and implementation of the proposed model is described in Section 4, while the results obtained for the proposed scenario are described in Section 5. Finally, an evaluation by mean of a sensitivity analysis and an experiment in a living lab is presented in Section 6, and Section 7 describes the conclusions drawn from this paper, as well as possible lines of future work.

2. Background

In this section, an overview of the leading theories about stress modeling is given. Then, current research on agent-based stress simulation is reviewed.

2.1. Stress theories

Stress is one of the most important emotions to analyze in the workplace context, as its adverse effects on an employee have a substantial impact on productivity loss [11]. Furthermore, the significant correlation between mental fatigue and impairment of physical performance in humans has been proved [12].

There is not a universally recognized definition for stress, since this notion is used in different contexts. According to the work of Cox and Griffiths [1,13,14], there are three different approaches to the definition of work-related stress: engineering, physiological and psychological. In the engineering approach, the stress is seen as a stimulus of the environment in the form of level of demand. Regarding the physiological, stress can be defined according to the changes that occur in a human under a stress state. Finally, from a psychological point of view, stress is defined as the dynamic process that occurs as an individual interacts with the environment.

In the same way that there are numerous definitions of stress, there are also numerous stress theories [15]. The Person-Environment (P-E) fit theory [16], based on the works by Lewin [17] and Murray [18], has been the source of other approaches to stress and well-being. This theory argues that stress arises not from the person or environment separately, but rather by the fit or congruence between them. It defines stress as a lack of match between their abilities (knowledge and skills) and the demands placed on them.

In 1982, Lazarus & Holroyd proposed the Transactional Model of Stress [19], that considers stress as a relationship between the person and the environment. This environment is appraised by the person as taxing, thus threatening well-being [20]. Lazarus & Holroyd distinguish two types of appraisal: primary and secondary. At the primary, a person acknowledges that there is something at stake; while at the secondary, individuals search for the benefit in a demanding encounter [21].

A different approach is the Job Demands-Control-Support model, proposed by Robert Karasek [22]. This model proposes an interaction where high demands and low control would predict high strain, but that high control would buffer the negative effect of demands on outcomes. The model has good predictive validity at the macro level, but it does not take into account individual differences in susceptibility to stressors. Consequently, it cannot explain the different behavioral or health outcomes under the same levels of demand and control in two individuals [23].

Another popular view of stress at work is the Effort-Reward Imbalance (ERI) model, proposed by Siegrist [24]. This model introduces some key features, as it emphasizes subjective perceptions of the environment. The central concept of the ERI model is that effort at work should be compensated by suitable rewards, and a mismatch between them will lead to stressful experiences.

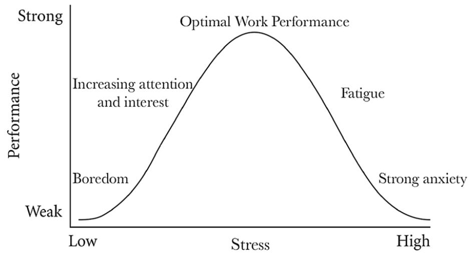


Fig. 1. The Yerkes-Dodson Inverted-U pressure performance curve [25].

With regards to the impact of stress on work performance, psychologists Robert Yerkes and John Dodson proposed the Yerkes-Dodson Inverted-U law, which is an empirical relationship between stress and performance in an attempt to obtain optimal efficiency from an individual [25]. Yerkes-Dodson Inverted-U law argues that individual performance increases with stress until the moment a limit value is reached, and decreases if this limit is exceeded. This law is often depicted as a normally distributed curve on a graph as seen in Fig. 1.

Fig. 1 shows how as the stress level increases, the individual performance also increases. The individual goes from a state of boredom to a state of increasing attention, until the moment in which arrives to the optimal state. From this level, if the stress continues to increase, the individual will enter in a fatigue state and even can arrive to a level of strong anxiety.

2.2. Agent-based simulation of stress

Based on the above theories, different agent-based simulation systems have been designed in order to study stress and its impact on productivity on several scenarios. These systems take into account job-related sources [9,26]. Page and Ashlock [9] proposes a way to calculate the stress factor as a relation between the work hours of an agent in a week (t), the maximum possible work a maximal stress agent will achieve (β) and the rate at which stress impacts an agent's performance (α). This relation is shown in Eq. (1). Then, the stress factor is used to calculate the performance:

$$S(t) = \beta + (1 - \beta)\frac{1}{1 + (\alpha t)^2} \tag{1}$$

Instead, the work carried out by Duggirala et al. [26] stems from the Silverman's model [27]. This model, which has been particularly defined for modeling human behavior in agents, states that stress is composed of three components: effective fatigue, time pressure, and event stress, as shown in Eq. (2):

$$\text{Stress} = \frac{\text{EF} + \text{ES} + \text{TP}}{3} \tag{2}$$

Based on this model, different agent-based simulation systems have been proposed for the analysis of different work-related policies. Harshal Hatatnagarkar et al. [28] analyze the implications of using a workload buffering strategy to manage work-related stress; Duggirala et al. [26] conclude that understanding the impact of stress can help managers to minimize the risk of crises while maintaining a team. All these solutions assume that all stress components contribute equally to stress.

Silverman's model integrates the classic Yerkes-Dodson Inverted-U model [25] of stimuli and performance with Janis and Mann's taxonomy [29] of coping strategies under stress, in order to develop a rich model describing the integrated stress and performance relationship. As

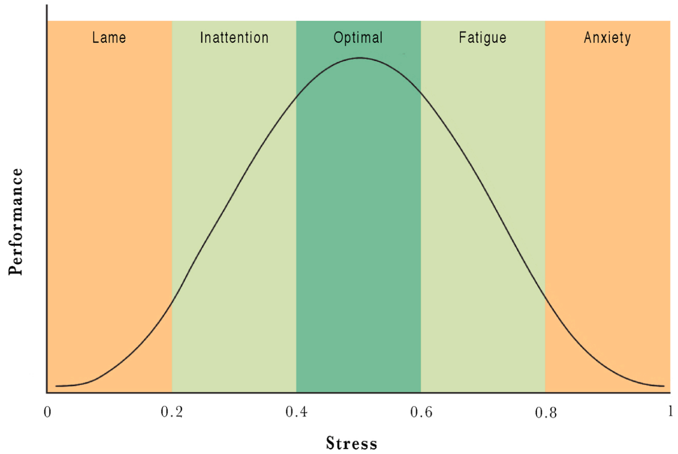


Fig. 2. Stress states.

commented in the previous section, the Inverted-U model shows how an optimal arousal level is required for effective performance: if arousal becomes too high, performance decreases.

Our work proposes to advance in the state of the art of agent-based simulation models for work-related stress by including ambient conditions, as well as other work conditions such as working overtime hours or the interruptions caused by electronic communication (email, messaging apps, etc.).

3. Scenario

To better understand the scope of the simulation system proposed, consider the following scenario. DOC is a highly skilled IT SME that detects that the stress level of the workers is too high, and this entails a decrease in their productivity and consequently in the company revenue. The company wants to identify the causes of the increase in the workers' stress level, and to evaluate several regulation policies. Implementing stress level detection techniques and regulation policies in the real scenario entails significant costs for the company, so they decide to address this challenge with agent-based simulation techniques.

With the purpose of analyzing and measuring the effectiveness of different stress regulation policies, the human resources department designs four different policies:

Baseline . It is used as the reference of the current method of operation in the office, and represents the situation to improve. There is not an adaptation nor log system to detect workers' state or ambient parameters. There are several equipment available to adapt the system, such as TV, lights, and Heating, ventilation and air conditioning (HVAC) system. These equipment are manually controlled by the workers, so ambient conditions such as noise, temperature or humidity may not have optimum values.

Ambient Adaptation . This strategy represents an environment where automation of ambient conditions control has been implemented. This guarantees that ambient conditions (such as temperature, humidity or noise) will always be near optimum values.

Workload Adaptation . The environment represented by this policy is an office where workload related tasks have been automated. This allows workers to configure rules that automate some of their daily tasks, which turns out in a relevant time saving and an increase in productivity.

Full Adaptation . This policy represents a whole automated office where there is an adaptation of both ambient and workload conditions.

Workers in DOC follow a schedule with the arriving time to work, the leaving time, the overtime hour limit, the free time and the sleep time. In this way, the activity of workers is divided into four intervals: working, working overtime hours, free time, and sleep time. In the *working* interval (from 9 a.m. to 5 p.m.), workers will work on the tasks that have been assigned to them and will perform electronic communication related tasks such as reading emails or messages applications. Sometimes they have finished all their tasks, and they can go home earlier. However, there are also days in which they do not complete these tasks before leaving time, so they have to work on them in the overtime interval. This overtime interval may last until 2 h each day. Once they have finished their tasks or have worked for a maximum of 2 overtime hours, they go home and rest for the remainder of the day.

Following the Inverted-U Theory [25], each worker can be in one of five states depending on the stress level: lame [0, 0.2), inattention [0.2, 0.4), optimal [0.4, 0.6), fatigue [0.6, 0.8) and anxiety [0.8, 1], as can be seen in Fig. 2.

A worker with high productivity or performance can carry out more tasks in a shorter period than an agent with a low level of performance. In that way, a worker with a no optimal level of stress may result in absenteeism, that refers to the unscheduled number leaves taken by a worker.

4. Agent-based simulation system

Once the scenario has been clarified, we proceed with the description of the simulation system design. In this section, an agent-based simulation system is proposed for the analysis of stress regulation policies at the workplace. Details about the proposed model and its implementation are given in the following subsections.

4.1. Model definition

The developed stress model is based on the Silverman's stress model

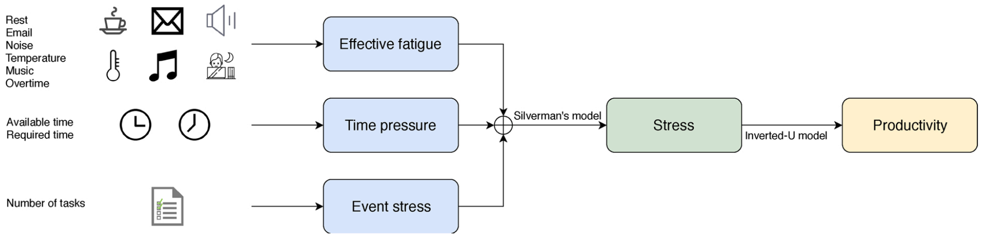


Fig. 3. Stress model.

[27], which argues that stress consists of three main components: *Event Stress* (*ES*), *Time Pressure* (*TP*), and *Effective Fatigue (EF)*. On the basis of this model, we have studied the influence of different ambient and work factors in order to extend it. The diagram shown in Fig. 3 represents an overview of the model design, showing the three components that compose stress along with the features that contribute to them.

As can be seen in the figure, several factors that affect to effective fatigue have been included. These factors are resting, electronic communication, noise, temperature, music and overtime hours, and their contribution to effective fatigue will be described in the following. The stress level is calculated following the Silverman's model from the effective fatigue, time pressure, and event stress. Finally, productivity is calculated from the stress level using the Inverted-U model.

In this work we are considering a workplace scenario in order to measure workers' stress and productivity levels. In this context, *Event Stress* can be attributed to the arrival of a high volume of work, *Time Pressure* to the relation between the time needed to complete the pending tasks in an agent's task queue and the remaining work time, and *Effective Fatigue* to the accumulated tiredness that results of the execution of work-related tasks and the impact of external conditions. Following the assumptions made by related works [26,28], we assume that all stress components have the same weight in the calculation of the total stress:

$$S = \frac{\mathrm{ES} + \mathrm{TP} + \mathrm{EF}}{3} \tag{3}$$

Silverman's stress model [27] states that all the elements that compound stress can vary between 0 and 1.0, with 0 being over-confident and unstressed, 0.5 neutral, and 1.0 representing the state of maximum stress.

Silverman proposes to calculate *Event Stress* as a relation between the number of tasks received in day and the mean number of tasks to which the worker is used to. This relation is shown in Eq. (4), where NT represents the Number of Task per day and MNT the Mean Number of Tasks per day:

$$\mathrm{ES} = \frac{\frac{\mathrm{NT}}{2}}{\mathrm{MNT}} \tag{4}$$

Therefore, if a worker whose average number of daily tasks is 20, one day receives 30, the contribution of ES will be 0.75. In the same way, if another day receives 10, the contribution will be 0.25.

Time Pressure derives from the relation between the required time to perform a task accurately and the available real-time for the task. Silverman represents this relation with Eq. (5), according to the work proposed by Hursh and McNally [30]. In this equation, T_i represents the ideal time required to perform a task and T_a represents the available time:

$$\mathrm{TP} = \frac{T_i}{(T_i + T_a)} \tag{5}$$

Hence, if an agent has pending tasks whose ideal performing time is 30 min, and only makes 20 min available, the stress effect of TP will be 0.6.

Lastly, *Effective Fatigue* consists of the contributions of several parameters called Performance Moderator Functions (PMF), that may be positive or negative stressors. Negative stressors are those which increase effective fatigue, and include ambient temperature and humidity, emails or messages received, sleep deprivation, noise, etc. Positive stressors, which decrease effective fatigue, may be relaxing music, video, rest time, etc. Specifically, in this work, the PMFs included will be overtime hours, temperature, humidity, noise, rest time, music and messages received. Silverman calculates EF as a normalized sum of all PMFs as shown in Eq. (6), where $WPFM_i$ represents the weight of the PMF, FT_i represents the fatigue tolerance of the agent to the PMF ($0 \le FT_i \le 1$), and N represents the number of PMFs:

$$\mathrm{EF} = \frac{1}{N}\sum_{i=1}^{N}\left(\overbrace{\frac{\mathrm{WPMF}_i}{\mathrm{WPMF}_i + \mathrm{FT}_i}}^{\text{negative stressors}} - \underbrace{\frac{\mathrm{WPMF}_i}{10}}_{\text{positive stressors}}\right) \tag{6}$$

In order to include the most relevant PMFs for the smart office context, the current literature about stress sources has been exhaustively analyzed. The drawn conclusions from this analysis are explained below.

An increase or a decrease of ambient temperature can result in stress for the worker, causing a low-performance capacity. The impact of temperature and humidity on stress can be determined using wet-bulb globe temperature (WBGT) [31]. WBGT is a type of apparent temperature used to estimate the effect of temperature and humidity on humans, whose simplified form, proposed by the Australian Bureau of Meteorology (ABM),[1] can be calculated from ambient temperature (T_a) and water vapor pressure (V_p), as shown in Eq. (7):

$$\mathrm{WBGT} = 0.567 \cdot T_a + 0.393 \cdot V_p + 3.94 \tag{7}$$

ABM also proposes V_p to be calculated as a function of the temperature and relative humidity (*RH*), as shown in Eq. (8). Relative humidity refers to the amount of the moisture in the air, compared to the potential saturation level. This equation is a

$$V_p = 6.105 \cdot \frac{\mathrm{RH}}{100} e^{\frac{17.27 \cdot T_a}{237.7 + T_a}} \tag{8}$$

Studies have demonstrated that the optimal level for the temperature factor is around 24 °C with relative humidity of 40% [32], resulting in a WBGT of 22.23 °C. In addition, based on [33] it has been estimated that stress level increases between 4 and 9 percent per degree on days when WBGT is above 27 °C. If the WBGT level is quite high or low, it will perceived as a negative stressor and will increase effective fatigue. However, if the WBGT level is in the ideal limits, between 20 and 25 °C, it will be perceived as a positive stressor, decreasing effective fatigue. One way of coping with a demanding work situation is to work longer hours. However, the relationship between fatigue and the number of work hours has been widely documented in past research, demonstrating that overtime and longer working hours result in higher levels of

[1] Australian Bureau of Meteorology: www.bom.gov.au.

Table 1
Performance moderator functions contribution to effective fatigue.

Feature	Contribution	Description	Source
Temperature and humidity	If WBGT $\leq 20\,^{\circ}$C or WBGT $\geq 25\,^{\circ}$C: $WPMF_t = 0.04 \cdot \lvert WBGT - 22 \rvert$	An increase in the difference between the current WBGT and the ideal WBGT (22 °C) will increase the effective fatigue	Somanathan et al. [33]
Overtime hours	$WPMF_o = 0.021 \cdot O_h$	Working overtime hours (O_h) will increase the effective fatigue	Singh et al. [34]
Rest time	$WPMF_r = 0.016 \cdot R_t$	Taking a break (R_t) will decrease effective fatigue	Dababneh et al. [35]
Noise	If $N_L \geq 65$ dB: $WPMF_n = 0.03 \cdot (N_L - 65)$	If the noise level (N_L) increases over a certain limit (65 dB), the effective fatigue will also increase	Maxwell [37]
Electronic communication	$WPMF_e = 0.0029 \cdot E_r$	The effective fatigue will increase with the number of emails or messages received (Er)	Kushlev et al. [42]

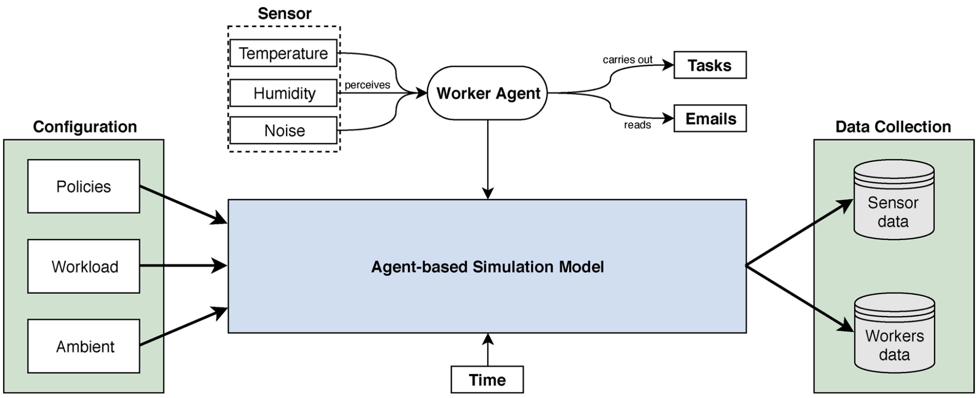


Fig. 4. Agent-based simulation system architecture.

fatigue, errors and lower productivity. Singh et al. [34] use a relation where fatigue level increases around 1% for each overtime working hour, so it will be considered a negative stressor.

Rest breaks have been proposed as a mean of reducing discomfort, indicating that short breaks may be beneficial for worker productivity and well-being at work [35]. Consequently, rest time will be considered a positive stressor that acts as the opposite to work overtime hours. Numerous research studies have confirmed noise as a primary cause of a reduction in workers productivity, as it is regarded as a source of distraction, frustration, and stress. Besides, it can contribute to stress and illness which, in turn, may produce absenteeism and turnover of staff [36]. Studies show that ideal noise level at work must be between 48 and 52 dBA, and a level greater than 65 dBA involves an important reduction of productivity [37], so high noise levels will be considered as a negative stressor. On the other side, listening to relaxing music before a stress task differently affects biological stress response domains, and helps to recover from a stressor more efficiently [38]. In this way, music will be considered a positive stressor.

Other aspects that affect stress are time management [39] and the pressure of constants notifications [40]. In particular, mail processing has been detected as one of the most stressful aspects in the workplace, as it involves an interruption of the current work. People who daily receive a high number of emails perceive email as a great source of stress, experiencing lower job satisfaction [41] and reporting greater work overload. Kushlev et al. proved that people experienced reduced stress when the number of times they checked their email was limited [42]. Also, email reception affects workers productivity as it involves an interruption of the tasks in which they were working, taking an average of 23 minutes to get back to the task [43]. In addition, nowadays social media and messaging apps have become popular in the workplace, and they also result a way of interruption due to switching contexts. Interruptions that disrupt concentration in a task are often detrimental, and switching tasks has been associated with higher stress [51]. Following these studies, we model the reception of email or other messages as electronic communication. With the purpose of designing the model as reliable as possible, the contributions of the explained features to workers stress have been estimated following the explained researches, as shown in Table 1.

The parameters defined in Table 1 influence stress, as they influence effective fatigue. Once the three components of stress are calculated, the stress can be determined. In addition to the calculation of the stress level of the worker, the rate at which this stress impacts worker's productivity results very interesting, as productivity will usually be the feature to maximize in a company. The integration of the Yerkes-Dodson Inverted-U model of stimuli and performance enables the description of the relation between stress and performance. The Inverted-U model shows that in order to achieve the maximum performance it is necessary to have an optimal arousal level. If the arousal level overcomes that optimal level or if it is not high enough, performance will decrease. Following the Inverted-U model [25], the productivity of an agent can be calculated as a function of the stress, as shown in Eq. (9), where S represents the stress:

$$P = \frac{1}{0.4 \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2}\left(\frac{S-0.5}{0.2}\right)^2} \tag{9}$$

Eq. (9) has been adjusted from the probability density function of a normal distribution $\mathcal{N}(0.5, 0.04)$. The parameters of the normal distribution have been chosen in order to have a function as similar as possible to the proposed by Yerkes-Dodson, limited between 0 and 1 (as the stress level). Productivity (P) will reach the maximum value of 1 when the stress level is 0.5 and will decrease as the value of stress moves away from its optimal value.

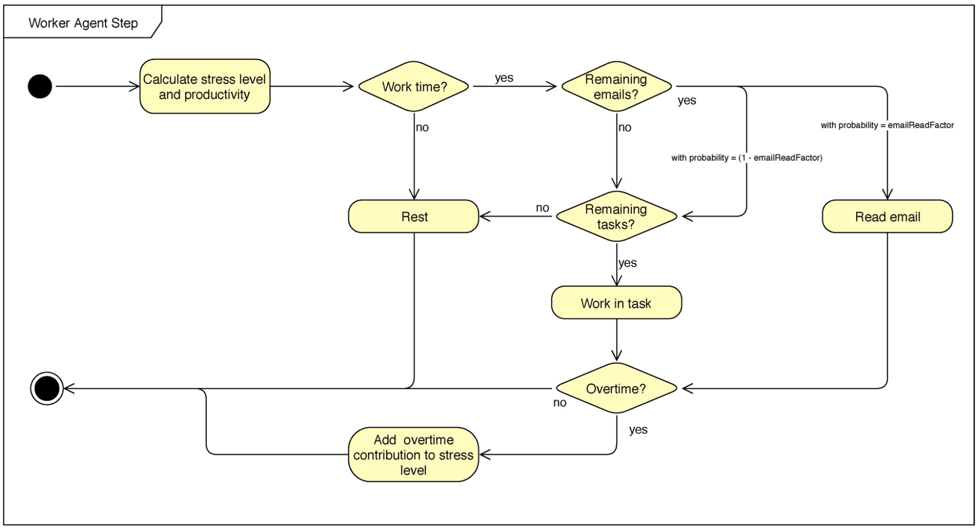


Fig. 5. Worker simulation step flowchart.

4.2. Model development

Once the theoretical aspects of the model have been explained, this subsection describes the development of the simulation system where implementing the simulation model. The simulation system has been designed with the primary purpose of enabling the representation of a smart office in a reliable way. It has been implemented with Python 3.5 using open source libraries. The core of the system has been developed using MESA [44], an open-source Python package that enables the creation of agent-based models using built-in core components or customized implementations. Furthermore, this framework enables the analysis of the results using Python's data analysis tools.

At a high level, the system consists of three main components: (1) a core that provides the simulation services; (2) a data storage component that manages the data generated by the simulation; and (3) a graphical user interface that allows the user to control the experiment execution (starting, pausing, stopping it); selecting parameters (work and ambient conditions, regulation policies, etc.); and executing strategies (number of workers, simulation duration, etc.). The core has been developed using MESA; which provides some base classes for agents, models, visualization and data collection. The user interface has been implemented using Tkinter library.[2] Finally, for the analysis and visualization of the data, Pandas and Matplotlib[3] have been used. In this way, the developed system allows the user to easily configure several aspects about the simulation using the graphical user interface. Once the required parameters have been chosen, the system will create the model and the agents, and the simulation will be run. The system will provide real time data via the graphical interface while the simulation is running, and once it has finished, it will store all the data for its later analysis and visualization.

The proposed system architecture is shown in Fig. 4. The architecture emphasizes on the modularization of the system and on the easy configuration of the different scenarios. The system structure depicted in Fig. 4 shows that the input-data is configurable from a separated module, and the output-data is stored in a repository. This enables an easy configuration of the different scenarios and regulation policies; as well as the data to be extracted from the simulation. In the following, the different modules of the system are described in detail.

Table 2
Objects properties.

Object	Properties
Task	Estimated time, remaining time, start time, assigned worker
Email	Reading time, remaining time, reception time, assigned worker
Sensor	Temperature, humidity, noise level
Time	Date, hour, is new day, is new hour, interval

The system makes use of two different classes for modeling the entities of the simulation: objects and agents. Tasks, time, sensors and emails/messages are modeled as objects as they do not behave as autonomous entities; while workers are modeled as agents.

Worker agent represents workers of the smart office environment that perform several actions such as carrying out tasks, reading messages or resting. In addition, these agents may suffer from stress when work or ambient conditions are not optimal, resulting in a decrease of their productivity. Fig. 5 shows a diagram with all the actions performed by this agent: at each step, the stress and productivity of the agent are calculated. Then, if it has remaining unread emails or messages, under a certain probability defined in the configuration it reads them during that step. Otherwise, in that step the agent works in tasks or rests depending on the hour and on the pending tasks associated to it.

Besides this agent, four objects have been developed for modeling tasks, messages, sensors, and time. These objects represent entities with properties that are relevant for the simulation, and their main properties are listed in Table 2.

- **Task object**. It models the assignments that each worker has to complete, and each one is assigned to a certain worker at a certain time. In addition, they have the estimated and the remaining time required to finish them. Each task has an assignment hour, and they are assigned to each worker along all day. All tasks have an estimated duration of 25 min, and the number of them that are commissioned to an agent in one day varies depending on the workload scenario.

[2] https://docs.python.org/3/library/tkinter.html.
[3] https://matplotlib.org.

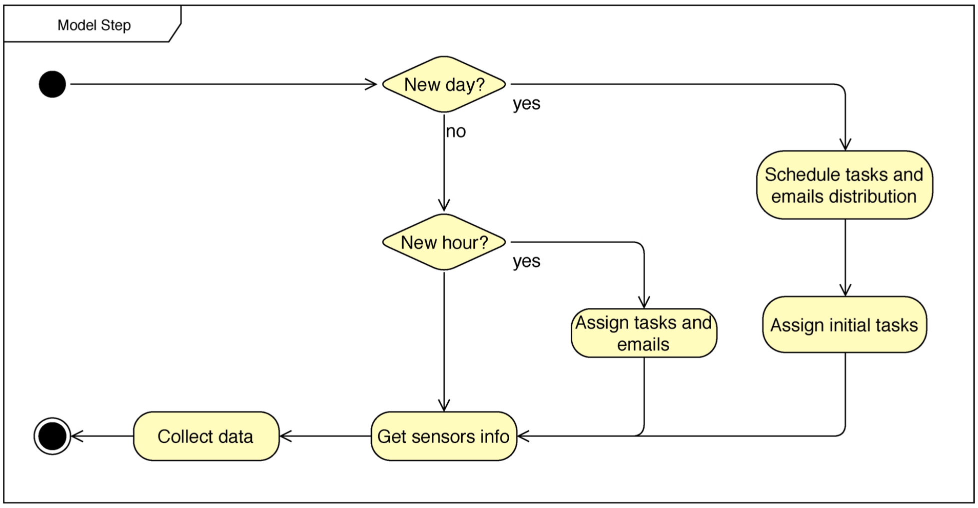


Fig. 6. Model simulation step flowchart.

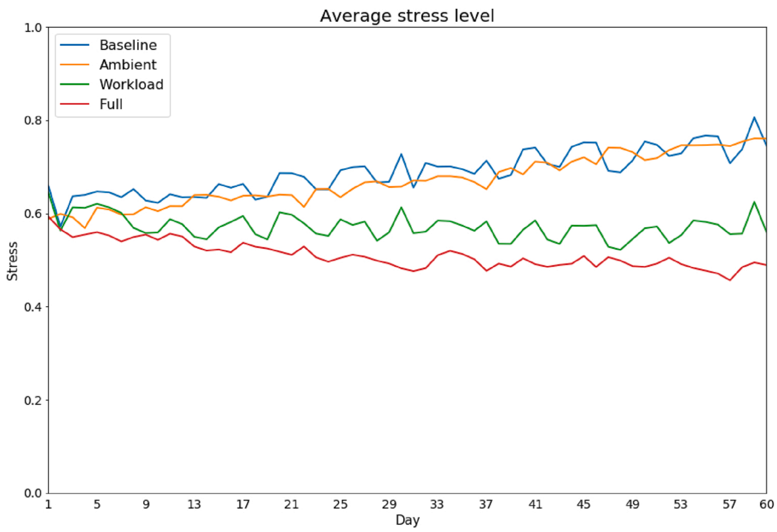


Fig. 7. Average stress level of workers for the different regulation policies.

- **Email object**. This object represents the reception of emails or other communication messages. Emails are very similar to tasks, they have the reception time and the estimated reading time for each one. The estimation of the number of emails read in a day is based on the work done by Kushlev and Dunn [42]. This work proposes that the number of emails read in a day follows a normal distribution with parameters $\mathcal{N}(12.54, 8.02)$. The email and message reception is distributed along the day for each worker. The time dedicated to an email by the agent (in minutes) is calculated in function of the time in reading the email (about 25 s), the time in answering the email and the time in getting back to the task; resulting in a normal distribution with parameters $\mathcal{N}(3, 0.5)$.
- **Sensor object**. Each sensor is responsible for measuring certain ambient conditions (e.g. temperature, humidity, noise), that are relevant to stress estimation. In order to provide the simulation with a greater reliability, the values for these variables are obtained from real values by mean of the open weather API[4] and the noise dataset generated by The National Institute for Occupational Safety and Health (NIOSH).[5] This dataset provides several noise levels in different office contexts.

[4] Weather API: https://www.apixu.com.
[5] Noise dataset: https://www.cdc.gov/niosh/data/datasets/rd-1005-2014-0/default.html.

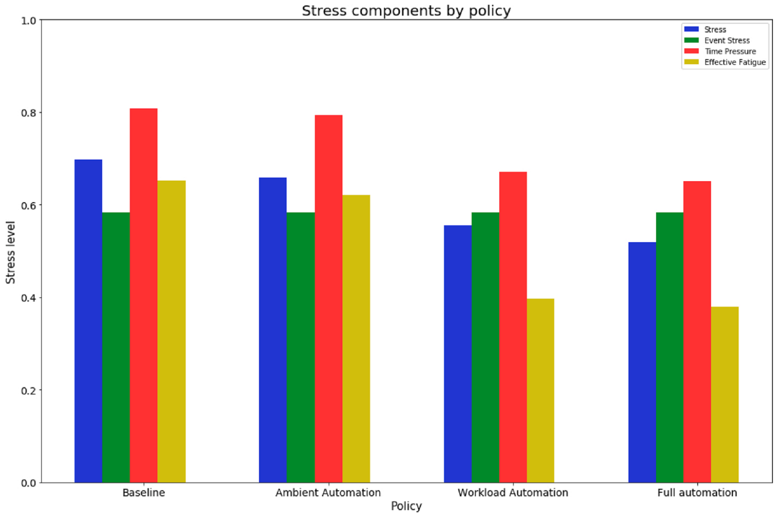


Fig. 8. Average stress components levels for each policy.

- **Time object**. It provides the simulation with a measuring of the time, converting the simulation steps to time. It has as properties the date and the hour, and methods for defining when a new day or a new hour starts. Finally, also enables the identification of the current interval of the simulation according to the time (working time, overtime or resting).

Once described the agent and objects, we proceed to describe the core of the simulation system: the agent-based simulation model. This module is responsible for managing and coordinating the simulation. It creates and control objects and worker agents, and manages all the changes on them. This module gets the simulation settings specified in the configuration (input-data), collects data during the simulation, and store these data at the repositories (output-data). The input-data include the work conditions (required time for each task, tasks arrival distribution, emails and messages reception distribution and the time spent in reading them, etc.); regulation policies; ambient conditions (temperature, humidity, and noise) and general parameters related to the simulation (number of workers, the equivalence in seconds of each step).

Fig. 6 describes the process performed by the model at each step during the simulation. At each step, the model checks if a new day has started. In this case, the model schedules the tasks and email reception distribution for each worker for that day. This defines the reception hour of each task and email. Then, it assigns the initial tasks to each worker.

Every hour, the model assigns tasks and emails to each worker according to the values defined in the scenario; measures the different global parameters (average stress, remaining tasks, etc.) and collects the data. This sequence will be repeated until the simulation is finished. At this moment, the model retrieves the generated data and stores it. The collected data are: stress level (with its three components), ambient data (temperature, humidity and noise), productivity, remaining and finished tasks, read and pending emails, and overtime hours worked.

5. Simulation results

In this section, the results of the simulations run for verifying the effectiveness of the different regulation policies described in Section 3 are presented.

The first metric to analyze is the stress level for each regulation policy. The team average level of stress along the two months of simulation is plot in Fig. 7. As shown in this figure, the workload automation policy has better results than the ambient automation. While for the baseline policy the average stress of workers considerably increases during weekdays, reaching values near to 0.8; the full automation policy achieves to stabilize the stress level of workers near to the optimum value. It is also interesting to appreciate how the ambient adaptation results effective for preventing peaks in the stress level, caused by peaks in the ambient conditions.

Table 3
Comparison between policies for productivity related parameters.

	Baseline	Ambient automation	Workload automation	Full automation
Average productivity (last day)	0.780469	0.793563	0.981176	0.986847
Average productivity (all days)	0.841826	0.870541	0.977049	0.986653
Average stress (last day)	0.753546	0.705455	0.529207	0.510618
Average stress (all days)	0.691424	0.662709	0.563505	0.534662
Average overtime hours	33.925708	33.175153	26.535347	25.945861
Average pending tasks	16.806667	15.369167	5.323333	4.810833

Analyzing the average values of the stress components for all workers during all the simulation, shown in Fig. 8, the effect of each policy can be appreciated. The figure depicts that while event stress remains stable for all policies, effective fatigue and time pressure are significantly decreased with the regulation policies.

These figures show that from the point of view of reducing stress, the implementation of an automation platform in the office is worth. The automation of both ambient and workload conditions controlling is the best policy to follow. However, seen that the workload automation has a significantly bigger impact that the ambient automation, the company could consider to implement only the former, depending on its budget. Moreover, the simulation also gives some relevant extracted parameters related to productivity, that are shown in Table 3.

The table shows different metrics obtained from the simulation for

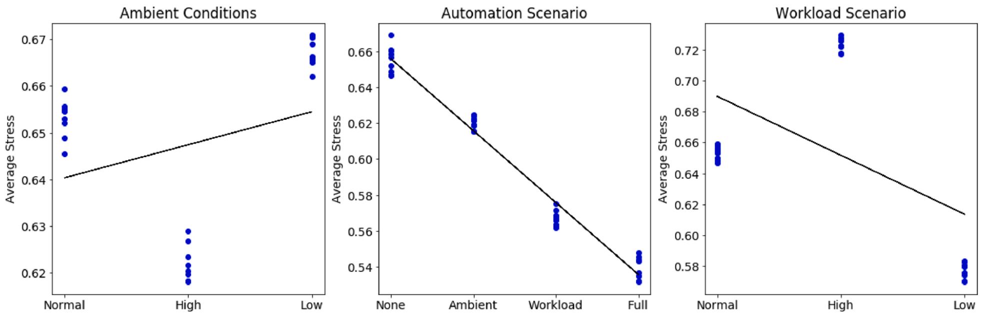


Fig. 9. Local analysis applied to the stress level.

the different regulation policies, as the average productivity and stress; the average number of unread emails or messages; and the average overtime hours worked; and average number of pending tasks for each worker. As shown, these parameters significantly improve with the application of regulation policies.

6. Validation

This section describes the conducted experiments that have been carried out with the purpose of validating the proposed model. The system has been validated using two different methods: a sensibility analysis and a real experiment in a living lab. The sensibility analysis has been done with the purpose of identifying the relationship between the involved variables and their effects on the problem; while the real experiment allows us to determine the reliability of the proposed model.

6.1. Sensibility analysis

The model has been evaluated using two different sensitivity analysis methods: local analysis approach and global analysis. The former studies small input perturbations on the model output, while the later the whole variation range of the inputs [45]. In order to analyze the influence of the different inputs of the system, several scenarios have been defined:

Ambient conditions : three different scenarios have been proposed (low, medium and high temperatures) obtaining temperature and humidity data from an open weather API for different months.

Automation : four scenarios have been defined, representing the four stress regulation policies previously described in Section 3.

Workload scenario : three different scenarios have been proposed, with different workload values (low, medium and high).

To bring about the One-at-Time (or local analysis) method, 100 simulations have been launched considering the different scenarios (10 for each proposed scenario). Fig. 9 shows the results of the local analysis for the average stress level.

As can be appreciated, all scenarios significantly influence stress levels. Changes in the ambient conditions have an impact of between 3% and 5% on average stress level. With regards to the automation scenarios, Fig. 9 shows that while ambient automation only decreases stress between 2% and 4%, workload automation decreases it around a 12%, and the combination of both policies achieves an improvement of near a 15%.

The global analysis has been carried out using Morris method [46], with the SALib library for Python [47]. This method enables the classification of the inputs in three groups: with negligible effects, with large linear effects without interactions and with large non-linear and/or interaction effects [48]. A total of 200 trajectories were built for the model, with the different scenarios described above. The Morris indices are detailed in Table 4. These indices show the influence of the different scenarios on stress level, and state that workload conditions are the variable that has a greater impact on it. Fig. 10 plots these results on a graph, enabling their visual analysis.

Table 4
Morris indices for the stress level.

Parameter	μ	μ^*	σ
Ambient conditions	0.0161784	0.0449908	0.0332322
Workload scenario	−0.0788225	0.167113	0.158649
Automation scenario	−0.112608	0.112608	0.0150309

The results of the Morris analysis show that the highest μ^* value is for the workload scenario, confirming that it is the most influencer input parameter for the stress. These results also show the importance of the automation scenario, with a μ^* value of 0.11. In the Fig. 10 we can also analyze the ratio σ/μ^*, that enables the characterization of the input parameters in terms of linearity and monotonicity. Both ambient and workload are non-linear and non-monotonic, while automation is non-linear but monotonic.

The presented methods attempt to validate that the influence that the different input variables have on the output is reasonable with the expected from the model design.

6.2. Experiment

As already stated, the main contribution of this work is the design of a stress model that takes into account ambient and workload conditions, and its implementation in an agent-based simulation system that enables the analysis of different stress regulation policies. For the analysis of these policies, three hypotheses have been raised:

- H1: The use of adaptation techniques helps to regulate stress level of workers who are under stressful conditions.
- H2: The proposed model is reliable and adequate.
- H3: The use of stress regulation policies improves workers productivity.

In order to evaluate the proposed system with respect to these hypotheses, an experiment with real users has been performed. For this experiment, a prototype of the proposed system has been deployed, which includes the following components. The stress of the participants is detected from biometrical signals (by mean of the Empatica E4 wristband) and a survey. Some actuators implement both hearing and visual signals using a variety of devices. Detailed information on materials is given in Section 6.2.2. This section covers the design, results,

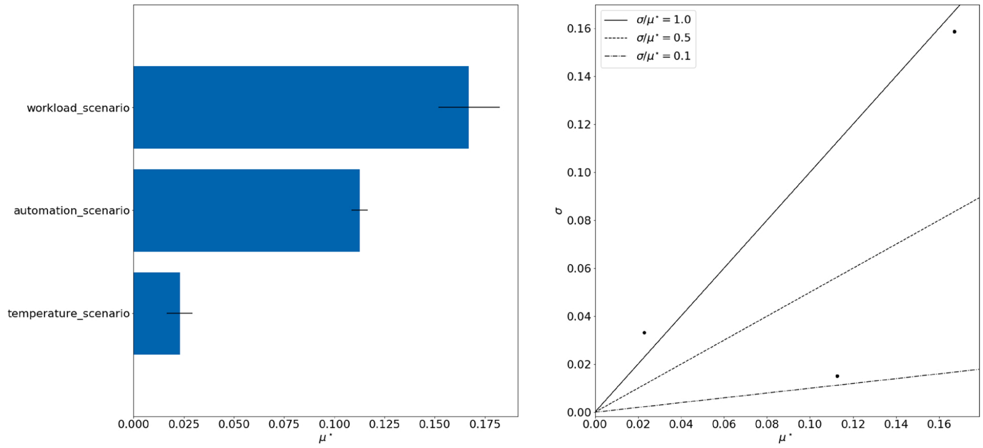


Fig. 10. Morris analysis applied to the stress level.

and conclusions drawn from the experiment, focusing on its scope.

6.2.1. Participants

The experiment included 20 participants. Their ages ranged from 18 to 28 years, all of them university students with a technical background, of both genders. Since the proposed system is primarily oriented to technical work positions, this selection is oriented to validate the system with participants that are currently working in technical environments or will do in the future. With this purpose, the selected participants were fellows in our research group or other research groups in the university. In this way, we tried to simulate the office conditions in order to make the scenario of the experiment as similar as possible to a real scenario. Participants were informed that they were performing an experiment, and provided with detailed information about the data collected in the experiment. However, they were not aware of the purpose of the study in order to exclude the influence of knowledge on the results. The ethical approval for the experiment was obtained by the Ethics Committee of the Technical University of Madrid.

6.2.2. Materials

The material used for this experiment is varied, as the proposed system needs several devices to properly function:

6 Empatica E4 wristband.[6] The E4 wristband is a wearable research device that offers real-time physiological data acquisition and software for in-depth analysis and visualization.
7 Room lighting (WS2812B LED strip controlled by WeMos ESP8266 board) that is used as an actuator on the light level of the room, with the possibility of using several lighting patterns.
8 Google Chromecast [49] that transmits content in a local computer network.
9 LG TV 49UJ651V. This device is used for displaying images and videos.
10 Google Home. The system uses this device for simulating office noise.

Participants accessed a Jupyter notebook using a computer with the Firefox browser.[7]

Table 5
Questions raised to the participants.

No.	Question formulation
Q1	From 0 to 10, what is your current stress level?
Q2	From 0 to 10, how stressful the number of pending tasks is?
Q3	From 0 to 10, how stressful is the remaining time available for finishing all tasks?
Q4	From 0 to 10, how stressful are current ambient and personal conditions?

6.2.3. Procedure

During the experiment, the participants have to complete a total of six tasks, which consist of programming problems obtained from Project Euler.[8] These tasks have an average duration of 5 min each one, so participants should be able to resolve all of them in 30 min. After the 30 minutes, participants could work on the tasks for other 7 min, in order to simulate overtime hours.

Two tasks are assigned to the participant at the beginning of the experiment, while the others are assigned at a certain hour, simulating the task arrival in a real workplace. When a participant finishes a task, he/she gives the response to the evaluator, who checks if the answer is correct and records the time that the participant has spent on the task. If the answer to the task is incorrect, it is given back to the participant. An example of a task is: *"If we list all the natural numbers below 10 that are multiples of 3 or 5, we get 3, 5, 6 and 9. The sum of these multiples is 23. Find the sum of all the multiples of 3 or 5 below 1000"*.

In addition, participants have to perform short but urgent tasks which consist of short questions that must be searched on the Internet, with an estimated duration of 1 minute. These short tasks simulate interruptions such as distractions or email reception. An example of a short task is: *"Which was the release date of the first version of Python?"*.

Also, a questionnaire is given to each participant in three times: at the beginning of the experiment, after 15 minutes and after 30 minutes. This questionnaire is intended to obtain the perceived stress level of each participant, and the raised questions are summarized in Table 5. The questions were designed following a 10-point scale. Different options were studied for the choice selection and, after analyzing the different options, we opted for a 10-point scale. Some of the benefits that we found in the use of a 10-point scale were better variance and better opportunity to detect changes [52]. In addition, the participants in the

[6] https://www.empatica.com/research/e4.
[7] https://www.mozilla.org/en-US/firefox/desktop/.
[8] https://projecteuler.net/.

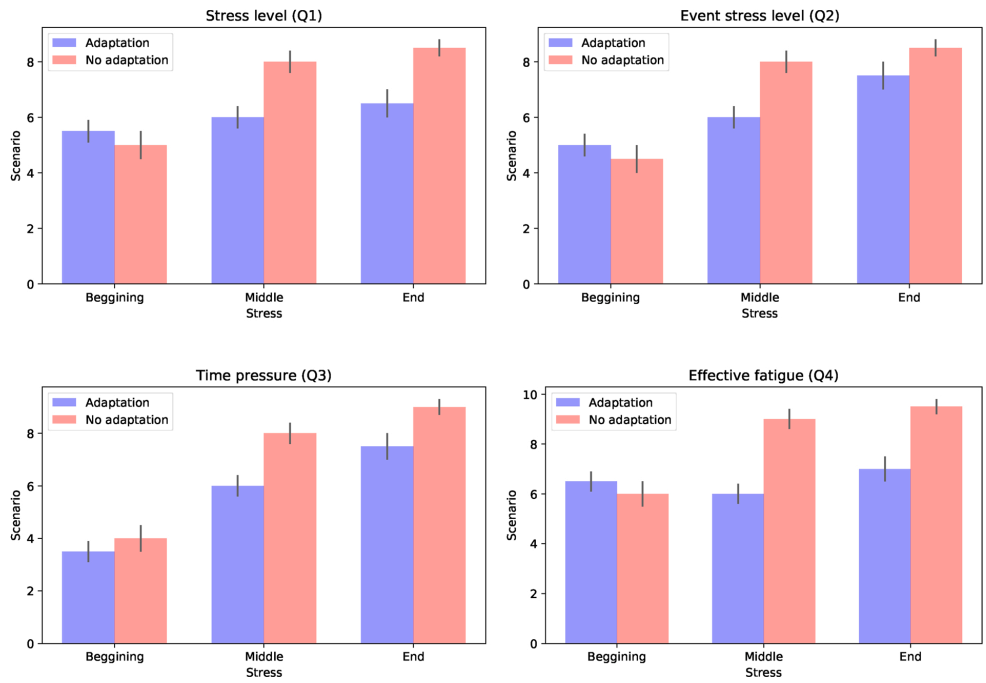


Fig. 11. Questionnaire results.

experiment are used to a 10-point scale, as it is the scale used for the grades in Spain schools and universities).

At the beginning of the experiment, the Empatica E4 wristband is given to each participant, in order to capture biometric data while they perform the tasks. This biometric data is used for analyzing the stress level of each participant.

The experiment finishes when the participant has completed all the tasks or when the 37 min (regular time of 30 min plus overtime) have finished.

6.2.4. Design

The experiment was a within-subject design. As previously stated, the controlling factor is the use of the adaptation techniques, which has two levels, activated and not activated. The automation use factor is counterbalanced using a Latin square so that the participants are divided into two groups: the control group and the treatment group. Participants in the control group performed the experiment without any kind of adaptation or automation; while participants in the treatment group performed the experiment with the adaptation system enabled. Both groups were formed by the same number of participants.

The workplace scenario consists of a small office, where participants have a computer with an opened Jupyter notebook o in the Firefox browser, where they must resolve the programming problems. In order to simulate office noise, an audio containing 2 hours of office noise[9] is played in the Google Home device. When the adaptation is activated, a relaxing video[10] is cast to the smart tv and the room lighting changes using lighting patterns with colored lights that slowly vary in intensity and color.

[9] Office noise: https://www.youtube.com/watch?v=D7ZZp8XuUTE.

[10] Relaxing video: https://www.youtube.com/watch?v=4inYSp-RWtU.

6.2.5. Results and discussion

In order to tackle the hypothesis, both biometric data and questionnaire results have been analyzed. The Empatica device provides BVP data extracted from a PPG (Photoplethysmography), the average heart rate (extracted from the BVP signal), and the time between individual heart-beats (IBI data, also extracted from the BVP data). Two features for detecting stress level can be obtained from these data: SDNN (standard deviation of heart rate intervals) and pNN50 (ratio of pairs of successive heart rate intervals that differ by 50 ms or less). On the other side, the questionnaire results provide the perceived stress of participants of the experiment during the performing of the tasks.

The results from the questionnaire show that the stress level in the adaptation part is lower than with no adaptation, as shown in Fig. 11. These results support H1, concluding that users feel more inclined to use the adaptation system rather than performing the task without adaptation. The analysis of the results of this question reveals that users feel more comfortable with the environment adaptation for the completion of the task.

With regards to the biometric analysis, SDNN and pNN50 results can be seen in Fig. 12. The following table, proposed by Madhuri et al. [50], can be helpful for analyzing the results of SDNN. As shown in Table 6, a bigger SDNN means a lower stress level, while a lower SDNN means a higher stress level. Our results shown in Fig. 12 show that in the three periods the SDNN has a greater value for the adaptation scenario. In the case of pRR50, we can see how its value significantly increases during the middle period for the no adaptation scenario, indicating an increase in stress level.

Finally, we present some quantitative data related to the experiment. The average number of tasks finished by the participants is 3.74 for the adaptation scenario, and 2.87 without adaptation; the average time dedicated to each task is of 9.8 min for the adaptation scenario and 12.8 for the no adaptation scenario; and the minimum time required to finish a task was of 4 min for the adaptation scenario and 6 for the no

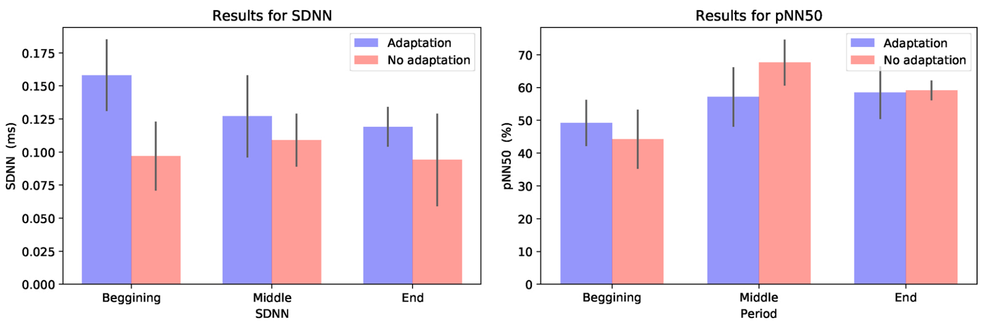


Fig. 12. Biometric analysis results.

Table 6
Stress level from SDNN [50].

SDNN (ms)	Stress level
25–55	Highly tense
55–110	Slightly tense
110–180	Mildly calm
180–215	Quietly relaxed
215–230	Deeply relaxed

adaptation scenario.

These results support H1 in the same way that the questionnaire; and also H2, indicating that users' productivity is improved with the use of stress regulation policies. In addition, these results are adequate with the expected from the simulation results, confirming that the use of agent-based simulation can help to identify stress causes, design stress regulation policies, and thus, to prevent stress.

7. Conclusions

In this work, an agent-based simulation system to analyze stress regulation policies at the workplace has been developed. This system implements a stress model also proposed in this paper, that takes into account both workload and ambient conditions for determining stress and productivity levels of workers. The proposed system enables the simulation of a smart office in order to analyze the evolution of workers' stress and productivity over time; and allows users to configure custom adaptation policies in an easy way.

In addition, this paper also shows the result for the analysis of four different stress regulation policies, measuring the influence of different adaptation levels on stress and productivity. These results show that ambient conditions adaptation does not result particularly effective for regulating stress at the workplace by itself, but combined with workload adaptation, it achieves a decrease of a 10% in workers' stress. The sensibility analysis performed to the model confirms this result, showing that between all the variables, the workload condition is the one that affects in a more significant way to stress and productivity. Finally, the results of the experiment in a living lab validates these results. The answers to the self-reported questionnaires demonstrate the positive effect of the adaptation for regulating stress, and also the results obtained from the analysis of the biometric data show that stress levels are lower when there is adaptation.

We are currently exploring some lines of future work that stem from this work. Firstly, we aim at integrating different personalities for the agents that represent the workers, with different ways of coping with the workload and ambient conditions. Also different ways of coping with tasks or messages reception depending on the agents' mood could be studied. Secondly, another possible line of future work that we are currently investigating is the integration of room offices and locations in the model, in order to analyze the stress contagion between workers who work together. Finally, the third line of future work is to improve the validation method with more experiments, increasing the number of participants, and also including a pairwise comparison method in order to improve the accuracy. These advances would give more reliability to the model.

Declaration of Competing Interest

The authors report no declarations of interest.

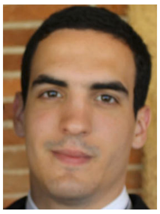
Acknowledgements

This work is supported by the Spanish Ministry of Economy and Competitiveness under the R&D project SEMOLA (TEC2015-68284-R).

References

[1] T. Cox, A. Griffiths, Assessment of potentially stressful work and organisational environments: theory and method, in: Evaluation of Human Work, 4th ed., CRC Press, 2015, pp. 565–582 (Chapter 13).
[2] A. Parent-Thirion, E. Fernández-Macías, J. Hurley, G. Vermeylen, Fourth European Working Conditions Survey, Tech. Rep. European Foundation for the Improvement of Living and Working Conditions (2007).
[3] E. A. for Safety, H. at Work, European Opinion Poll on Occupational Safety and Health, Tech. Rep., European Agency for Safety and Health at Work, 2013.
[4] E. A. for Safety, H. at Work, Campaign Guide: Managing Stress and Psychosocial Risks at Work, Tech. Rep., European Agency for Safety and Health at Work, 2013.
[5] S. Saute, L. Murphy, M. Colligan, N. Swanson, J. Hurrell, F. Scharf, R. Sinclair, P. Grubb, L. Goldenhar, T. Alterman, J. Johnston, A. Hamilton, J. Tisdale, Stress at Work, Tech. Rep., National Institute for Occupational Safety and Health, 1999.
[6] P. Sobkowicz, Agent based model of effects of task allocation strategies in flat organizations, Phys. A: Stat. Mech. Appl. 458 (2016) 17–30.
[7] E. Bonabeau, Agent-based modeling: methods and techniques for simulating human systems, Proc. Natl. Acad. Sci. U.S.A. 99 (Suppl. 3) (2002) 7280–7287.
[8] S. Elkosantini, Toward a new generic behavior model for human centered system simulation, Simul. Modell. Pract. Theory 52 (2015) 108–122.
[9] M. Page, D. Ashlock, Stress and productivity performance in the workforce modelled with binary decision automata, 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (2015) 1–8.
[10] M. Duggirala, M. Singh, H. Hayatnagarkar, S. Patel, V. Balaraman, Understanding impact of stress on workplace outcomes using an agent based simulation, in: Proceedings of the Summer Computer Simulation Conference, SCSC'16, Society for Computer Simulation International, San Diego, CA, USA, 2016, pp. 35:1–35:10.
[11] S. Lupien, B. McEwen, M. Gunnar, C. Heim, Effects of stress throughout the lifespan on the brain, behaviour and cognition, Nat. Rev. Neurosci. 10 (6) (2009) 434–445.
[12] S. Marcora, W. Staiano, V. Manning, Mental fatigue impairs physical performance in humans, J. Appl. Psychol. 106 (2009) 857–864.
[13] T. Cox, A. Griffiths, The Nature and Measurement of Work Stress: Theory and Practice, The Evaluation of Human Work: A Practical Ergonomics Methodology, Taylor & Francis, London, 2005.
[14] T. Cox, A. Griffiths, Work-related stress: nature and assessment, IEE Colloquium on Stress and Mistake-Making in the Operational Workplace (1995), 1/1–1/4.
[15] P.J. Dewe, M.P. O'Driscoll, C.L. Cooper, Theories of psychological stress at work, Handbook of Occupational Health and Wellness (2012) 23–38.
[16] J.R. French, R.D. Caplan, R. Van Harrison, The Mechanisms of Job Stress and Strain, vol. 7, J. Wiley, Chichester, 1982.
[17] K. Lewin, A dynamic theory of personality, J. Hered. 27 (11) (1936) 441–442.

[18] H.A. Murray, Explorations in Personality, Oxford Univ. Press, 1938.
[19] K.A. Holroyd, R.S. Lazarus, Stress, coping and somatic adaptation. Handbook of Stress: Theoretical and Clinical Aspects, 1982, pp. 21–35.
[20] K. Glanz, B.K. Rimer, K. Viswanath, Health Behavior and Health Education: Theory, Research, and Practice, John Wiley & Sons, 2008.
[21] R.S. Lazarus, Stress and Emotion: A New Synthesis, Springer Publishing Company, 2006.
[22] T.T. Robert Karasek, Healthy work: stress, productivity, and the reconstruction of working life, Quart. Rev. Biol. 66 (4) (1990) 525–526.
[23] P.L. Perrewé, K.L. Zellars, An examination of attributions and emotions in the transactional approach to the organizational stress process, J. Organ. Behav. 20 (5) (1999) 739–752.
[24] J. Siegrist, Adverse health effects of high-effort/low-reward conditions, J. Occup. Health Psychol. 1 (1) (1996) 27.
[25] R. Yerkes, J. Dodson, The relation of strength of stimulus to rapidity of habit-formation, J. Comp. Neurol. Psychol. 18 (1908) 459–482.
[26] M. Duggirala, M. Singh, H. Hayatnagarkar, S. Patel, V. Balaraman, Understanding impact of stress on workplace outcomes using an agent based simulation, in: Proceedings of the Summer Computer Simulation Conference, Society for Computer Simulation International, 2016, p. 35.
[27] B.G. Silverman, More Realistic Human Behavior Models for Agents in Virtual Worlds: Emotion, Stress, and Value Ontologies, Center for Human Modeling and Simulation, 2001, p. 34.
[28] H. Hayatnagarkar, M. Singh, S. Kumar, M. Duggirala, V. Balaraman, Can a buffering strategy reduce workload related stress? in: L. Zhang, X. Song, Y. Wu (Eds.), Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems Springer Singapore, Singapore, 2016, pp. 378–388.
[29] I.L. Janis, L. Mann, Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment, Free Press, 1977.
[30] S. Hursh, R. McNally, Modeling Human Performance to Predict Unit Effectiveness, Army Science: The New Frontiers, Military and Civilian Applications, Borg Biomedical Books, Saratoga, WY, 1993, pp. 309–328.
[31] C.P. Yaglou, D. Minaed, Control of heat casualties at military training centers, Arch. Indust. Health 16 (4) (1957) 302–316.
[32] A. Rasdan, C. Mohammad, M. Hanifiah, M. Haniff, The impact of workers productivity under simulated environmental factor by Taguchi analysis, Proc. Soc. Behav. Sci. 10 (2014) 263–268.
[33] E. Somanathan, R. Somanathan, A. Sudarshan, M. Tewari, et al., The Impact of Temperature on Productivity and Labor Supply: Evidence from Indian Manufacturing, Tech. Rep., Indian Statistical Institute, New Delhi, India, 2015.
[34] M. Singh, M. Duggirala, H. Hayatnagarkar, V. Balaraman, A multi-agent model of workgroup behaviour in an enterprise using a compositional approach, CEUR Workshop Proceedings 1561 (ModSym) (2016) 10–15.
[35] A.J. Dababneh, N. Swanson, R.L. Shell, Impact of added rest breaks on the productivity and well being of workers, Ergonomics 44 (2) (2001) 164–174.
[36] D. Abbott, Calming the office cacophony, The Safety & Health Practitioner 22 (1) (2004) 34–36.
[37] L.E. Maxwell, Noise in the Office Workplace, Department of Design and Environmental Analysis, College of Human Ecology, Cornell University, 1999.
[38] M.V. Thoma, R.L. Marca, R. Brönnimann, L. Finkel, U. Ehlert, M. Urs, The effect of music on the human stress response, PLOS ONE 8 (8) (2013) 1–12.
[39] T.H. Macan, C. Shahani, R.L. Dipboye, A.P. Phillips, College students' time management: correlations with academic performance and stress, J. Educ. Psychol. 82 (4) (1990) 760.
[40] G. Mark, D. Gudith, U. Klocke, The cost of interrupted work: more speed and stress, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2008) 107–110.
[41] A.C. Jerejian, C. Reid, C.S. Rees, The contribution of email volume, email management strategies and propensity to worry in predicting email stress among academics, Comput. Human Behav. 29 (3) (2013) 991–996.
[42] K. Kushlev, E.W. Dunn, Checking email less frequently reduces stress, Comput. Human Behav. 43 (2015) 220–228.
[43] K. Pattison, Worker, Interrupted: The Cost of Task Switching, Fast Company, 2008 (Accessed 11 November 2019), https://www.fastcompany.com/944128/worker-interrupted-cost-task-switching.
[44] D. Masad, J. Kazil, Mesa: an agent-based modeling framework, Proceedings of the 14th Python in Science Conference (SciPy 2015) (2015) 53–61.
[45] B. Iooss, P. Lemaître, A review on global sensitivity analysis methods, Uncertainty Management in Simulation-Optimization Of Complex Systems (2015) 101–122.
[46] M.D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (2) (1991) 161–174.
[47] J. Herman, W. Usher, Salib: an open-source python library for sensitivity analysis, J. Open Source Softw. 9 (2017).
[48] B. Iooss, P. Lemaitre, A review on global sensitivity analysis methods, in: Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications, Springer US, Boston, MA, 2015, pp. 101–122.
[49] K. Williams, The technology ecosystem: fueling Google's chromecast, IEEE Women Eng. Mag. 8 (1) (2014) 30–32.
[50] V. Madhuri, M.R. Mohan, R. Kaavya, Stress management using artificial intelligence, 2013 Third International Conference on Advances in Computing and Communications, IEEE (2013) 54–57.
[51] Duncan P. Brumby, Christian P. Janssen, Gloria Mark, How do interruptions affect productivity? Rethinking Productivity. Software Engineering, Apress, Berkeley, CA, 2019, pp. 85–107.
[52] James Carifio, Rocco J. Perla, Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes, J. Soc. Sci. 3 (3) (2007) 106–116.



Sergio Muñoz received the graduate and master's degrees in Telecommunications Engineering from the Technical University of Madrid (Universidad Politécnica de Madrid), Spain, in 2016 and 2017, respectively, where he is currently pursuing the Ph.D. degree. He is currently a Teaching Assistant with the Technical University of Madrid. His research interests include ambient intelligence and agent-based simulation. The main topic of his thesis is the adaptation of smart environments to users' emotions, to enhance well-being and performance.

Carlos A. Iglesias received the Telecommunications Engineering degree and the Ph.D. degree in telecommunications from the Universidad Politécnica de Madrid (UPM), Spain, in 1993 and 1998, respectively. He is currently an Associate Professor with the Telecommunications Engineering School, UPM. He has been leading the Intelligent Systems Group, UPM, since 2014. He has been the Principal Investigator on numerous research grants and contracts in the field of advanced social and the IoT systems, funded by the regional, national, and European bodies. His main research interests include social computing, multiagent systems, information retrieval, sentiment and emotion analysis, linked data, and web engineering.

3.2.5 An Emotion-Aware Learning Analytics System Based on Semantic Task Automation

Title	An Emotion-Aware Learning Analytics System Based on Semantic Task Automation
Authors	Muñoz, Sergio and Sánchez, Enrique and Iglesias, Carlos A.
Journal	Electronics
Impact factor	JCR 2021 2.690
ISSN	2079-9292
Volume	9
Year	2020
Keywords	automation
Pages	1194
Online	https://www.mdpi.com/2079-9292/9/8/1194
Abstract	E-learning has become a critical factor in the academic environment due to the endless number of possibilities that it opens for the learning context. However, these platforms often suppose to increase the difficulties for the communication between teachers and students. Without having real contact between teachers and students, the former finds it harder to adapt their methods and content to their students, while the students also find complications for maintaining their focus. This paper aims to address this challenge with the use of emotion and engagement recognition techniques. We propose an emotion-aware e-learning platform architecture that recognizes students' emotions and attention in order to improve their academic performance. The system integrates a semantic task automation system that allows users to easily create and configure their own automation rules to adapt the study environment. The main contributions of this paper are: (1) the design of an emotion-aware learning analytics architecture; (2) the integration of this architecture in a semantic task automation platform; and (3) the validation of the use of emotion recognition in the e-learning platform using partial least squares structural equation modeling (PLS-SEM) methodology.

electronics

MDPI

Article

An Emotion-Aware Learning Analytics System Based on Semantic Task Automation

Sergio Muñoz *, Enrique Sánchez and Carlos A. Iglesias

Intelligent Systems Group, Universidad Politécnica de Madrid, 28040 Madrid, Spain; enriquetolby@gmail.com (E.S.); carlosangel.iglesias@upm.es (C.A.I.)
* Correspondence: sergio.munoz@upm.es

Received: 6 July 2020; Accepted: 22 July 2020; Published: 25 July 2020

check for updates

Abstract: E-learning has become a critical factor in the academic environment due to the endless number of possibilities that it opens for the learning context. However, these platforms often suppose to increase the difficulties for the communication between teachers and students. Without having real contact between teachers and students, the former finds it harder to adapt their methods and content to their students, while the students also find complications for maintaining their focus. This paper aims to address this challenge with the use of emotion and engagement recognition techniques. We propose an emotion-aware e-learning platform architecture that recognizes students' emotions and attention in order to improve their academic performance. The system integrates a semantic task automation system that allows users to easily create and configure their own automation rules to adapt the study environment. The main contributions of this paper are: (1) the design of an emotion-aware learning analytics architecture; (2) the integration of this architecture in a semantic task automation platform; and (3) the validation of the use of emotion recognition in the e-learning platform using partial least squares structural equation modeling (PLS-SEM) methodology.

Keywords: e-learning; emotion aware; semantic; task automation; emotion detection; emotion regulation

1. Introduction

Nowadays, e-learning [1] is a key factor in the learning context, due to the great number of advantages and possibilities that it entails in different academic environments (i.e., distance learning, self-directed learning, or face-to-face learning) [2]. For this reason, it has inspired the research and development efforts of numerous companies and universities that seek to improve different learning methods to obtain a better academic performance [3]. However, these platforms also entail specific challenges [4]. One of these challenges is that maintaining the focus and motivation becomes difficult for certain students [5]. In addition, these platforms also introduce a handicap in the communication between teacher and students [6]. In this kind of learning, teachers find it hard to know the students and to check whether the methods and contents are appropriate for them.

This is particularly important in recent years, where the analysis of the role that emotions play in academic environments is gaining momentum [7]. In this context, two technologies play a key role: emotion-aware systems [8] and learning analytics [9] techniques. Emotion-aware systems, based on cognitive computing [10], are capable of detecting human emotions through techniques such as image recognition or speech analysis. By using these systems, human emotions can be detected on a larger scale, with better results, and reducing human–machine interactions. Learning analytics systems are fed with data from the learning environment to conduct a comprehensive analysis using machine learning [11] techniques and algorithms. Moreover, these systems can be improved thanks to a better understanding of the relevance and relationship between the different factors that appear in the learning process. However, interoperability between different platforms and components has become a key issue in these systems [12]. The vast amount of generated data requires highly agile data models to make better decisions or any other actionable results. These models enable monitoring, processing, optimizing, and analyzing data to gain insights [13]. Semantic technologies propose a suitable approach for interoperability by sharing common vocabularies, enabling the interoperable representation of inferred data.

In order to address the above challenges and advance in the state of the art e-learning platforms, we propose an emotion-aware system with learning analytics capabilities integrated into a semantic task automation platform. The main goal is to improve the learning process by including emotion and attention recognition techniques [14]. The proposed system recognizes students' emotions and engagement, allowing educators to take into account this information for adapting their content and methodologies. Furthermore, taking advance of the growing emergence of intelligent devices, the platform also integrates ambient intelligence technologies for adapting the environment to the students' mood [15]. This is accomplished by means of the integration of an emotion-aware task automation platform called EWETasker, developed in a previous work [16]. This platform offers an easy way to integrate smart devices and services in an environment and use them for automating daily tasks. In this way, the design and development of an emotion aware architecture for e-learning platforms is the main contribution of this paper. In addition, a semantic task automation platform has been integrated into our system in order to provide it with mood and engagement regulation by the use of ambient intelligence techniques. Finally, the usefulness of an e-learning platform that takes into account students' emotions and attention has been validated using partial least squares structural equation modeling (PLS-SEM) methodology [17].

The rest of this paper is organized as follows. In the first place, an overview of the related work in emotion aware e-learning systems and learning analytics is given in Section 2. Then, Section 3 describes the reference architecture of the proposed emotional aware e-learning system, describing the main components and modules as well as their implementation. Section 4 describes the case studies defined for the proposed architecture. Validation of the use of an emotion-aware e-learning system for improving academic performance is described in Section 5. Finally, the conclusions drawn from this work, along with possible future work lines, are described in Section 6.

2. Background

This section describes the background and related work for the system proposed in this paper. First, Section 2.1 describes the motivation and objectives in the study of emotions related to the learning process. Then, an overview of the most used emotion recognition tools is given in Section 2.2. Section 2.3 presents the related works in emotion aware e-learning platforms. Finally, Section 2.4 gives an overview of the state-of-the-art regarding learning analytics.

2.1. Emotion Theories Applied to Education

Over the last years, emotions in the learning process have become an important field of study by psychologists. The influence that emotions have on the way people act [18] is known. In the academic context, the connection between emotions and encouragement has been proved [19]. Understanding and managing emotions result in a key factor in improving academic results due to the close relationship between attention and performance [20]. The individual characteristics of each student produce differences in cognition, motivation, and learning styles [21], and consequently on their learning outcomes [22].

Pekrun et al. [23] proposed a model on how emotions influence students' learning and achievement. This model studies the effect of mood on performance and how cognitive and motivational mechanisms may mediate it. They state that the influence of emotions can be mediated by several mechanisms with cumulative or contradictory effects, in order to predict overall effects on performance supposes a challenge. The study of Ranellucci et al. [24] provided empirical support for this model, evaluating the relations between achievement goals, emotions, learning strategies, and performance; and showing both direct and indirect effects of students' goals on academic achievement through their emotions and learning strategies. This study presented the benefits of mastery-approach goals for students' emotions and demonstrated enjoyment was beneficial for most learning strategies, while boredom predicted poorer time management, and anxiety predicted lower self-monitoring. In addition, performance-approach goals predicted less critical thinking, and performance-avoidance goals predicted more anxiety, boredom, critical thinking, and lower achievement gains.

In addition, an affective model for education was proposed by Kort et al. [14]. This model argues that a typical learning experience involves a range of positive and negative emotions. According to the authors, this range of emotions occurs naturally and is an inevitable part of the learning cycle, so it is important to help students understand and recognize this process and to propel themselves. Social emotional intelligence and competence development literature are linked in the Social Emotional Development (SED) model, proposed by Seal et al. [25]. This model includes self awareness, consideration of others, connection to others, and impacting change. Furthermore, the work aims to provide a framework to understand and facilitate increasing student social and emotional capacity to recognize emotional cues, process emotional information, and utilize emotional knowledge to adapt to social challenges in higher education. In addition, the work by Frenzel et al. [26] argue the existence of positive reciprocal links between teachers' and students' enjoyment, and that these links are mediated by the observations of each other's classroom behaviors.

Following these theories, recognizing students' emotions seems to be essential to stimulate them in a learning process. However, despite the importance of recognizing students' emotions in e-learning platforms, there are no many real implementations of these platforms that include emotion recognition. The inclusion of this feature can multiply the benefits of distance learning and add value to the platforms in charge of providing this service. With this purpose, it is essential to know how emotion awareness can be carried out.

2.2. Emotion Recognition Tools

Feidakis [8] defines emotion awareness as the implicit or explicit collection of emotion data and the recognition of emotional patterns. This definition is carefully explained by the author distinguishing three types of emotion captures: explicit (subjective report of feelings), implicit (voice, face, and physiological signals), and with emotion patterns recognition (identification of the people emotions based on observation).

The first steps were carried out by expert psychologists who identified the emotions of the patient based on their body posture, tone of voice, or facial expression [27]. Then, self-reporting tools became widely used because of their simplicity and low cost, apart from being a non-intrusive tool [28]. However, there are also some disadvantages to be taken into account, such as lack of honesty among participants, misinterpretation of the questions, or the use of unadjusted rating scales. The traditional approach of detecting emotions through questionnaires answered by the participants does not yield very efficient methods [29]. That is the reason for focusing on automatic emotion detection using multimodal approaches (i.e., facial recognition, speech analysis, and biometric data) as the ensemble of different information sources from the same mode [30].

In recent years, improvements in the field of Artificial Intelligence [10] have enabled the same analysis of emotions using emotion recognition devices and without human intervention. They take advantage of being non-invasive and not requiring specialized equipment since most of them use common devices such as webcams or microphones. Algorithms to predict emotions based on facial expressions are mature and considered accurate. Currently, there are two main techniques to realize facial expression recognition depending on its way of extracting feature data: appearance-based features, or geometry-based features [31]. Both techniques have in common the extraction of some features from the images which are fed into a classification system. They differ mainly in the features extracted from the video images and the classification algorithm used [32]. Geometric based techniques find specific features such as the corners of the mouth or eyebrows and extract emotional data from them. Otherwise, appearance-based extraction techniques describe the texture of the face caused by expressions and extract emotional data from skin changes [33]. Emotion recognition from speech analysis is an area that is gaining momentum in recent years [34]. Speech features are divided into four main categories: continuous features (pitch, energy, formants), qualitative features (voice quality, harsh, breathy), spectral features (linear predictive codes, Mel frequency cepstral coefficients), and Teager energy operator-based features such as TEO-decomposed FM variation (TEO-FM-Var) and TEO Autocorrelation Envelope (TEO-Auto-Env) [35].

Physiological signals are another data source for recognizing people's emotions [36]. The idea of wearables that detect the wearer's affective state dates back to the early days of affective computing [37]. For example, skin conductance changes if the skin is sweaty, which is related to stress situations and other effects. Skin conductance is used as an indicator of arousal, to which it is correlated [38]. A low level of skin conductivity suggests a low arousal level. Heart rate is also a physiological signal connected with emotions, as its variability increases with arousal. Generally, heart rate is higher for pleasant and low arousal stimuli compared to unpleasant and high arousal stimuli [38].

2.3. Emotion Recognition in E-Learning

The tools presented above are the basis of the emotion recognition research and have been applied in the last years to create emotion-aware learning systems. Reviewing the most relevant works, the proposal of Derick et al. [39] stands out. They implemented a set of visualizations intending to allow learners to reflect their affective states and their connection with specific learning activities through self-reporting. The results they obtained were that the simplest visualizations helped the students in a better way since most of them did not know the interpretation of complex ones. Also making use of the self-report, and complementing it with data about interactions between students and teachers, Ruiz et al. [40,41] proposed a method to measure students mood based on a model of twelve emotions (six positives and six negatives). With the data obtained, a dashboard was implemented, resulting in the improvement of the academic results of almost 75% of the students.

The work of [42] opted for using a library known as Clmtrackr through a web-cam to detect emotions based on facial expression. In this way, it was possible to relate the students' mood with the learning activity they were doing and overcoming some of the difficulties of affective communication in e-learning environments. Ez-Zaoui et al. [43] propose a multimodal and contextual approach to recognize emotions in e-learning platforms. With this purpose, four sources of information were captured and analyzed: audio, video, self-report, and students' interactions. This analysis resulted in a dashboard designed for teachers with the correlation between activities and emotions.

Finally, the work carried out by Robal et al. [44] tried to evaluate the performance of gaze tracking tools to detect attention losses of students in Massive Open Online Courses (MOOC). The conclusion reached was that hardware-based tools perform better than software-based ones. Despite this, the platform proposed in this work implements a software solution, known as Webgazer [45] because of the impossibility of using dedicated hardware outside the lab and its non-negligible results (an error of 175 pixels and an average visual angle of 4.17°). The significant majority of these works have in common the use of a dashboard to show the emotions captured in the learning process [42]. These works do not only look for collecting as much as possible data about students' emotions, but they also pretend to show the data in an effective way to be used by teachers and students to improve their results.

2.4. Learning Analytics

The growing interest in applying machine learning techniques in different fields, coupled with the emergence of numerous data capture techniques, has resulted in a growing trend to use both to improve virtually any area of society. Within this trend appears the concept of learning analytics, which can be defined as "*the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning*" [9]. This definition takes on an even higher dimension if the environment of an e-learning platform is considered, in which students' interactions are captured simply and on a large scale.

In the approach proposed by [46], the learning analytics process is defined in three steps. The first one is to collect the most relevant data from the educational environment and pre-process it to clean, anonymize, and transform data into a format suitable for later use. The second step is based on applying different data analysis techniques must be applied in order to find a solution to improve the learning process. Once a solution appears, it is needed to take action to achieve the goals proposed. Finally, the third step, named post-processing, consists of evaluating the entire process. This evaluation can result in obtaining new data that had not been taken into account, pre-processing better oriented to the final objective, defining new attributes and metrics, modifying analysis variables, or choosing different prediction and analysis methods. As mentioned, learning analytics is a research field with growing interest today, which is reflected in the amount of research that seeks to improve the academic outcomes of students participating in experiments. Beginning with the work of [47], this research aimed to combat dropout in the first year of college. With this purpose, a system was developed with the capabilities of making predictions about students' academic future based on academic and demographic information. The results showed that students that had been advised about their trends obtained better grades and a lower dropout rate than their peers in the same semester.

Pardos et al. [48] analyzed the relationship between affective state and student behavior. With this aim, measurements about student's work context, actions, utterances, facial expressions, body language, and interactions with teachers or fellow students were taken. Once it was made, a machine learning model was implemented, whose results demonstrated that it is possible to predict students' final grades based on their behavior and course performance. Lastly, the work of [49] tried to apply learning analytics in order to improve the learning process when students work in teams. They proposed a system capable of monitoring and evaluating teamwork competencies through data about e-learning platform forums, cloud storage platform interactions and files, and wiki tools. The results showed a strong relationship between active interactions, such as writing a message in the course forum, and academic performance. According to this, teachers thought that, using this system, they were in a better position to evaluate each student's teamwork.

2.5. Semantic Task Automation

A recent shift in ambient intelligence is to move from the model of full and transparent automation to smart collaboration since autonomous often leave users feeling out of control [50]. A popular approach to interconnect and personalize both IoT and Internet services is the use of Event-Condition-Action (ECA) rules, also known as trigger-action rules. A number of now prominent web sites, mobile, and desktop applications feature this rule-based task automation model, such as If-this-then-that (IFTTT) [51]. These systems, called Task Automation Service (TAS) [52], are typically web platforms or smartphone applications that provide an intuitive visual programming environment where inexperienced users seamlessly create and manage their automations.

In this context, we developed EWE Tasker [53], a semantic task automation platform which aims to board rule capabilities and address two of the major drawbacks of these systems: (1) the only incoming data streams available are those the platform is prepared for; and, (2) the lack of a mechanism to use and reason over large scale data outside their platform such as the Linked Open Data (LOD) cloud or context data [54].

EWE Tasker is based on the Evented WEb (EWE) ontology [52], a standardized data schema designed to model the most significant aspects of Task Automation Services. EWE provides a common model to define and describe task automation services, representing its rules and enabling rule interoperability. In addition, it provides a base vocabulary in order to build domain-specific vocabularies. Four major classes make up the core of EWE: *Channel*, *Event*, *Action*, and *Rule*. The class

Channel defines individuals that either generate Events, provide Actions, or both (such as an emotion recognition sensor or a smart light). The class *Event* defines a particular occurrence of a process and allows users to describe under which conditions should rules be triggered. The recognition of sadness generated by the emotion detector sensor is an example of an entity that belongs to this class. The class *Action* defines an operation provided by a Channel that is triggered under some conditions (i.e., to change the light color). Finally, the class *Rule* defines an *ECA* rule, triggered by an Event that produces the execution of an Action. An example of a rule is: "*If sadness is detected, then change the light color.*".

Users in the EWE Tasker platform are able to create and configure their own automation rules in an easy way by means of a mobile phone application or a web interface. The platform integrates several external devices and services (such as Gmail or Spotify) and can be adapted to different smart environments in an easy way benefiting from the advantages of semantic technologies.

3. Emotion-Aware E-Learning Platform Architecture

Once the state of the art is analyzed, we present the proposed architecture along with the implementation details. We aim to build a system that enables the personalization of the learning experience based on students' emotions. With this purpose, and taking into account some e-learning literature [55–57], we have identified the following requirements to fulfill by the architecture: (1) to be integrable into the chosen e-learning platform; (2) to recognize and store the students' emotions; (3) to be able to convert data into valuable information for users; and (4) to enable automatic adaptation of the student environment. The proposed architecture matching these requirements is shown in Figure 1.

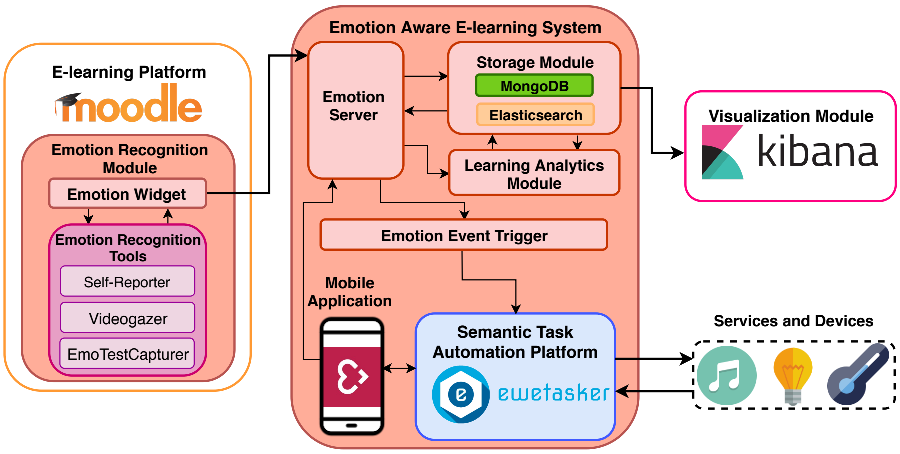


Figure 1. Emotion-aware e-learning platform architecture.

The platform architecture is composed of two main components: emotion-aware e-learning module and emotion recognition module. The former is the core of the platform, and it provides the system with the capability of collect, store, process, analyze, and visualize all the data regarded to the system. In addition, it integrates a semantic task automation platform. On the other hand, the emotion recognition module dotes the system with emotion and engagement detection. This module is integrated into an e-learning platform.

3.1. Emotion Recognition Component

The Emotion Recognition component is responsible for collecting data from the e-learning platform. It consists of different views developed as web components that must be integrated into the original e-learning platform. For the proposed architecture, we have used Moodle as the e-learning platform. Moodle allows us to create online courses quickly and easily, thanks to its friendly web interface. In addition, its use is widespread in the academic world, so it is not a barrier for teachers or students, giving relevance to our case study. Moodle also allows us to integrate the emotion recognition component easily.

This component is composed of two modules: emotion widget and emotion recognition tools. The former is responsible for sending all the captured data in the platform to the emotion-aware e-learning component, while the latter is responsible for collecting data regarding students' emotion and engagement. Three emotion recognition tools have been integrated into this component: self-reporter, videogazer, and emotestcapturer.

The self-report tool implements a form in which students can record the emotions they are experiencing each day of the course, as well as the level of intensity of these emotions. Students can fill the report everyday indicating how they feel by means of selecting an emotion and its intensity. These data allow us to perform constant monitoring of students' moods throughout the course using real metrics that have their origin in the subject itself. In addition, these data allow us to get a correlation with measurements obtained from other sub-modules, and with the academic results achieved in each phase of the course. The videogazer tool, based on WebGazer library [45], allows us to detect students' attention in video lessons. Thanks to this tool, it is possible to determine whether the user is paying attention to the lesson or not based on his/her looking position. In this way, it is possible to relate the different phases of the video lesson with the attention and interest produced by each of them. Finally, the emotestcapturer implements a tool that allows us to capture students' emotions when they are solving a test. This tool is empowered by the Clmtracker library [58], which is able to detect the emotion of users through the position of the coordinates that draw their faces. All the data collected by these tools are sent to the emotion-aware e-learning component, where it will be processed.

3.2. Emotion-Aware E-Learning Component

The Emotion Aware E-learning component's main functions are collecting, storing, processing, analyzing, and visualizing the data obtained from the emotion recognition component. In addition, the module is able to adapt the smart environment according to automation rules, thanks to the integration of a task automation platform. An overview of the data interchange carried out by this module is seen in Figure 2. The data are received by the emotion server submodule.

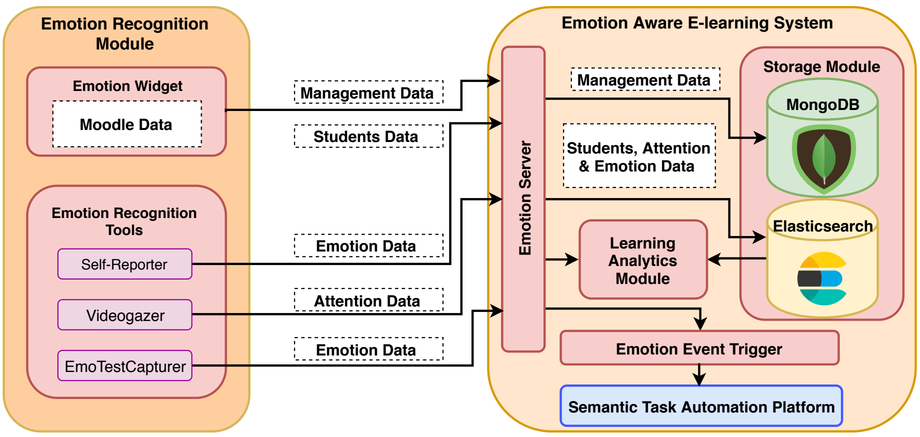


Figure 2. Emotion aware e-learning module flow diagram.

This submodule consists of a web service that receives and handles requests from the emotion recognition module. It receives information regarding management or students, and pass this information to the corresponding submodule of the system depending on its nature. This information is stored by the storage submodule and sent to the learning analytics submodule and the emotion event trigger. The storage submodule consists of two repositories that store the system logic data and the users' related data. One repository, consisting of a MongoDB database [59], is used for storing personal user information (age, sex, etc.) and academic information (courses, lessons, etc.). The other repository, based on the distributed analytics and search engine Elasticsearch [60], is used to store the emotion and attention-related information. This data contains all the information about the students' mood, personal information, and grades.

In order to visualize these data, the Elasticsearch based repository connects with the visualization module, based on the dashboard-based visualization tool Kibana [61]. The data must be shown in an accessible and easy to understand way, without the need for previous knowledge by the users. To make this possible, Kibana has been selected as an implementation tool for these visualizations. Kibana provides a web-based environment, on which we can implement our visualizations. We have used different chart types for defining several visualizations. These visualizations, which can be seen in Section 4, are grouped into three different dashboards: two for teachers and one for students. The visualizations designed for these dashboards are average self-report emotions per time (area); average grades, maximum and minimum grades per topic (table); the number of students enrolled (gauge); success rate (goal); grade levels (heatmap); average attention per video, grade average per topic, average emotions per test (bar); emotion evolution per minute in a test, average attention per second, average self-report emotions per time (line); the number of students per gender, and the number of students per grade (pie) and student country (region).

Other than analyzing data by means of the visualization, the system also provides a way to perform data analysis using machine learning techniques. The learning analytics submodule has this goal and allows the users to draw on conclusions that cannot be obtained through ordinary study and visualization techniques. This submodule performs an exploratory analysis of the data using K-Means and finding correlations between features, enabling the detection of anomalous cases, trends, and correlations in the students' data.

Finally, to perform automatic adaption of the students' smart environment, a semantic task automation platform has been integrated. The integrated automation platform is EWE Tasker, developed in a previous work [16]. This platform offers an easy way to integrate smart devices and services in an environment and use them for automating daily tasks. These integrations stem from devices as smart lights, sensors, and smartphones to services as Gmail or Twitter. Users in the platform are able to create and configure their own automation rules, and integrate new devices and services thanks to its semantic core, based on the EWE ontology [52]. In order to import the received data into the platform, the emotion event trigger submodule applies over it a semantic layer. Listing 1 shows an example of emotion detected event, along with its corresponding parameters and the generator channel, modeled according to work in Muñoz et al. [62]. The event has been generated by the detection of "*sadness*" with an average intensity of 84%. The example is written using Notation3, a shorthand non-XML serialization of Resource Description Framework (RDF) models designed with human-readability in mind [63].

The task automation platform processes these events and performs the corresponding actions on the environment, according to the automation rules predefined by the users. The platform is linked to different services and devices, enabling the adaption of the environment to students' emotion or engagement. In addition, it includes a smartphone app that can connect with different devices in order to open new possibilities. Some of the developed possibilities are stress detection from Empatica E4 wristband [64] and indoor geolocation using Estimote iBeacons [65]. The integration of the automation platform enables the automatic improvement of students' mood and comfort by adapting their environment.

Listing 1. Semantic representation of emotion detected event written in Notation3.

```
ewe:hasEmotion rdfs:subPropertyOf ewe:hasParameter .
ewe:hasUser rdfs:subPropertyOf ewe:hasParameter .
onyx:Emotion rdfs:subClassOf ewe:Parameter .

ewe:WebCamSensor a owl:Class;
rdfs:subClassOf ewe:Channel;
dcterms:title "Web Camera";
dcterms:description "This~channel represents a web camera sensor able
to recognize emotions from video.";
ewe:generatesEvent ewe:EmotionDetectionEvent.

ewe:EmotionDetectionEvent a owl:Class ;
rdfs:subClassOf ewe:Event ;
dcterms:description "This~event fires every time an emotion is detected
by a certain user." ;
dcterms:title "New emotion detected" ;
ewe:hasEmotion onyx:Emotion;
ewe:hasUser ewe:User .

ewe:web-cam1 rdf:type ewe:WebCamSensor.

onyx:sadness1 rdf:type onyx:Emotion;
dcterms:title "Sadness emotion detected";
dcterms:value "sadness";
onyx:source ewe:web-cam1;
onyx:hasEmotionIntensity 0.84 .

ewe:student3238 rdf:type ewe:User;
foaf:accountName "Student 3238".

ewe:sadnessdetected1 rdf:type ewe:EmotionDetectionEvent ;
ewe:generatedBy ewe:web-cam1;
ewe:hasEmotion onyx:sadness1 ;
ewe:hasUser ewe:student3238 .
```

4. Case Study

In order to better understand the usefulness of the proposed system, we have defined three case studies: Academic and Emotional Dashboards for teachers, Emotional Dashboard for students, and Smart Automation for students and teachers. The context in which these case studies take place is an online programming course. This course, entitled "*Basic Python for Beginners*" [66] is available on the Moodle.net. The course offers a set of video lessons and topic-based questionnaires.

These case studies are described in the following, explaining how they allow us to test the main features of the system: emotion and engagement recognition; data analysis, collection, and visualization; and smart automation.

4.1. Academic and Emotional Dashboards for Teachers

This case study is meant to remove communication barriers between teachers and students in distance learning environments. An overview of this use case is given in Figure 3.

The student's goal is to complete the lessons that compound the course. With this purpose, the student will carry out three types of activities: theory study, video lessons, and questionnaires. Meanwhile, student's engagement, stress, and mood will be monitored. In addition, self-reported emotions' questionnaires will be performed during the course. While the student is doing the video lessons, his/her attention is being measured. This enables us to capture the student engagement related to the topic and to the current video time. In addition, the student can perform a test, answering a series of questions related to the studied topic. While the user performs the test, emotions and stress level experienced are captured, as can be seen in Figure 4.

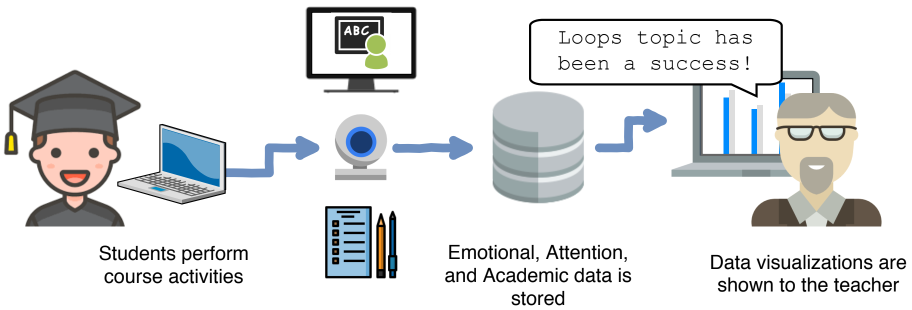


Figure 3. Academic and emotional dashboards for teachers' case study.

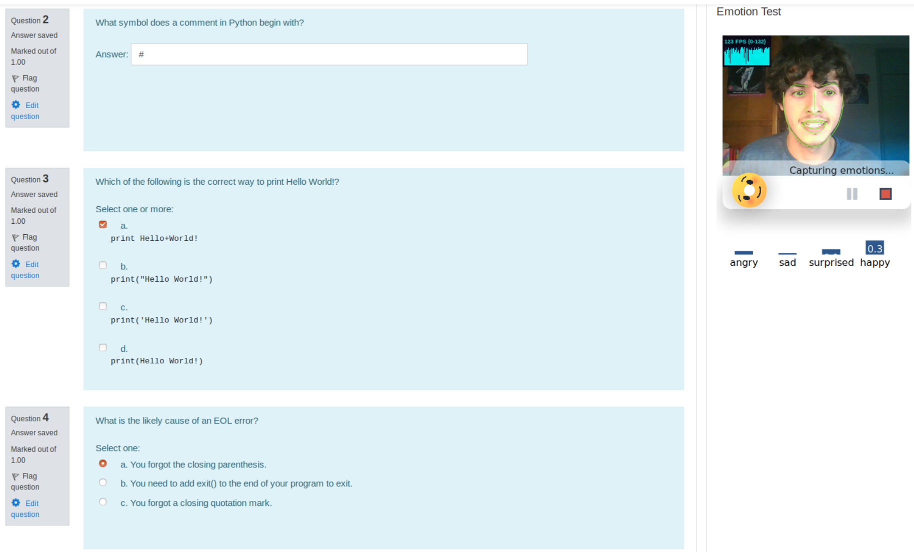


Figure 4. Emotional capture in the e-learning platform.

The collected data will be shown to the teacher in a dashboard, enabling a better understanding of the student's learning experience. This allows the teacher to perform adaptations in methodology, content, or activities in order to improve the student mood, and, consequently, his academic results.

Figure 5 shows an example of the charts enabled to the teacher in the dashboard. These charts show the statistics obtained from students' during the different activities of the course. This information stems from grades to emotional data and associated with each topic. In this way, teachers can analyze how each emotion and engagement affects the learning outcomes obtained. In this example, the chart shows the students' average engagement, allowing the teacher to analyze the average of attention that the students have paid to the video lesson of a specific topic. It can be seen how the attention in the first seconds of the video is high, while as time goes by, it varies until it drops sharply at the end.

Through these visualizations, the teacher can better understand the characteristics of the group of students who are participating in the course. This allows him to adapt the contents to the type of students who have greater difficulties, taking into account a great variety of factors such as emotion, attention, age, or nationality. In this way, the communication barriers of an e-learning environment are overcome.

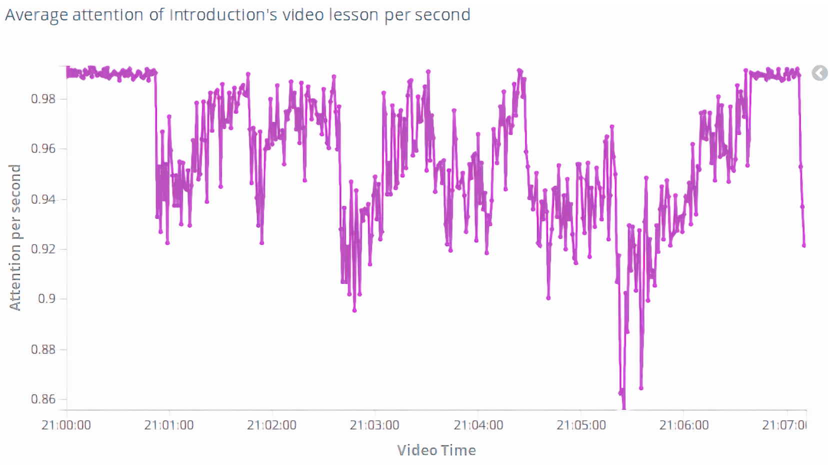


Figure 5. Average attention of students for a certain lesson.

4.2. Emotional Dashboard for Students

This case study is based on the previous one but also includes an emotional dashboard for students. This dashboard allows students to perform an introspection exercise, in which they can learn how emotions affect their performance, and, consequently, look for solutions to make the most of the course. An overview of this study case is presented in Figure 6.

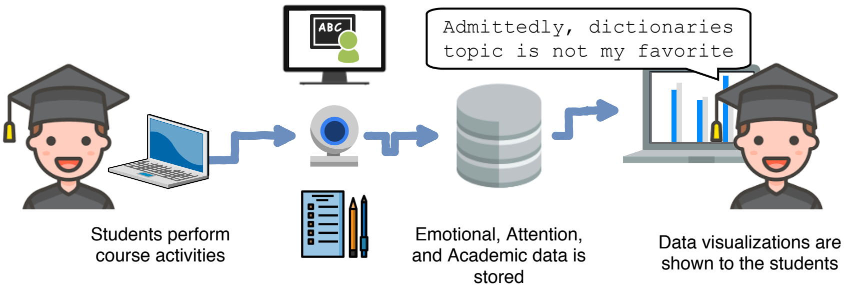
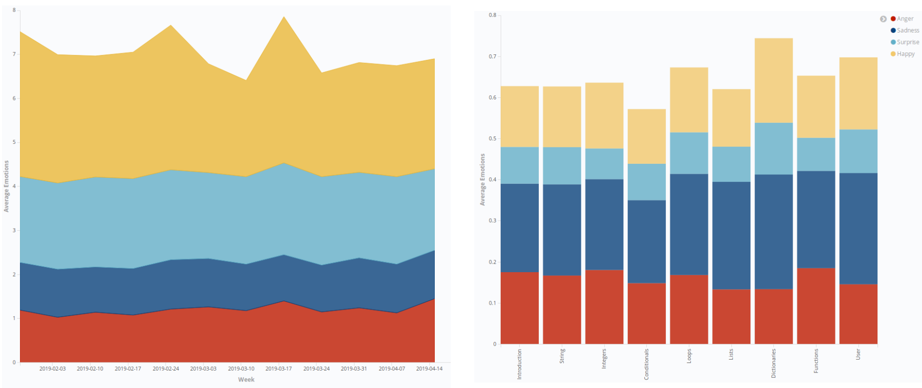


Figure 6. Emotional dashboard for students' case study.

The Student's Emotional Dashboard is designed to show the evolution of the students' emotional state in a clean and straightforward way. In addition, the visualizations include the grades obtained during the course. Figure 7 shows some of the developed visualizations.

Particularly, Figure 7a shows the evolution in the emotions that the student has experienced while performing tests. The color code, used in this visualization and in others in which emotions appear, is inspired by Borth et al. [67]. In this example, it can be seen how the most reported emotion is happiness. It can also be seen how surprise is a frequently reported emotion and how negative emotions have small values in comparison.

On the other side, Figure 7b shows how the feelings expressed by the student have evolved during the different lessons of the course. In this example, we can see that sadness is the emotion that the student experiences with higher intensity, followed by anger.



(**a**) (**b**)

Figure 7. Self-reported emotion and average captured emotions for a test. (**a**) emotions self-reported by students shown per week; (**b**) example of the average captured emotions.

4.3. Smart Automation for Students and Teachers

The last case study has as a main goal to improve the comfort and mood of students while they are studying, through the use of smart devices and task automation. For this purpose, three different scenarios have been designed, as shown in Figure 8. These scenarios are mood regulation, attention improvement, and report generation.

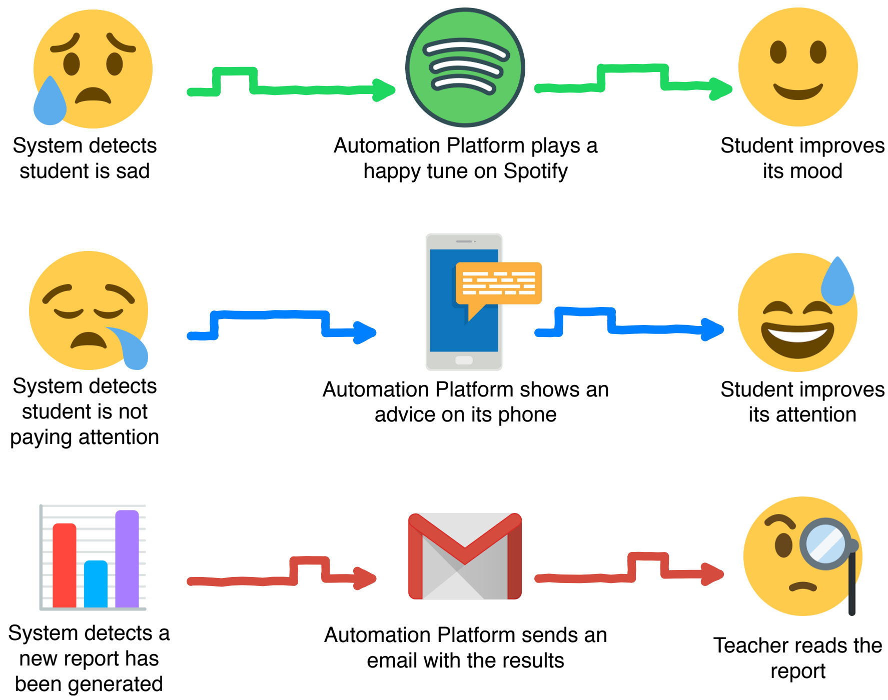


Figure 8. Smart automation scenarios.

The mood regulation scenario aims to adapt the environment to the emotions experienced by a student through the use of smart devices located in his/her workplace. Students can create and configure their own mood regulation rules using the task automation platform. In this way, the environment of the student can automatically change according to his/her emotions, with the purpose of enhancing positive feelings. The automation rules can involve several devices or services, such as smart lights or speakers. In this way, while the student is studying, the work environment will adapt to his/her mood, improving his/her comfort and, consequently, his/her performance. An example of a rule to achieve this automation is: *If a student is sad, then play happy music*.

Another possible automation is attention improvement. This case study aims to improve students' attention when they are studying the course contents through the e-learning platform. As in the mood regulation scenario, automation rules will be created and configured by the students. In this scenario, the system will detect if the attention paid to the slides is under a certain threshold. When this happens, an action will be done by the system for warning the student and improving his/her attention. These actions can go from showing notifications on the phone to light or sound signals, depending on the student's preferences. Through these automations, the student can improve his concentration, optimizing the study time, and obtaining better academic results. A rule example for attention improvement is: *If the student's attention falls, then show a notification*.

Finally, the report generation scenario has the goal of facilitating the exploratory analysis of the course data presented in the previous section. This data analysis, when we are faced with a large amount of data, becomes a process that can take from the order of minutes to hours. Waiting until the end of this process can discourage teachers from using these analysis techniques, leaving them in disuse or undervalued. For this reason, and taking advantage of the capabilities of the automation platform, the teachers can configure rules for being warned about the completion of the analysis. In this way, when the system detects an analysis task has finished, the teacher can be notified using his preferred method configured in the automation rule (email, mobile notification, etc.), and can also obtain the generated report. A possible rule following this scenario is: *If a new analysis has been performed, then send an email to the teacher*.

These automations are carried out through the definition of semantic rules from the automation platform. Listing 2 shows the semantic definition of a sample mood regulation rule, with its corresponding event, action, and parameters. In this case, the rule is: *If a student is sad, then play happy music*, composed by the event of sadness detection and the action of playing a happy song. The song to play and the emotion detected are the parameters of the rule.

In this section, we have presented the case studies designed to improve the mood of students participating in an online course through the study of their emotions. The proposed platform achieves this goal by the integration of several visualization dashboards, analysis techniques, and smart automation. To this end, firstly, the use of visualization tools has been proposed to broaden the knowledge of both teachers and students about the interactions that occur during the course activities. These interactions include data pertaining to emotions, attention, and course grades.

Listing 2. Rule instances.

```
ewe:PlaySong a owl:Class ;
rdfs:subClassOf ewe:Action ;
dcterms:description "Play a song on a speaker.";
dcterms:title "Play song";
ewe:hasSong mo:Track.

ewe:Speaker a owl:Class;
rdfs:subClassOf ewe:Channel;
dcterms:title "Audio speaker";
dcterms:description "This~channel represents a smart speaker
able to play sounds.";
ewe:providesAction ewe:PlaySong.

ewe:play-happy-song rdf:type ewe:PlaySong ;
ewe:providedBy ewe:speaker1;
ewe:hasSong mo:all-together-now.

ewe:speaker1 rdf:type ewe:Speaker.

ewe:sadnessdetected1 rdf:type ewe:EmotionDetectionEvent ;
ewe:generatedBy ewe:web-cam1;
ewe:hasEmotion onyx:sadness1 ;
ewe:hasUser ewe:student3238 .

mo:all-together-now rdf:type mo:Track ;
dc:title "All~together now" ;
foaf:maker mo:the-beatles.

ewe:regulate-sadness a ewe:Rule ;
dcterms:title    "Sadness regulation rule" ;
dcterms:description "This~rule aims to regulate sadness by mean of
playing a happy song." ;
ewe:triggeredByEvent    ewe:sadnessdetected1 ;
ewe:firesAction         ewe:play-happy-song .
```

5. Validation

As presented before, the main experimental contribution of this work was the design and implementation of an emotion-aware Learning Analytics system for e-learning platforms, based on Machine Learning techniques, and integrated with a semantic task automation platform. With the purpose of knowing the scope of this research, a set of hypotheses was raised:

Hypothesis 1 (H1). *To be aware of his/her emotions improves user satisfaction.*

Hypothesis 2 (H2). *The system quality of an e-learning platform directly affects user satisfaction.*

Hypothesis 3 (H3). *To be aware of their emotions helps students to obtain better academic results.*

Hypothesis 4 (H4). *The system quality of an e-learning platform is a key factor for academic results.*

To evaluate the proposed system, it was tested in an experiment that was carried out in the scope of a Python's course in Moodle platform [68] with real users. In this experiment, the entire system was deployed, including the emotion recognition tools, the visualization, and analysis modules and the smart automation system. The following sections cover the participants, materials, design results, and conclusions drawn from the experiment, focusing on its objective.

5.1. Participants

The experiment was conducted among 30 students from a technical university, including both genders. The ages of the participants ranged from 20 to 25. Their selection was motivated by the characteristics of the course in which the experiment was set, a Python programming course for beginners. The participants' knowledge of programming was extensive, while their knowledge of Python language depended on each case.

The ethical approval for the experiment was obtained by the Ethics Committee of the Technical University of Madrid, and the study was carried out in accordance with the ethical guidelines [69]. Participants were aware of the nature of the experiment, and gave informed consent prior to participation in the study. All subjects could abort the experiment at any time.

5.2. Materials

The material needed to carry out this experiment includes a wide variety of devices, sensors, and computer equipment, detailed below:

- Camera (Gucee HD92) that feeds the video to the emotion recognizer submodule.
- Room lighting (WS2812B LED strip controlled by WeMos ESP8266 board) is used as an actuator on the light level of the room, with the possibility of using several lighting patterns.
- Google Chromecast [70] that transmits content in a local network.
- LG TV 49UJ651V. This device is used for displaying images.
- Google Home. The system uses this device to play music.

Participants accessed the web HTML-based interface using a desktop computer with the Firefox browser [71].

5.3. Methodology

During the experiment, each participant performed a lesson in an online programming course. This course, entitled "*Basic Python for Beginners*" [66], is available on the Moodle.net. This course has been chosen mainly because it adapts easily to the case studies that we defined in Section 4.

While participants were doing the course, the e-learning system is fed with the information provided by the different sensors that are continually monitoring the participant emotional and attention states. While the student is doing the course, different adaption rules are triggered depending on his/her emotion and attention, such as changing light color or playing music. In addition, the student can visualize his/her emotions utilizing the developed dashboard. The experiment finishes when the lesson is finished. In addition, a questionnaire is given to the participants just after the sessions conclude. These questions use five-point Likert scales and are oriented to offer the participant's view of the system. The questions raised are summarized in Table 1.

Table 1. Questionnaire for the participants.

No.	Question Formulation	Code
Q1	The system is easy to use	SQ1
Q2	System reliability is correct	SQ2
Q3	The system provides interactive features between user and system	SQ3
Q4	The system provides custom information	SQ4
Q5	The system features are attractive for the user	SQ5
Q6	The system provides information in a rapid way	SQ6
Q7	System interface is clear and intuitive	SQ7
Q8	The system provides the needed information related to emotions	EIQ1
Q9	The information provided by the system is accessible at the required moment	EIQ2
Q10	The information related to emotions provided by the system is relevant for your task	EIQ3
Q11	The system provides enough information related to emotions	EIQ4
Q12	The information provided by the system is easy to understand	EIQ5
Q13	The system provides updated data of your emotions	EIQ6
Q14	The system enables me to accomplish the tasks more efficiently	US1
Q15	The system helps me learn effectively	US2
Q16	The system improves my learning performance	US3

Table 1. *Cont.*

No.	Question Formulation	Code
Q17	The system features are improvements with regards to standard e-learning platforms	US4
Q18	Overall, the system is useful	US5
Q19	The system saves my time in searching for materials	BE1
Q20	The system has increased my knowledge and helped me to be successful in the course	BE2
Q21	The system makes communication easier with the teacher and other students	BE3
Q22	The system allows the teacher to adapt to students' needs	BE4
Q23	The system has helped me to achieve the learning goals of the module	BE5

5.4. Results

We tested our research hypotheses using partial least squares structural equation modeling [17], with SmartPLS 3.2.9 [72]. This software allows us to easily check the psychometric properties of the measurement and structural model. It enables the analysis of the reliability and validity, and estimate the strength of the relationships among different model variables simultaneously.

PLS-SEM is suitable for small samples, obtaining good results with a sample size at least 10 times the largest number of structural paths directed at a particular construct in the structural model [73]. Figure 9 shows an overview of our model, according to the proposed hypotheses. As in our model, the maximum number of paths is two, and our minimum sample size should be 20, so we consider PLS-SEM an appropriate method according to our sample.

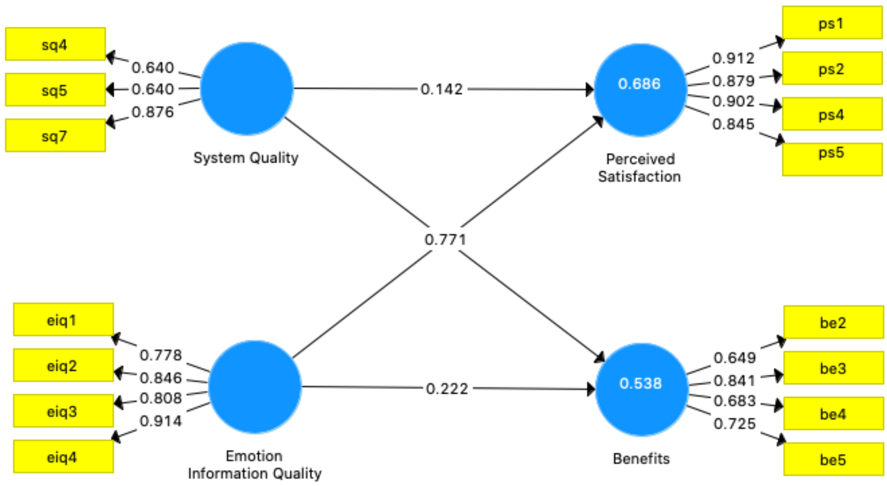


Figure 9. Emotion-aware e-learning platform use validation model.

The measurement model was assessed using the following criteria: First, we will check the indicator reliability, verifying that outer loading for the indicators is greater than 0.70 [74]. Secondly, we will use Cronbach's alpha (α) and Composite Reliability (CR) for proving internal consistency reliability. The cut-off value is to be greater 0.70 for both tests [75]. Finally, we will check for convergent and discriminant validity. The former one with the average variance extracted (AVE), which should be greater than 0.50 [76]. The later, discriminant validity, will be checked: using Fornell–Larcker criterion [76]; Cross-loadings [75]; and the Heterotrait–Monotrait ratio (HTMT) [77].

For the test of individual item reliability, we examined the outer loadings. This gives us the loading of the measures on the constructs they intend to measure. According to the work of [74], we deleted all indicators with a threshold lower than 0.40 and retained those which were higher than 0.70. For those whose value was between the two thresholds, we retained the indicators which did not decrease the value of AVE and CR. As a result, eight indicators were eliminated, resulting in the indicators shown in Table 2. Next, we retrieved CR, Cronbach's alpha, and AVE in order to check for internal consistency and convergent validity. In Table 2, we can see how all the values meet the minimum threshold for internal consistency. The composite reliability for all scales was between 0.76 and 0.93, greater than the minimum required value of 0.70. In addition, the average variance extracted AVE employed to assess the convergent validity was greater than 0.50 for all constructs. Third, we evaluate discriminant validity using the correlation matrix for the Fornell–Larcker method, whose results are presented in Table 4. Discriminant validity indicates the extent to which a given construct differs from other latent constructs [78]. The table shows that the diagonal values are larger than the other values inside the one column, which is the expected behavior according to the AVE.

Table 2. Internal consistency and convergent validity.

Construct and Items	Outer Loading	Outer t-Statistic	CR	AVE
System quality			0.767	0.529
SQ4: The system provides custom information	0.640	2.460		
SQ5: The system features are attractive for the user	0.640	3.732		
SQ7: System interface is clear and intuitive	0.876	2.603		
Emotional related information quality			0.904	0.703
EIQ1: The system provides the needed information related to emotions	0.778	3.986		
EIQ2: The information provided by the system is accessible at the required moment	0.846	1.238		
EIQ3: The related to emotions information provided by the system is relevant for your task	0.808	1.747		
EIQ4: The system provides enough information related to emotions	0.914	1.243		
Perceived usefulness			0.935	0.783
US1: The system enables me to accomplish the tasks more efficiently	0.912	1.577		
US2: The system helps me learn effectively	0.879	1.186		
US4: The system features are improvements with regard to standard e-learning platforms	0.902	1.194		
US5: Overall, the system is useful	0.845	1.177		
Benefits			0.817	0.530
BE2: The system has increased my knowledge and helped me to be successful in the course	0.649	3.420		
BE3: The system makes communication easier with the teacher and other students	0.841	2.846		
BE4: The system allows the teacher to adapt to students' needs	0.683	3.292		
BE5: The system has helped me to achieve the learning goals of the module	0.725	2.074		

The cross-loadings formed the second method utilized to assess discriminant validity. Table 3 shows the values of these loadings for each indicator. The results show that each indicator all the items correlated more strongly with their latent variable than with any other variables in the model [79]. Finally, we check for discriminant validity problems using the HTMT criterion, proposed by [77]. HTMT is equal to the average Heterotrait–Heteromethod correlation relative to the average Monotrait–Heteromethod correlations. The Heterotrait–Heteromethod correlations are correlations of indicators across constructs measuring different phenomena, while the Monotrait–Heteromethod correlations are correlations of indicators measuring the same construct. Table 4 shows that all the HTMT values are within the accepted threshold values (less than 0.90).

Table 3. Cross loadings for indicators.

	Emotion Information Quality	Benefits	System Quality	Perceived Satisfaction
EIR1	**0.778**	0.428	0.466	0.664
EIR2	**0.846**	0.250	0.045	0.691
EIR3	**0.808**	0.387	0.261	0.575
EIR4	**0.914**	0.373	0.314	0.791
BE2	0.078	**0.649**	0.479	0.048
BE3	0.317	**0.841**	0.643	0.290
BE4	0.527	**0.683**	0.460	0.269
BE5	0.283	**0.725**	0.439	0.173
SQ4	0.084	0.408	**0.640**	0.087
SQ5	0.308	0.248	**0.640**	0.347
SQ7	0.387	0.731	**0.876**	0.384
PS1	0.730	0.200	0.302	**0.912**
PS2	0.657	0.138	0.325	**0.879**
PS4	0.743	0.313	0.302	**0.902**
PS5	0.753	0.336	0.464	**0.845**

Bold numbers represent correlation of each item with its latent variable, which is stronger than with any other variable in the model.

Table 4. Fornell–Larcker method and HTMT correlations.

	Fornell–Larcker				**HTMT Correlations**			
	EIQ	**BE**	**SQ**	**PS**	**EIQ**	**BE**	**SQ**	**PS**
Emotion Information Quality	**0.838**							
Benefits	0.430	**0.728**			0.549			
System Quality	0.331	0.703	**0.727**		0.513	0.883		
Perceived Satisfaction	0.817	0.284	0.396	**0.885**	0.817	0.365	0.502	

Bold numbers show that the diagonal values of correlation are larger than the other values in the column, which is the expected behavior and validates discriminant validity.

Thus, our measurement model verifies all three conditions suggested by prior research, providing strong support for its reliability. Having confirmed that the measurement model is reliable and valid, we next evaluated the structural model results. For the structural model, we have followed the following criteria, proposed by [74]:

- Assess the structural model for collinearity issues (VIF < 5).
- Assess the significance and relevance of the structural model relationships (t > 1.65).
- Assess the level of R2 (The cut-off levels are: 0.190 weak; 0.333 moderate; and 0.6702 substantial).
- Assess the level of Q (cut-off point larger than zero).
- Assess the model's fit.

Thus, first, we check collinearity symptoms, to ensure that the results are valid. We do this by analyzing the variance inflation factor (VIF). A VIF greater than 5.00 would indicate overly high collinearity, and, consequently, a potential collinearity problem. The retrieved VIF values are all between 1.17 and 3.98, so they always meet the accepted threshold values (VIF < 5). Thus, we can affirm that collinearity was not a problem in our data.

Then, we proceed to evaluate the significance of each path coefficient. Path coefficients (β) representation relationships between the constructs in the model. These relationships are shown in Figure 9. The significance of each path coefficient, the standard errors, and the t-values were calculated by the algorithm of bootstrapping in PLS, with 5000 samples. From the analysis, we obtain that the emotional information contributes to improve user satisfaction ($\beta = 0.771, p = 0$) and benefits ($\beta = 0.23, p = 0.04$), supporting H1 and H3. In addition, we also observe that the system quality also improves user satisfaction and benefits, supporting H2 ($\beta = 0.142, p = 0.03$) and H4 ($\beta = 0.63, p = 0$).

We then assessed our model performance using the explanatory power (R^2) and predictive relevance (Q^2). Table 5 shows the results obtained with SmartPLS for each endogenous variable. The predictive relevance was obtained using Blindfolding in SmartPLS, with an omission distance D being 7. The values of R^2, according to the cut-off points of 0.67, 0.33, and 0.19 for substantial, moderate, and weak levels, respectively [80], gives us a moderate level for emotion information quality, and a substantial level for system quality. According to Hair et al. [73], Q^2 measures the capability of the path model to predict the endogenous measuring items indirectly from the prediction of their latent variables using the related structural relations. The cut-off point is to be larger than 0, so, as can be seen in Table 5, the model has strong predictive relevance for the endogenous constructs.

Table 5. Explanatory power (R^2) and predictive relevance (Q^2).

	R^2	Q^2
Emotion Information Quality	0.538 (moderate level)	0.244
System Quality	0.686 (substantial level)	0.460

As an additional indicator, we used the global goodness-of-fit (GoF) criterion proposed by [81] to assess the model fit. GoF is defined as "how well the specified model reproduces the observed covariance matrix among the indicator items" [74], so it allows us to evaluate the overall performance of the model in both measurement and structural levels. The GoF index is ranged between 0 and 1, and is calculated using the following formula:

$$GoF = \sqrt{(\overline{R^2} \cdot \overline{Q^2})} = \sqrt{(0.612 \cdot 0.352)} = 0.46 \quad (1)$$

There is no inference based criteria to assess the statistical significance of the GoF index, due to its descriptive nature [82]. However, Wetzels et al. [83] propose several baselines in order to validate the PLS models globally. These baselines are, according to different effect sizes of R^2: $GoF_{small}(0.10)$, $GoF_{medium}(0.25)$, and $GoF_{large}(0.36)$. The obtained value of 0.46, higher than the cut-off of 0.36 proposed by Wetzels et al. [83] for large effect sizes of R^2 indicates a good overall fit, and adequately validates the complex PLS model globally.

These results also validate the structural model, confirming the proposed hypotheses and validating that the use of an emotion-aware e-learning platform improves students' satisfaction and helps to improve their academic performance.

5.5. Teacher Validation

Finally, in order to validate the usefulness of the system for teachers, it has been tested by five professors at the Technical University of Madrid. Participants were asked to try the system, and then answer a set of questions for evaluating the usefulness of the system.

These questions use five-point Likert scales and are oriented to offer the participant's view of the system. The questions raised, along with the mean values obtained, are summarized in Table 6. As shown in the table, the system results in being quite useful for the professors, achieving high scores for all the questions. However, the score for the Q4 is a little lower compared with the others, which suggests that an effort must be still done in improving communication between the teacher and the students.

Table 6. Questionnaire for the professors.

No.	Question Formulation	Avg.
Q1	The information provided by the system related to students' emotion is useful	4.2
Q2	The system features are improvements with regard to standard e-learning platforms	4.8
Q3	The system allows the teacher to adapt the content to students' needs	4.6
Q4	The system makes communication easier with the teacher and other students	3.8
Q5	Overall, the system is useful	4.6

6. Conclusions and Future Work

This paper proposes an architecture for an emotion aware e-learning platform based on semantic technologies, which enables the emotion analysis in study places and the automated adaption of these study places to the students' needs. The proposed architecture allows teachers to study these analyses in order to adapt the methodology to this information, with the purpose of improving students' mood and consequently help them to achieve better academic performance. In this way, our platform advances in decreasing communication barriers between teachers and students in distance learning environments; while it also provides the students with mood regulation capabilities. In addition, the application of a semantic layer enables data interoperability and portability of components. Finally, the system has been implemented and evaluated using PLS-SEM methodology.

Through the experimentation, we have verified a set of hypotheses. In summary: (1) using the proposed emotion-aware e-learning system helps students to obtain better academic results; (2) the system quality of an e-learning platform directly affects user satisfaction; (3) using the proposed emotion-aware e-learning system improves user satisfaction; and, finally, (4) the system quality of an e-learning platform is a critical factor for academic results. These results encourage the use and improvement of this kind of e-learning systems, as they seem to provide users with several advantages, such as analysis of their mood and attention or adaption of their smart environment to this mood.

As future work, many lines can be followed to continue this work. One of these lines is to deep in the context of emotion and mood regulation. The integration of more sophisticated ways for smart environment adaption to emotions could be a key improvement for these systems. Currently, we are working on this line exploring new ways of emotion regulation. In addition, another line of future work would be the enhancement of the learning analytics side of the system, studying new methods, and improving the current analysis in order to offer new relevant data. Furthermore, we also plan to integrate new ways of emotion recognition in the system by means of the analysis of text interactions in students. This adds more reliability to the system as it incorporates a mature way of recognizing emotions. In addition, it would be relevant to implement the system in a real college course, in order to study and evaluate its benefits in a real scenario. In addition, in this line, the implementation of the system in companies which provide courses for employees' training would open new opportunities to the validation and improvement of the system. Finally, it would be also interesting to deeply evaluate the benefits of the system for the teachers, analyzing if the information about the students' provided by the system entails a saving in terms of effort and time costs to the adaptation of the material and courses to students, and if this adaptation provides better results.

Author Contributions: Conceptualization, S.M., E.S., and C.A.I.; Data curation, S.M. and E.S.; Funding acquisition, C.A.I.; Investigation, S.M.; Methodology, S.M.; Software, S.M. and E.S.; Supervision, C.A.I.; Validation, S.M.; Visualization, E.S.; Writing—original draft, S.M. and E.S.; Writing—review and editing, C.A.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Spanish Ministry of Economy and Competitiveness under the R&D project SEMOLA (TEC2015-68284-R).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Moore, J.L.; Dickson-Deane, C.; Galyen, K. e-Learning, online learning, and distance learning environments: Are they the same? *Internet High. Educ.* **2011**, *14*, 129–135. [CrossRef]
2. Arkorful, V.; Abaidoo, N. The role of e-learning, advantages and disadvantages of its adoption in higher education. *Int. J. Instr. Technol. Distance Learn.* **2015**, *12*, 29–42.
3. Garrison, D.R. *E-Learning in the 21st Century: A Community of Inquiry Framework for Research and Practice*; Routledge: London, UK, 2016.
4. Welle-Strand, A.; Thune, T. E-learning policies, practices and challenges in two Norwegian organizations. *Eval. Program Plan.* **2003**, *26*, 185–192. [CrossRef]
5. Andersson, A.; Grönlund, Å. A conceptual framework for e-learning in developing countries: A critical review of research challenges. *Electron. J. Inf. Syst. Dev. Ctries.* **2009**, *38*, 1–16. [CrossRef]
6. Cantoni, V.; Cellario, M.; Porta, M. Perspectives and challenges in e-learning: Towards natural interaction paradigms. *J. Vis. Lang. Comput.* **2004**, *15*, 333–345. [CrossRef]
7. Mayer, R.E. Searching for the role of emotions in e-learning. *Learn. Instr.* **2019**, 101213. [CrossRef]
8. Feidakis, M. A review of emotion-aware systems for e-learning in virtual environments. In *Formative Assessment, Learning Data Analytics and Gamification*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 217–242.
9. Siemens, G. Learning analytics: Envisioning a research discipline and a domain of practice. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada, 29 April–2 May 2012; ACM: New York, NY, USA, 2012; pp. 4–8.
10. Kelly III, J.E.; Hamm, S. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*; Columbia University Press: New York, NY, USA, 2013.
11. Witten, I.H.; Frank, E. Data mining: Practical machine learning tools and techniques with Java implementations. *ACM Sigmod Rec.* **2002**, *31*, 76–77. [CrossRef]
12. Bermudez-Edo, M.; Elsaleh, T.; Barnaghi, P.; Taylor, K. IoT-Lite: A lightweight semantic model for the Internet of Things. In Proceedings of the 2016 International IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), Toulouse, France, 18–21 July 2016; pp. 90–97.
13. Palavalli, A.; Karri, D.; Pasupuleti, S. Semantic internet of things. In Proceedings of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 4–6 February 2016; pp. 91–95.
14. Kort, B.; Reilly, R.; Picard, R.W. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In Proceedings of the IEEE International Conference on Advanced Learning Technologies, Madison, WI, USA, 6–8 August 2001; pp. 43–46.
15. Cook, D.J.; Augusto, J.C.; Jakkula, V.R. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive Mob. Comput.* **2009**, *5*, 277–298. [CrossRef]
16. Muñoz, S.; Llamas, A.F.; Coronado, M.; Iglesias, C.A. Smart Office Automation Based on Semantic Event-Driven Rules. In Proceedings of the Intelligent Environments (Workshops), London, UK, 12–16 September 2016; pp. 33–42.
17. Claes, F.; Cha, J. Partial least squares. *Adv. Methods Mark. Res.* **1994**, *6446*, 52–78.
18. Sylwester, R. How emotions affect learning. *Educ. Leadersh.* **1994**, *52*, 60–65.
19. Pintrich, P.R. A motivational science perspective on the role of student motivation in learning and teaching contexts. *J. Educ. Psychol.* **2003**, *95*, 667–686. [CrossRef]
20. Värlander, S. The role of students' emotions in formal feedback situations. *Teach. High. Educ.* **2008**, *13*, 145–156. [CrossRef]
21. Cano, F. Consonance and dissonance in students' learning experience. *Learn. Instr.* **2005**, *15*, 201–223. [CrossRef]
22. de la Fuente Arias, J.; Justicia, F.J. The DEDEPRO [TM] Model for Regulating Teaching and Learning: Recent Advances. *Electron. J. Res. Educ. Psychol.* **2007**, *5*, 535–564.

23. Pekrun, R. The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Appl. Psychol.* **1992**, *41*, 359–376. [CrossRef]
24. Ranellucci, J.; Hall, N.C.; Goetz, T. Achievement goals, emotions, learning, and performance: A process model. *Motiv. Sci.* **2015**, *1*, 98. [CrossRef]
25. Seal, C.R.; Naumann, S.E.; Scott, A.N.; Royce-Davis, J. Social emotional development: A new model of student learning in higher education. *Res. High. Educ. J.* **2011**, *10*, 114–115.
26. Frenzel, A.C.; Becker-Kurz, B.; Pekrun, R.; Goetz, T.; Lüdtke, O. Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *J. Educ. Psychol.* **2018**, *110*, 628–639. [CrossRef]
27. Shan, C.; Gong, S.; McOwan, P.W. Beyond Facial Expressions: Learning Human Emotion from Body Gestures. In Proceedings of the BMVC, Warwick, UK, 10–13 September 2007; pp. 1–10.
28. Wiedemann, G.; Rayki, O.; Feinstein, E.; Hahlweg, K. The Family Questionnaire: Development and validation of a new self-report scale for assessing expressed emotion. *Psychiatry Res.* **2002**, *109*, 265–279. [CrossRef]
29. Matlovic, T.; Gaspar, P.; Moro, R.; Simko, J.; Bielikova, M. Emotions detection using facial expressions recognition and EEG. In Proceedings of the 2016 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), Thessaloniki, Greece, 20–21 October 2016; pp. 18–23.
30. Araque, O.; Corcuera-Platas, I.; Sánchez-Rada, J.F.; Iglesias, C.A. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Syst. Appl.* **2017**, *77*, 236–246. [CrossRef]
31. Pantic, M.; Bartlett, M., Machine Analysis of Facial Expressions. In *Face Recognition*; I-Tech Education and Publishing: Seattle, WA, USA, 2007; pp. 377–416.
32. Sebe, N.; Cohen, I.; Gevers, T.; Huang, T.S. Multimodal approaches for emotion recognition: A survey. *Proc. SPIE* **2005**, *5670*, 5670-1–5670-12.
33. Mehta, D.; Siddiqui, M.F.H.; Javaid, A.Y. Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality. *Sensors* **2018**, *18*, 416. [CrossRef]
34. Anagnostopoulos, C.N.; Iliou, T.; Giannoukos, I. Features and Classifiers for Emotion Recognition from Speech: A Survey from 2000 to 2011. *Artif. Intell. Rev.* **2015**, *43*, 155–177. [CrossRef]
35. Ayadi, M.E.; Kamel, M.S.; Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognit.* **2011**, *44*, 572–587. [CrossRef]
36. Vinola, C.; Vimaladevi, K. A Survey on Human Emotion Recognition Approaches, Databases and Applications. *ELCVIA Electron. Lett. Comput. Vis. Image Anal.* **2015**, *14*, 24–44. [CrossRef]
37. Picard, R.W.; Healey, J. Affective wearables. *Pers. Technol.* **1997**, *1*, 231–240. [CrossRef]
38. Brouwer, A.M.; van Wouwe, N.; Mühl, C.; van Erp, J.; Toet, A. Perceiving blocks of emotional pictures and sounds: Effects on physiological variables. *Front. Hum. Neurosci.* **2013**, *7*, 295. [CrossRef]
39. Derick, L.; Sedrakyan, G.; Munoz-Merino, P.J.; Delgado Kloos, C.; Verbert, K. Evaluating emotion visualizations using AffectVis, an affect-aware dashboard for students. *J. Res. Innov. Teach. Learn.* **2017**, *10*, 107–125. [CrossRef]
40. Ruiz, S.; Charleer, S.; Urretavizcaya, M.; Klerkx, J.; Fernández-Castro, I.; Duval, E. Supporting learning by considering emotions: Tracking and visualization a case study. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh, UK, 25–29 April 2016; ACM: New York, NY, USA, 2016; pp. 254–263.
41. Ruiz, S.; Urretavizcaya, M.; Fernández-Castro, I.; López-Gil, J.M. Visualizing Students' Performance in the Classroom: Towards Effective F2F Interaction Modelling. In *Design for Teaching and Learning in a Networked World*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 630–633.
42. GhasemAghaei, R.; Arya, A.; Biddle, R. A dashboard for affective e-learning: Data visualization for monitoring online learner emotions. In *EdMedia+ Innovate Learning*; Association for the Advancement of Computing in Education (AACE): Waynesville, NC, USA, 2016; pp. 1536–1543.
43. Ez-Zaouia, M.; Lavoué, E. EMODA: A tutor oriented multimodal and contextual emotional dashboard. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada, 13–17 March 2017; ACM: New York, NY, USA, 2017; pp. 429–438.
44. Robal, T.; Zhao, Y.; Lofi, C.; Hauff, C. Webcam-based Attention Tracking in Online Learning: A Feasibility Study. In Proceedings of the 23rd International Conference on Intelligent User Interfaces, Tokyo, Japan, 7–11 March 2018; ACM: New York, NY, USA, 2018; pp. 189–197.

45. Papoutsaki, A.; Sangkloy, P.; Laskey, J.; Daskalova, N.; Huang, J.; Hays, J. WebGazer: Scalable Webcam Eye Tracking Using User Interactions. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI), New York, NY, USA, 9–15 July 2016; pp. 3839–3845.
46. Chatti, M.A.; Dyckhoff, A.L.; Schroeder, U.; Thüs, H. A reference model for learning analytics. *Int. J. Technol. Enhanc. Learn.* **2013**, *4*, 318–331. [CrossRef]
47. Arnold, K.E.; Pistilli, M.D. Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada, 29 April–2 May 2012; ACM: New York, NY, USA, 2012; pp. 267–270.
48. Pardos, Z.A.; Baker, R.S.; San Pedro, M.O.; Gowda, S.M.; Gowda, S.M. Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *J. Learn. Anal.* **2014**, *1*, 107–128. [CrossRef]
49. Fidalgo-Blanco, Á.; Sein-Echaluce, M.L.; García-Peñalvo, F.J.; Conde, M.Á. Using Learning Analytics to improve teamwork assessment. *Comput. Hum. Behav.* **2015**, *47*, 149–156. [CrossRef]
50. Mennicken, S.; Vermeulen, J.; Huang, E.M. From today's augmented houses to tomorrow's smart homes: New directions for home automation research. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 105–115.
51. Ovadia, S. Automate the internet with "if this then that" (IFTTT). *Behav. Soc. Sci. Libr.* **2014**, *33*, 208–211. [CrossRef]
52. Coronado, M.; Iglesias, C.A.; Serrano, E. Modelling rules for automating the Evented WEb by semantic technologies. *Expert Syst. Appl.* **2015**, *42*, 7979–7990. [CrossRef]
53. Muñoz López, S.; Fernández, A.; Coronado, M.; Iglesias, C.A., Smart Office Automation based on Semantic Event-Driven Rules. In *Proceedings of Workshop on Smart Offices and Other Workplaces, Colocated with 12th International Conference on Intelligent Environments (IE&039;16)*; Ambient Intelligence and Smart Environments; IOS Press: Amsterdam, The Netherlands, 2016; Volume 21, pp. 33–42.
54. Coronado, M.; Iglesias, C.A. Task automation services: Automation for the masses. *IEEE Internet Comput.* **2016**, *20*, 52–58. [CrossRef]
55. Horton, W. *E-Learning by Design*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
56. Bentley, T.; Johnston, L.; von Baggo, K. Putting some emotion into requirements engineering. In Proceedings of the 7th Australian Workshop on Requirements Engineering, Melbourne, Australia, 2–3 December 2002; pp. 227–244.
57. Keller, J.; Suzuki, K. Learner motivation and e-learning design: A multinationally validated process. *J. Educ. Media* **2004**, *29*, 229–239. [CrossRef]
58. Øygard, A.M. CmlTrcker, Javascript library for Precise Tracking of Facial Features via Constrained Local Models. Available online: https://github.com/auduno/clmtrackr (accessed on 14 April 2020).
59. Banker, K. *MongoDB in Action*; Manning Publications Co.: Shelter Island, NY, USA, 2011.
60. Dixit, B.; Kuc, R.; Rogozinski, M.; Chhajed, S. *Elasticsearch: A Complete Guide*; Packt Publishing Ltd.: Birmingham, UK, 2017.
61. Srivastava, A. *Kibana 7 Quick Start Guide: Visualize Your Elasticsearch Data with Ease*; Packt Publishing Ltd.: Birmingham, UK, 2019.
62. Muñoz, S.; Araque, O.; Sánchez-Rada, J.F.; Iglesias, C.A. An emotion aware task automation architecture based on semantic technologies for smart offices. *Sensors* **2018**, *18*, 1499. [CrossRef] [PubMed]
63. Berners-Lee, T.; Connolly, D.; Kagal, L.; Scharf, Y.; Hendler, J. N3logic: A logical framework for the world wide web. *arXiv* **2007**, arXiv:0711.1533.
64. McCarthy, C.; Pradhan, N.; Redpath, C.; Adler, A. Validation of the Empatica E4 wristband. In Proceedings of the 2016 IEEE EMBS International Student Conference (ISC), Ottawa, ON, Canada, 29–31 May 2016; pp. 1–4.
65. Gilchrist, C. *Learning iBeacon*; Packt Publishing Ltd.: Birmingham, UK, 2014.
66. Thompson, C. Python for Beginners. Available online: https://archive.moodle.net/mod/glossary/showentry.php?eid=93 (accessed on 14 April 2020).
67. Borth, D.; Ji, R.; Chen, T.; Breuel, T.; Chang, S.F. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, Spain, 21–25 October 2013; ACM: New York, NY, USA, 2013; pp. 223–232.
68. Büchner, A. *Moodle 3 Administration*; Packt Publishing Ltd.: Birmingham, UK, 2016.

69. UPM. Reglamento del Comité de Ética de Actividades de I+D+i de la Universidad Politécnica de Madrid. Available online: https://www.upm.es/sfs/Rectorado/Vicerrectorado%20de%20Investigacion/Servicio%20de%20Investigacion/Ayudas_y_Convocatorias/C.ETICA-REGLAMENTO.pdf (accessed on 17 September 2019).
70. Williams, K. The Technology Ecosystem: Fueling Google's Chromecast [WIE from Around the World]. *IEEE Women Eng. Mag.* **2014**, *8*, 30–32. [CrossRef]
71. Oualline, S.; Oualline, G. Web Browsing with Firefox. In *Practical Free Alternatives to Commercial Software*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–26.
72. Ringle, C.M.; Wende, S.; Becker, J.M. *SmartPLS 3*; SmartPLS: Bönningstedt, Germany, 2015.
73. Hair, J.F., Jr.; Sarstedt, M.; Hopkins, L.; Kuppelwieser, V.G. Partial least squares structural equation modeling (PLS-SEM). *Eur. Bus. Rev.* **2014**, *26*, 106–121. [CrossRef]
74. Hair, J.F.; Anderson, R.E.; Babin, B.J.; Black, W.C. *Multivariate Data Analysis: A Global Perspective*; Pearson Education: London, UK, 2010; Volume 7.
75. Urbach, N.; Ahlemann, F. Structural equation modeling in information systems research using partial least squares. *J. Inf. Technol. Theory Appl.* **2010**, *11*, 5–40.
76. Fornell, C.; Larcker, D.F. Structural equation models with unobservable variables and measurement error: Algebra and statistics. *J. Mark. Res.* **1981**, *18*. [CrossRef]
77. Henseler, J.; Ringle, C.M.; Sarstedt, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* **2015**, *43*, 115–135. [CrossRef]
78. Duarte, P.A.O.; Raposo, M.L.B. A PLS model to study brand preference: An application to the mobile phone market. In *Handbook of Partial Least Squares*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 449–485.
79. Gefen, D.; Straub, D. A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Commun. Assoc. Inf. Syst.* **2005**, *16*, 91–109. [CrossRef]
80. Al-Fraihat, D.; Joy, M.; Sinclair, J.; Masa'deh, R. Evaluating E-learning systems success: An empirical study. *Comput. Hum. Behav.* **2020**, *102*, 67–86. [CrossRef]
81. Tenenhaus, M.; Vinzi, V.E.; Chatelin, Y.M.; Lauro, C. PLS path modeling. *Comput. Stat. Data Anal.* **2005**, *48*, 159–205. [CrossRef]
82. Vinzi, V.E.; Trinchera, L.; Amato, S. PLS path modeling: From foundations to recent developments and open issues for model assessment and improvement. In *Handbook of Partial Least Squares*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 47–82.
83. Wetzels, M.; Odekerken-Schröder, G.; Van Oppen, C. Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Q.* **2009**, *33*, 177–195. [CrossRef]

3.2.6 An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices

Title	An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices
Authors	Muñoz, Sergio and Araque, Oscar and Sánchez-Rada, J. and Iglesias, Carlos
Journal	Sensors
Impact factor	JCR 2021 3.847
ISSN	1424-8220
Publisher	
Volume	18
Year	2018
Keywords	automation
Pages	1499
Online	`http://www.mdpi.com/1424-8220/18/5/1499`
Abstract	The evolution of the Internet of Things leads to new opportunities for the contemporary notion of smart offices, where employees can benefit from automation to maximize their productivity and performance. However, although extensive research has been dedicated to analyze the impact of workers' emotions on their job performance, there is still a lack of pervasive environments that take into account emotional behaviour. In addition, integrating new components in smart environments is not straightforward. To face these challenges, this article proposes an architecture for emotion aware automation platforms based on semantic event-driven rules to automate the adaptation of the workplace to the employee's needs. The main contributions of this paper are: (i) the design of an emotion aware automation platform architecture for smart offices; (ii) the semantic modelling of the system; and (iii) the implementation and evaluation of the proposed architecture in a real scenario.

sensors

MDPI

Article

An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices

Sergio Muñoz *, **Oscar Araque**, **J. Fernando Sánchez-Rada** and **Carlos A. Iglesias**

Intelligent Systems Group, Universidad Politécnica de Madrid, 28040 Madrid, Spain; o.araque@upm.es (O.A.); jf.sanchez@upm.es (J.F.S.-R.); carlosangel.iglesias@upm.es (C.A.I.)

* Correspondence: sergio.munoz@upm.es

Received: 19 March 2018; Accepted: 8 May 2018; Published: 10 May 2018

Abstract: The evolution of the Internet of Things leads to new opportunities for the contemporary notion of smart offices, where employees can benefit from automation to maximize their productivity and performance. However, although extensive research has been dedicated to analyze the impact of workers' emotions on their job performance, there is still a lack of pervasive environments that take into account emotional behaviour. In addition, integrating new components in smart environments is not straightforward. To face these challenges, this article proposes an architecture for emotion aware automation platforms based on semantic event-driven rules to automate the adaptation of the workplace to the employee's needs. The main contributions of this paper are: (i) the design of an emotion aware automation platform architecture for smart offices; (ii) the semantic modelling of the system; and (iii) the implementation and evaluation of the proposed architecture in a real scenario.

Keywords: ambient intelligence; smart office; emotion regulation; task automation; semantic technologies

1. Introduction

The emergence of Internet of Things (IoT) opens endless possibilities for the Information and Communication Technologies (ICT) sector, allowing new services and applications to leverage the interconnection of physical and virtual realms [1]. One of these opportunities is the application of Ambient Intelligence (AmI) principles to the workplace, which results in the notion of smart offices. Smart offices can be defined as "*workplaces that proactively, but sensibly, support people in their daily work*" [2].

A large body of research has been carried out on the impact that emotions have on decision making [3], health [4], emergencies [5] and working life [6]. This states the importance of recognizing and processing the emotions of people in intelligent environments. Particularly in the workplace, emotions play a key role, since the emotional state of workers directly affects other workers [7] and, consequently, company business. The application of emotion aware technologies to IoT environments entails a quantitative improvement in the workers' quality of life, since it allows the environment to be adaptive to these emotions and, therefore, to human needs [8]. In addition, this improvement in worker quality of life directly affects company performance and productivity [9].

Emotion Aware AmI (AmE) extends the notion of intelligent environments to detect, process and adapt intelligent environments to users' emotional state, exploiting theories from psychology and social sciences for the analysis of human emotional context. Considering emotions in the user context can improve customization of services in AmI scenarios and help users to improve their emotional intelligence [10]. However, emotion technologies are rarely addressed within AmI systems and have been frequently ignored [10,11].

A popular approach to interconnect and personalize both IoT and Internet services is the use of Event-Condition-Action (ECA) rules, also known as trigger–action rules [12]. Several now prominent websites, mobile and desktop applications feature this rule-based task automation model, such as IFTTT (https://ifttt.com/) or Zapier (https://zapier.com/). These systems, so-called Task Automation Services (TASs) [13], are typically web platforms or smartphone applications, which provide an intuitive visual programming environment where inexperienced users seamlessly create and manage their own automations. Although some of these works have been applied to smart environments [14,15], these systems have not been applied yet for regulating users' emotions in emotion aware environments.

This work proposes a solution that consists in an emotion aware automation platform that enables the automated adaption of smart office environments to the employee's needs. This platform allows workers to easily create and configure their own automation rules, resulting in a significant improvement of their productivity and performance. A semantic model for the emotion aware TASs based on the Evented WEb (EWE) [13] ontology is also proposed, which enables data interoperability and automation portability, and facilitates the integration between tools in large environments. Moreover, several sensors and actuators have been integrated in the system as a source of ambient data or as action performers which interact with the environment. In this way, the design of an emotion aware automation platform architecture for smart offices is the main contribution of this paper, as well as the semantic modelling of the system and its implementation and validation in a real scenario.

The rest of this paper is organized as follows. Firstly, an overview about the related work in smart offices, emotion regulation and semantic technologies is given in Section 2. Section 3 presents the semantic modelling of the system, describing different ontologies and vocabularies which have been used and the relationships between them. Then, Section 4 describes the reference architecture of the proposed emotional aware automation platform, describing the main components and modules as well as its implementation. Section 5 describes the evaluation of the system in a real scenario. Finally, the conclusions drawn from this work are described in Section 6.

2. Background

This section describes the background and related work for the architecture proposed in this paper. First, an overview of related work in AmE and specifically in smart offices is given in Sections 2.1 and 2.2, respectively. Then, the main technologies involved in emotion recognition and regulation are described in Sections 2.3 and 2.4. Finally, Section 2.5 gives an overview of the state of art regarding to semantic technologies.

2.1. Emotion Aware AmI (AmE)

The term AmE was coined by Zhou et al. [16]. AmE is "*a kind of AmI environment facilitating human emotion experiences by providing people with proper emotion services instantly*". This notion aims at fostering the development of emotion-aware services in pervasive AmI environments.

AmE are usually structured in three building blocks [10,17]: emotion sensing, emotion analysis and emotion services or applications.

Emotion sensing is the process of gathering affective data using sensors or auto-reporting techniques. There exists many potential sensor sources, including speech, video, mobile data [18], textual and physiological and biological signals. An interesting research for multimodal sensing in real-time is described in [19]. Then, the **Emotion analysis** module applies emotion recognition techniques (Section 2.4) to classify emotions according to emotion models, being the most popular the categorical and dimensional ones and optionally express the result in an emotion expression language (Section 2.5). **Emotion services or applications** exploit the identified emotions in order to improve user's life. The main applications are [17] emotion awareness and sharing to improve health and mental well-being to encourage social change [20], mental health tracking [21], behaviour change support [22], urban affective sensing to understand the affective relationships of people towards specific places [23] and emotion regulation [24] (Section 2.4).

The adaptation of AmI frameworks to AmE presents a number of challenges because of the multimodal nature of potential emotion sensors and the need for reducing ambiguity of emotion multimodal sources using fusion techniques. In addition, different emotion models are usually used depending on the nature of the emotion sources and the intended application. According to [25], most existing pervasive systems do not consider a multi-modal emotion-aware approach. As previously mentioned, despite the mushrooming of IoT, there are only few experiences in the development of AmE environments that take into account emotional behaviour, and most of them describe prototypes or proofs of concept [10,11,25–29].

From these works, emotion sensing has been addressed using emotion sources such as speech [25,26,29], text [10], video facial and body expression recognition [24] and physiological signals [24]. Few works have addressed the problem of emotion fusion in AmI [24] where a neural multimodal fusion mechanism is proposed. With regard to regulation techniques, fuzzy [24,29] and neurofuzzy controllers [11] have been proposed. Finally, the fields of application have been smart health [24], intelligent classroom [29] and agent-based group decision making [28].

Even though some of the works mention a semantic modelling approach [10], the reviewed approaches propose or use a semantic schema for modelling emotions. Moreover, the lack of semantic modelling of the AmI platform is challenging for integrating new sensors and adapt them to new scenarios. In addition, these works follow a model of full and transparent automation which could leave users feeling out of control [30], without supporting personalization.

2.2. Smart Offices

Although several definitions for smart offices are given in different works [2,31,32], all of them agree in considering a smart office as an environment that supports workers on their daily tasks. These systems use the information collected by different sensors to reason about the environment, and trigger actions which adapt the environment to users' needs by mean of actuators.

Smart offices should be aligned to the business objectives of the enterprise, and should enable a productive environment that maximizes employee satisfaction and company performance. Thus, smart offices should manage efficiently and proactively the IoT infrastructure deployed in the workplace as well as the enterprise systems. Moreover, smart offices should be able to interact with smartphones and help employees to conciliate their personal and professional communications [33].

Focusing on existing solutions whose main goal is the improvement of workers' comfort at the office, Shigeta et al. [34] proposed a smart office system that uses a variety of input devices (such as camera and blood flow sensor) in order to recognize workers' mental and physiological states, and adapts the environment by mean of output devices (such as variable colour light, speaker or aroma generator) for improving workers' comfort. In addition, HealthyOffice [35] deals with a novel mood recognition framework that is able to identify five intensity levels for eight different types of moods, using Silmee TM device to capture physiological and accelerometer data. Li [36] proposed the design of a smart office system that involves the control of heating, illuminating, lighting, ventilating and reconfiguration of the multi-office and the meeting room. With regard to activity recognition, Jalal et al. [37] proposed a depth-based life logging human activity recognition system designed to recognize the daily activities of elderly people, turning these environments into an intelligent space. These works are clear examples of using smart office solutions for improving quality of life, and they propose systems able to perform environment adaption based on users' mental state.

Kumar et al. [38] proposed a semantic policy adaptation technique and its applications in the context of smart building setups. It allows users of an application to share and reuse semantic policies amongst them-selves, based on the concept of context interdependency. Alirezaie et al. [39] presented a framework for smart homes able to perform context activity recognition, and proposed also a semantic model for smart homes. With regard to the use of semantic technologies in the smart office context, Coronato et al. [40] proposed a semantic context service that exploits semantic technologies to support smart offices. This service relies on ontologies and rules to classify several typologies of entities present in a smart office (such as services, devices and users) and to infer higher-level context information from low-level information coming from positioning systems and sensors in the physical environments (such as lighting and sound level).

One of the first mentions of emotion sensor was in the form of affective wearables, by Picard et al. [41]. As for semantic emotion sensors, there is an initial work proposed by Gyrard et al. [42]. However, to the extent of our knowledge, there is no work in the literature that properly addresses the topics of emotion sensors and semantic modelling in a unified smart automation platform. This paper aims to fill this gap, proposing a semantic automation platform that also takes into account users' emotion.

2.3. Emotion Recognition

Over the last years, emotion detection represents a significant challenge that is gaining the attention of a great number of researchers. The main goal is the use of different inputs for carrying out the detection and identification of the emotional state of a subject. Emotion recognition opens endless possibilities as it has wide applications in several fields such as health, emergencies, working life, or commercial sector. The traditional approach of detecting emotions through questionnaires answered by the participants does not yield very efficient methods. That is the reason for focusing on automatic emotion detection using multimodal approaches (i.e., facial recognition, speech analysis and biometric data), as well as ensemble of different information sources from the same mode [43].

Algorithms to predict emotions based on facial expressions are mature and considered accurate. Currently, there are two main techniques to realize facial expression recognition depending on its way of extracting feature data: appearance-based features, or geometry-based features [44]. Both techniques have in common the extraction of some features from the images which are fed into a classification system, and differ mainly in the features extracted from the video images and the classification algorithm used [45]. Geometric based techniques find specific features such as the corners of the mouth, eyebrows, etc. and extracts emotional data from them. Otherwise, appearance based extraction techniques describe the texture of the face caused by expressions, and extract emotional data from skin changes [46].

Emotion recognition from speech analysis is an area that is gaining momentum in recent years [47]. Speech features are divided into for main categories: continuous features (pitch, energy, and formants), qualitative features (voice quality, harsh, and breathy), spectral features (Linear Predictive Coefficients (LPC) and Mel Frequency Cepstral Coefficients (MFCC)), and Teager energy operator-based features (TEO-FM-Var and TEO-Auto-Env) [48].

Physiological signals are another data source for recognizing people's emotions [49]. The idea of wearables that detect the wearer's affective state dates back to the early days of affective computing [41]. For example, skin conductance changes if the skin is sweaty, which is related to stress situations and other affects. Skin conductance is used as an indicator of arousal, to which it is correlated [50]. A low level of skin conductivity suggests low arousal level. Heart rate is also a physiological signal connected with emotions, as its variability increases with arousal. Generally, heart rate is higher for pleasant and low arousal stimuli compared to unpleasant and high arousal stimuli [50].

2.4. Emotion Regulation

Emotion regulation consists in the modification of processes involved in the generation or manifestation of emotion [51], and results an essential component of psychological well-being and successful social functioning. A popular approach to regulate emotions is the use of colour, music or controlled breathing [52,53].

Xin et al. [54,55] demonstrated that colour characteristics such as chroma, hue or lightness produce an impact on emotions. Based on these studies and on the assumption of the power of colour to change mood, Sokolova et al. [52] proposed the use of colour to regulate affect. Participants of this study indicated that pink, red, orange and yellow maximized their feeling of joy, while sadness correlates with dark brown and gray. Ortiz-García-Cervigón et al. [56] proposed an emotion regulation system at home, using RGB LED strips that are adjustable in colour and intensity to control the ambience. This study reveals that warm colours are rated as more tensed, hot, and less preferable for lighting, while cold colours are rated as more pleasant.

With regard to music, several studies [57,58] show that listening to music influences mood and arousal. Van der Zwaag [59] found that listening to preferred music significantly improved performance on high cognitive demand tasks, suggesting that music increases efficiency for cognitive tasks. Therefore, it has been demonstrated that listening to music can influence regulation abilities, arousing certain feelings or helping to cope negative emotions [60]. In addition, it has been demonstrated that different types of music may have different demands on attention [61].

The commented studies show that the adaptation of ambient light colour and music are considerable solutions for regulating emotions in a smart office environment, as this adaptation may improve workers' mood and increase their productivity and efficiency.

2.5. Semantic Modelling

Semantic representation considerably improves interoperability and scalability of the system, as it provides a rich machine-readable format that can be understood, reasoned about, and reused.

To exchange information between independent systems, a set of common rules need to be established, such as expected formats, schemas and expected behaviour. These rules usually take the form of an API (application programming interface). In other words, systems need not only to define **what** they are exchanging (concepts and their relationship), but also **how** they represent this information (representation formats and models). Moreover, although these two aspects need to be in synchrony, they are not unambiguously coupled: knowing how data are encoded does not suffice to know what real concepts the refer to, and vice versa.

The semantic approach addresses this issue by replacing application-centric ad-hoc models and representation formats with a formal definition of the concepts and relationships. These definitions are known as ontologies or vocabularies. Each ontology typically represents one domain in detail, and they borrow concepts from one another whenever necessary [62]. Systems then use parts of several ontologies together to represent the whole breadth of their knowledge. Moreover, each concept and instance (entity) is unambiguously identified. Lastly, the protocols, languages, formats and conventions used to model, publish and exchange semantic information are standardized and well known (SPARQL, RDF, JSON-LD, etc.) [63–65].

This work merges two domains: rule-based systems and emotions. We will explore the different options for semantic representation in each domain.

There are plenty of options for modelling and implementing rule-based knowledge, such as RuleML [66], Semantic Web Rule Language (SWRL) [67], Rule Interchange Format (RIF) [68], SPARQL Inferencing Notation (SPIN) [69] and Notation 3 (N3) Logic [70].

EWE [13] is a vocabulary designed to model, in a descriptive approach, the most significant aspects of Task Automation Service (TAS). It has been designed after analyzing some of the most relevant TASs [71] (such as Ifttt, Zapier, Onx, etc.) and provides a common model to define and describe them. Based on a number of identified perspectives (privacy, input/output, configurability, communication, discovery and integration), the main elements of the ontology have been defined, and formalized in an ontology. Moreover, extensive experiments have been developed to transform the automation of these systems into the proposed ontology. Regarding inferences, EWE is based on OWL2 classes and there are implementations of EWE using a SPIN Engine (TopBraid (https://www.w3.org/2001/sw/wiki/TopBraid)) and N3 Logic (EYE (http://eulersharp.sourceforge.net/)).

Four major classes make up the core of EWE: *Channel*, *Event*, *Action* and *Rule*. The class *Channel* defines individuals that either generate *Events*, provide *Actions*, or both. In the smart office context, sensors and actuators such as an emotion detector or a smart light are described as channels, which produce events or provide actions. The class *Event* defines a particular occurrence of a process, and allows users to describe under which conditions should rules be triggered. These conditions are the configuration parameters, and are modelled as input parameters. Event individuals are generated by a certain channel, and usually provide additional details. These additional details are modelled as output parameters, and can be used within rules to customize actions. The recognition of sadness generated by the emotion detector sensor is an example of entity that belongs to this class. The class *Action* defines an operation provided by a channel that is triggered under some conditions. Actions provides effects whose nature depends on itself, and can be configured to react according to the data collected from an event by means of input parameters. Following the smart office context mentioned above, to change the light colour is an example of action generated by the smart light channel. Finally, the class *Rule* defines an *ECA*, triggered by an event that produces the execution of an action. An example of rule is: "*If sadness is detected, then change the light colour*".

There are also different options for emotion representation. EmotionML [72] is one of the most notable general-purpose emotion annotation and representation languages that offers twelve vocabularies for categories, appraisals, dimensions and action tendencies. However, as shown in previous works [73], the options for semantic representation are limited to a few options, among which we highlight the Human Emotion Ontology (HEO) [74], and Onyx [73], a publicly available ontology for emotion representation. Among these two options, we chose Onyx for several reasons: it is compatible with EmotionML; it tightly integrates with the Provenance Ontology [75], which gives us the ability to reason about the origin of data annotations; and it provides a meta-model for emotions, which enables anyone to publish a new emotion model of their own while remaining semantically valid, thus enabling the separation of representation and psychological models. The latter is of great importance, given the lack of a standard model for emotions. In EmotionML, emotion models are also separated from the language definition. A set of commonly used models is included as part of the vocabularies for EmotionML [76], all of which are included in Onyx.

Moreover, the Onyx model provides a model for emotion conversion, and a set of existing conversions between well known models. Including conversion as part of the model enables the integration of data using different models. Two examples of this would be working with emotion readings from different providers, or fusing information from different modalities (e.g., text and audio), which typically use different models. It also eases a potential migration to a different model in the future.

In addition, Onyx has been extended to cover multimodal annotations [77,78]. Lastly, the Onyx model has been embraced by several projects and promoted by members of the Linked Data Models for Emotion and Sentiment Analysis W3C Community Group [79].

There are three main concepts in the Onyx ontology that are worth explaining, as they are used in the examples in following sections. They are: *Emotion*, *EmotionAnalysis* and *EmotionSet*. They relate to each other in the following way: an EmotionAnalysis process annotates a given entity (e.g., a piece of text or a video segment) with an EmotionSet, and an EmotionSet is in turn comprised of one or more Emotions. Due to the provenance information, it is possible to track the EmotionAnalysis that generated the annotation.

3. Semantic Modelling for the Smart Office Environment

With the purpose of applying a semantic layer to the emotion aware automation system, several vocabularies and relationships between ontologies have been designed. This enables the semantic modelling of all entities in the smart office environment. Figure 1 shows the relationships between the used ontologies described above.

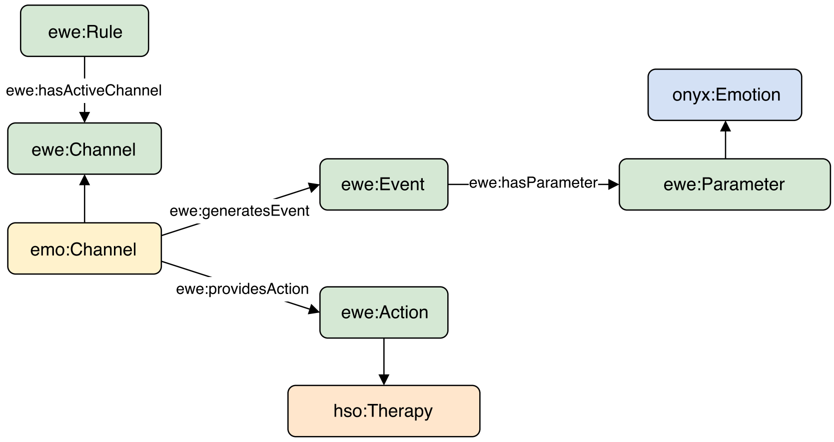


Figure 1. Main classes of the ontologies involved in the semantic modelling.

Automation rules (*ewe:Rule*) are modelled using EWE ontology [13], which presents them in event-condition-action form. Events (*ewe:Event*) and actions (*ewe:Action*) are generated by certain channels. In the proposed architecture, there are different channels that either generate events, provide actions, or both. The class *ewe:Channel* has been subclassed to provide an emotional channel class (*emo:Channel*), which is responsible for generating events and actions related to the emotion recognition and regulation. From this class, the channels *emo:EmotionSensor* and *emo:EmotionRegulator* have been defined. The former is responsible for generating events related to the emotion detection, while the later is responsible for providing certain actions that have the purpose of regulating the emotion. These two classes group all sensors and actuators able to detect or regulate emotions, but should be subclassed by classes representing each device concretely. In addition, events and actions may have parameters. The *emo:EmotionDetected* event has as Parameter the detected emotion. Emotions are modelled using Onyx [73], as described in Section 2.5, so the parameter must subclass *onyx:Emotion*.

The *emo:EmotionRegulator* channel can be subclassed for defining a *SmartSpeaker* or a *SmartLight*, able to provide actions to regulate the emotion such as *emo:PlayRelaxingMusic* or *emo:ChangeAmbientColor*, respectively. The action of playing relaxing music has as parameter (*ewe:Parameter*) the song to be played, while the action of change ambient colour has as parameter the colour to which the light must change. In addition, all these actions are also represented as therapies using Human Stress Ontology (HSO) ontology [80], so *hso:Therapy* has been subclassed. To give a better idea of how specific Channels, Events and Actions have been modelled; Table 1 shows the commented example written in Notation3, describing all its actions with their corresponding parameters.

An example of event and action instances with grounded parameters, which are based on the concepts defined in the listing given above, is presented in Table 2. This table describes the definition of sadness and the actions of playing music and changing ambient colour.

Similarly, automation rules are described using the punning mechanism to attach classes to properties of Rule instances. In the example shown in Table 3, the rule instance describes a rule that is triggered by the event of *sad emotion detection* and produces the action of *changing ambient colour to green* (both defined in Table 2).

Table 1. Semantic representation of Emotion Regulator channel written in Notation3.

```
emo:SmartSpeaker a owl:Class ;
    rdfs:label ‘‘Smart Speaker’’ ;
    rdfs:comment ‘‘This channel represents a smart speaker.’’ ;
    rdfs:subClassOf emo:EmotionRegulator .

emo:PlayRelaxingMusic a owl:Class ;
    rdfs:label ‘‘Play relaxing music’’ ;
    rdfs:comment ‘‘This action will play relaxing music.’’ ;
    rdfs:subclassOf ewe:Action ;
    rdfs:subclassOf hso:Therapy ;
    rdfs:domain emo:SmartSpeaker .

emo:SmartLight a owl:Class ;
    rdfs:label ‘‘Smart Light’’ ;
    rdfs:comment ‘‘This channel represents a smart light.’’ ;
    rdfs:subClassOf emo:EmotionRegulator .

emo:ChangeAmbientColor a owl:Class ;
    rdfs:label ‘‘Change ambient color’’ ;
    rdfs:comment ‘‘This action will change ambient color.’’ ;
    rdfs:subclassOf ewe:Action ;
    rdfs:subclassOf hso:Therapy ;
    rdfs:domain emo:SmartLight .
```

Table 2. Event and action instances.

```
:sad-emotion-detected a emo:EmotionDetected ;
    ewe:hasEmotion onyx:sadness .

:play-music a emo:PlayRelaxingMusic ;
    ewe:hasSong ‘‘the title of the song to be played’’ .

:change-ambient-color-green a emo:ChangeAmbientColor ;
    ewe:hasColor dbpedia:Green .
```

Table 3. Rule instance.

```
:regulate-stress a ewe:Rule ;
    dcterms:title     ‘‘Stress regulation rule’’^xsd:string ;
    ewe:triggeredByEvent     :sad-emotion-detected ;
    ewe:firesAction          :change-ambient-color-greenr .
```

4. Emotion Aware Task Automation Platform Architecture

The proposed architecture was designed based on the reference architecture for TAS [81],which was extended to enable emotion awareness. The system is divided into two main blocks: **emotional context recognizer** and **emotion aware task automation server**, as shown in Figure 2. Emotional context recognizer aims to detect and recognize users’ emotions and information related to context or Internet services and send them to the automation platform to trigger the corresponding actions.

The automation system that receives these data is a semantic event-driven platform that receives events from several sources and performs the corresponding actions. In addition, it provides several functions for automating tasks by means of semantic rules and integrates different devices and services.

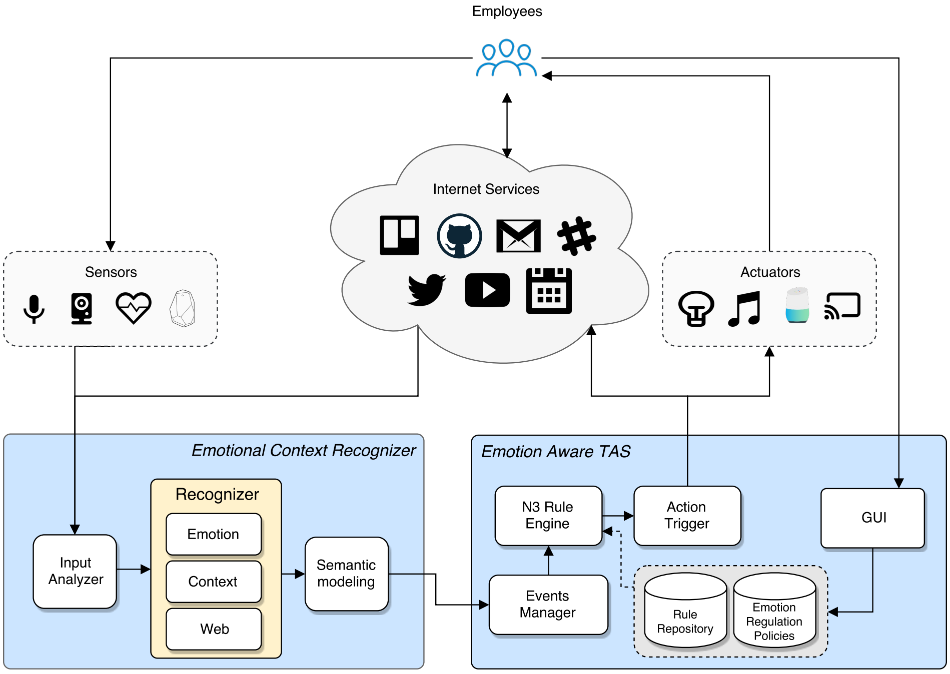


Figure 2. Emotion Aware Automation Platform Architecture.

4.1. Emotional Context Recognizer

The emotional context recognizer block is responsible for detecting users' emotions and contextual events, encoding emotions and events using semantic technologies, and sending these data to the automation platform, where they are evaluated. The block consists of three main modules: input analyzer, recognizer and semantic modelling. In addition, each module is composed of multiple independent and interchangeable sub-modules that provide the required functions, with the purpose of making the system easy to handle.

The input analyzer receives data from sensors involved in emotion and context recognition (such as camera, microphone, wearables or Internet services) and its pre-processing. With this purpose, the input analyzer is connected with the mentioned sensors, and the received data are sent to the recognizer module. The recognizer module receives data captured by the input analyzer. It consists in a pipeline with several submodules that perform different analysis depending on the source of the information. In the proposed architecture, there are three sub-modules: emotion recognizer, context recognizer and web recognizer. The emotion recognizer module provides functions for extracting emotions by means of real time recognition of facial expression, speech and text analysis, and biometric data monitoring; the context recognizer provides functions for extracting context data from sensors (e.g., temperature amd humidity); and the web recognizer provides functions for extracting information from Internet services. Once data have been extracted, they are sent to the semantic modelling module. The main role of semantic modelling is the application of a semantic layer (as described in Section 3), generating the semantic events and sending them to the automation platform.

4.2. Emotion Aware Task Automation Server

The automation block consists in an intelligent automation platform based on semantic ECA rules. The main goal is to enable semantic rule automation in a smart environment, allowing the user to configure custom automation rules or to import rules created by other users in an easy way. In addition, it provides integration with several devices and services such as a smart TV, Twitter, Github, etc., as well as an easy way for carrying out new integrations.

The platform handles events coming from different sources and triggers accordingly the corresponding actions generated by the rule engine. In addition, it includes all the functions for managing automation rules and the repositories where rules are stored, as well as functions for creating and editing channels. With this purpose, the developed platform is able to connect with several channels for receiving events, evaluating them together with stored rules and performing the corresponding actions.

To enable the configuration and management of automation rules, the platform provides a *graphical user interface* (GUI) where users can easily create, remove or edit rules. The GUI connects with the rule administration module, which is responsible for handling the corresponding changes in the repositories. There are two repositories in the platform: *rule repository*, where information about rules and channels is stored; and *emotion regulation policies repository*. The policies are sets of rules which aim to regulate the emotion intensity in different contexts. In the smart office context proposed, they are intended to regulate negative emotions to maximize productivity. The rules may be aimed towards automating aspects such as: ambient conditionsto improve the workers' comfort; work related tasks to improve efficiency; or the rules could adjust work conditions to improve productivity. Some examples of these rules are presented below:

(a) *If stress level of a worker is too high, then reduce his/her task number*. When a very high stress level in a worker has been detected, this rule proposes reducing his/her workload to achieve that his/her stress level falls and his/her productivity rises.

(b) *If temperature rises above 30 °C, then turn on the air conditioning*. To work at high level of temperatures may result in workers' stress, so this rule proposes to automatically control this temperature in order to prevent high levels of stress.

(c) *If average stress level of workers is too high, then play relaxing music*. If most workers have a high stress value, the company productivity will significantly fall. Thus, this rule proposes to play relaxing music in order to reduce the stress level of workers.

In addition, the company human resources department may implement their own emotion regulation policies to adjust the system to their own context. The system adapts rules based on channel description. Rule adaptation is based on identifying if the smart environment includes the channels used by a certain rule. The system detects available channels of the same channel class used by the rule and request confirmation from the user to included the "adapted rule". This enables the adaptation of rules to different channel providers, which can be physical sensors (i.e., different beacons) or internet services (i.e., Gmail and Hotmail). The EWE ontology allows us this adaptation by mean of OWL2 punning mechanism for attaching properties to channels [13].

With regards to event reception, these are captured by the *events manager* module, which sends them to the rule engine to be evaluated along with the stored rules. The rule engine module is a semantic engine reasoner [82] based on an ontology model. It is responsible for the reception of events from the *events manager* and the load of rules that are stored in the repository. When a new event is captured and the available rules are loaded, the reasoner runs the ontology model inferences and the actions based on the incoming events and the automation rules are drawn. These actions are sent to the *action trigger*, which connects to the corresponding channels to perform the actions.

The semantic integration of sensors and services is done based on the notion of adapters [83,84], which interact with both sensors and internet services, providing a semantic output. Adapters, as well as mobile clients, are connected to the rule engine through Crossbar.io (https://crossbar.io/), and IoT

Middleware that provides both REST-through Web Application Messaging Protocol (WAMP)- and Message Queuing Telemetry Transport (MQTT) interfaces.

Finally, the implementation of this architecture, called EWETasker, was made using PHP for the server, HTML/JavaScript for the web client (including the GUI), and Android SDK for a mobile client. The implementation was based on N3 technology and EYE reasoning engine (http://n3.restdesc.org/). Several sensors and services have already been integrated into EWETasker suitable for the smart office use case. In particular, EWETasker supports indoor and temperature sensors (Estimote bluetooth beacons (https://estimote.com)), smart object sensor (Estimote bluetooth stickers), electronic door control based on Arduino, video emotion sensors (based on Emotion Research Lab), social network emotion sensor (Twitter), and mobile-phone sensors (Bluetooth, location, wifi, etc.). With regards to corporate services, several services oriented to software consultancy firms have been integrated for collaboration (Twitter, GMail, Google Calendar, and Telegram) and software development (Restyaboard Scrum board (http://www.restya.com), GitHub (https://github.com) and Slack (https://slack.com)).

5. Experimentation

As already stated, the main experimental contribution of this work was the design and implementation of an emotion aware automation platform for smart offices. In this way, we raised four hypotheses regarding the effectiveness of the proposed system:

- H1: The use of the proposed platform regulates the emotional state of a user that is under stressful conditions.
- H2: The actions taken by the proposed platform do not disturb the workflow of the user.
- H3: The use of the proposed system improves user performance.
- H4: The use of the system increases user satisfaction.

To evaluate the proposed system with respect to these hypotheses, an experiment with real users was performed. For this experiment, a prototype of the proposed system was deployed, which includes the following components. The emotion of the participants was detected from a webcam feed, which feeds a video-based emotion recognizer. As for the semantic layers of the system, the events manager, rule engine and action trigger were fully deployed. Finally, the actuators implemented both hearing and visual signals using a variety of devices. Detailed information on materials is given in Section 5.2. This section covers the design, results and conclusions drawn from the experiment, focusing on its scope.

5.1. Participants

The experiment included 28 participants. Their ages ranged from 18 to 28 years, all of them university students with technical background, of both genders. All of them were unaware of this work, and no information regarding the nature of the experiment was given to the participants beforehand. Since the proposed system is primarily oriented to technical work positions, this selection is oriented to validate the system with participants that are currently working in technical environments, or will in the future.

5.2. Materials

The material used for this experiment is varied, as the proposed automation system needs several devices to properly function. Regarding the deployment of the automation system, the TAS ran in a commodity desktop computer, with sufficient CPU and memory for its execution. The same environment was prepared for the emotion recognizer system. For the sensors and actuators, the following were used:

- Emotion Research software (https://emotionresearchlab.com/). This module provides facial mood detection and emotional metrics that are fed to the automation system. This module is an implementation that performs emotion classification in two main steps: (i) it makes use of Histogram of Oriented Gradients (HOG) features that are used to train with a SVM classifier in order to localize face position in the image; and (ii) the second step consists in a normalization process of the face image, followed by a Multilayer Perceptron that implements the emotion classification. Emotion Research reports 98% accuracy in emotion recognition tasks.

- A camera (Gucee HD92) feeds the video to the emotion recognizer submodule.
- Room lighting (WS2812B LED strip controlled by WeMos ESP8266 board) is used as an actuator on the light level of the room, with the possibility of using several lighting patterns.
- Google Chromecast [85] transmits content in a local computer network.
- LG TV 49UJ651V is used for displaying images.
- Google Home is used for communicating with the user. In this experiment, the system can formulate recommendations to the user.

Participants accessed the web HTML-based interface using a desktop computer with the Firefox browser (https://www.mozilla.org/en-US/firefox/desktop/).

5.3. Procedure

During the experiment, each participant performed a task intended to keep the participant busy for approximately 10 min. This task consisted in answering a series of basic math related questions that were presented to the participant via a web interface (e.g., "Solve $24 \cdot 60 \cdot 60$"). We used a set of 20 questions of similar difficulty that have been designed so that any participant can answer them within 30 s. The use of a web-based interface allowed us to programmatically perform the session, and to record metrics associated with the experiment.

The workflow of the experiment is as follows. Each participant's session is divided into two parts. In each part of the session half of the task questions are sequentially prompted to the participant by the examiner system. Simultaneously, the automation system is fed with the information provided by the different sensors that are continually monitoring the participant emotional state. The experiment finishes when all the questions have been answered. In addition, a questionnaire is given to the participants just after the sessions concludes. These questions are oriented to offer the participant's view of the system. The raised questions are summarized in Table 4. Questions Q2, Q3, Q4 and Q5 are asked twice, once in regard to the no automation part, and the other time in relation to the part with the automation enabled. Questions Q1 and Q2 are designed so that a check of internal consistency is possible; as, if results from these two questions were to disagree, the experiment would be invalid [86].

Table 4. Questions raised to the participants at the end of the session.

No.	Hypothesis	Question Formulation
Q1	H1, H2	In which section have you been more relaxed?
Q2	H1, H2	What is your comfort level towards the environment?
Q3	H3	Do you think the environment's state has been of help during the completion of the task?
Q4	H4	Would you consider beneficial to work in this environment?
Q5	H4	What is your overall satisfaction with relation to the environment?

The workflow of the system in the context of the experiment is as follows. While the participant is performing the task, the emotion sensor is continuously monitoring the participant's emotional state. The emotion sensor uses the camera as information input, while the Google Home is used when the user communicates with the system. This emotion-aware data are sent to the TAS, which allows the system the have continuous reports. The TAS receives, processes, and forwards these events to the N3 rule engine. Programmed rules are configured to detect changes in the participant emotional state, acting accordingly. As an example, a shift of emotion, such as the change from happy to sad, is detected by the rule engine which triggers the relaxation actions. If a certain emotion regulation rule is activated, the corresponding action is then triggered through the communication to the action trigger module, which causes the related actuators to start its functioning. The configured actions are aimed at relaxing and regulating the emotion of the participant, so that the performance in the experiment task is improved, as well as the user satisfaction. The actions configured for this experiment are: (i) relaxation recommendations done by the Google Home, such as a recommendation to take a brief walk for two minutes; (ii) lighting patterns using coloured lights that slowly change its intensity and colour; and (iii) relaxing imagery and music that are shown to the user via the TV. A diagram of this deployment is shown in Figure 3.

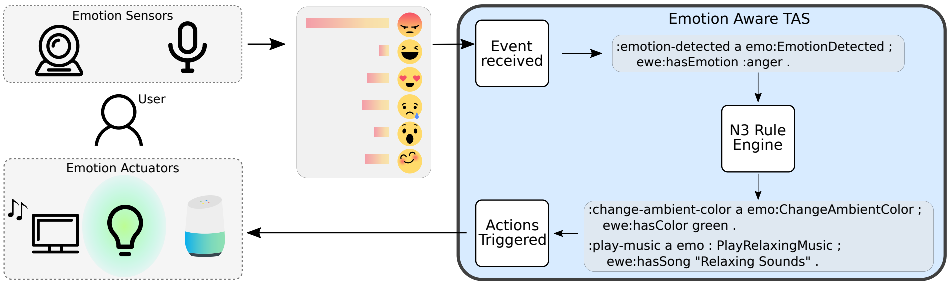


Figure 3. Deployment for the experiment.

While the participants are performing the proposed task, the actions of the automation system are controlled. During half of each session, the automation is deactivated, while, during the other half, the action module is enabled. With this, we can control the environmental changes performed by the automation system, allowing its adaptation at will.

Another interesting aspect that could be included is the integration of learning policies based on employee's emotional state. A related work that models learning policies and their integration with Enterprise Linked Data is detailed in [87].

5.4. Design

The experiment was a within-subject design. As previously stated, the controlled factor is the use of the automation system, which has two levels, activated and not activated. The automation use factor is counterbalanced using a Latin square so that the participants are divided into two groups. One group performs the first half of the session without the automation system, while, for the second half of the session, the system is used. The other group performs the task inversely.

5.5. Results and Discussion

To tackle Hypothesis 1 and Hypothesis 2, Questions 1 and 2 were analyzed. Regarding Question 1, 18 respondents declared that the section with the adaptation system enabled was the most relaxing for them. In contrast, seven users claimed that for them the most relaxing section of the experiment was that without the adaptation system. The results from Question 1 suggest that users prefer to use the adaptation system, although it seems that this is not the case for all the users. Regarding the Question 2, results show that the average in the adaptation part (3.5) is higher than with no adaptation whatsoever (2.5), as shown in Figure 4. An ANOVA analysis shows that this difference is statistically significant ($p = 0.015 < 0.05$). These results support H1 and H2, concluding that users feel more inclined to use the adaptation system rather than performing the task without adaptation.

Following, Question 3 addressed Hypothesis 3. The analysis of the results of this question reveals that users point higher the usefulness of the environment adaptation for the completion of the task, as shown in Figure 4. While the average for the adaptation section is 3.93, it is 2.07 for the no adaptation part. Through ANOVA, we see that this difference is considerably significant ($p = 2.96 \times 10^{-6} < 0.05$). As expected, Hypothesis 3 receives experimental support, indicating that the use of the automation system can improve the performance of the user in a certain task, as perceived by the users.

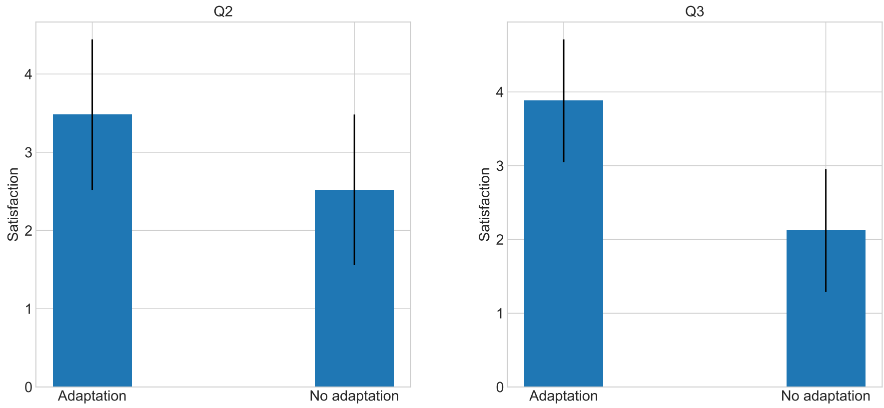


Figure 4. Results for Q2 and Q3.

In relation to Hypothesis 4, both Questions 4 and 5 are aimed to check its validity. As can be seen in Figure 5, users consider more beneficial to work with the adaptation system enabled. The average measure for the adaptation is 3.78, while the no adaptation environment is considered lower on average, with 2.21. Once again, the ANOVA test outputs a significant difference between the two types of environment ($p = 0.0002 < 0.05$). With regard to Question 5, the average for the satisfaction with the adapted environment is 3.83; in contrast, the satisfaction with the no adaptation environment is 2.17, as shown in Figure 5. After performing an ANOVA test, we see that this difference is greatly significant ($p = 1.02 \times 10^{-10} < 0.05$). Attending to this, users seem to consider the adaptation system for their personal workspace, and at the same time, they exhibit a higher satisfaction with an adapted work environment. These data indicate that Hypothesis 4 is true, and that users positively consider the use of the adaptation system.

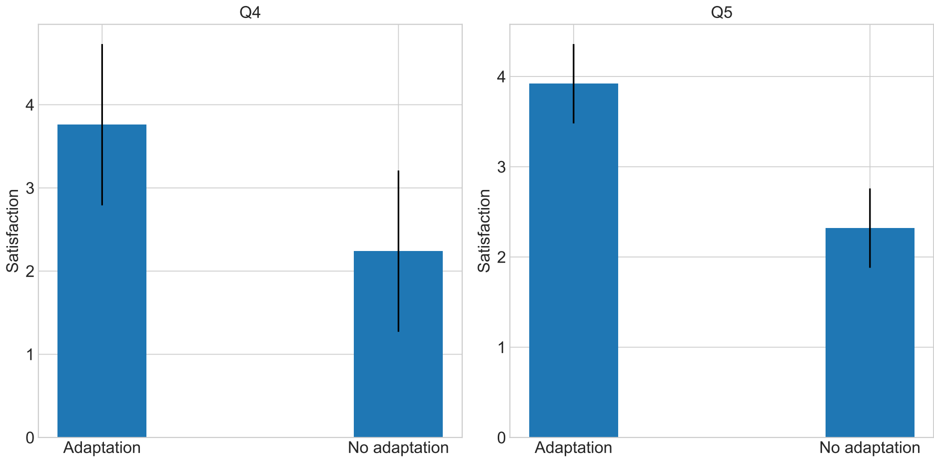


Figure 5. Results for Q4 and Q5.

6. Conclusions and Outlook

This paper presents the architecture of an emotion aware automation platform based on semantic event-driven rules, to enable the automated adaption of the workplaces to the need of the employees. The proposed architecture allows users to configure their own automation rules based on their emotions to regulate these emotions and improve their wellbeing and productivity. In addition, the architecture is based on semantic event-driven rules, so this article also describes the modelling of all components of the system, thus enabling data interoperability and portability of automations. Finally, the system was implemented and evaluated in a real scenario.

Through the experimentation, we verified a set of hypotheses. In summary: (i) using the proposed automation system helps to regulate the emotional state of users; (ii) adaptations of the automation system do not interrupt the workflow of users; (iii) the proposed system improves user performance in a work environment; and, finally, (iv) the system increases user satisfaction. These results encourage the use and improvement of this kind of automation systems, as they seem to provide users with a number of advantages, such as regulation of stress and emotions, and personalized work spaces.

As future work, there are many lines that can be followed. One of these lines is the application of the proposed system to other scenarios different from smart offices. The high scalability offered by the developed system facilitates the extension of both the architecture and the developed tools with the purpose of giving a more solid solution to a wider range of scenarios. Currently, we are working on its application to e-learning and e-commerce scenarios. In addition, another line of future work is the recognition of the activity, as it is useful to know the activity related to the detected emotion of the user.

Furthermore, we also plan to develop a social simulator system based on emotional agents to simplify the test environment. This system will enable testing different configurations and automations of the smart environment before implementing them in a real scenario, resulting in an important reduction of costs and efforts in the implementation.

Author Contributions: S.M. and C.A.I. originally conceived the idea; S.M. designed and developed the system; S.M., O.A. and J.F.S. designed the experiments; S.M. and O.A. performed the experiments; O.A. analyzed the data; O.A. contributed analysis tools; and all authors contributed to the writing of the paper.

Funding: This work was funded by Ministerio de Economía y Competitividad under the R&D projects SEMOLA (TEC2015-68284-R) and EmoSpaces (RTC-2016-5053-7), and by the Regional Government of Madrid through the project MOSI-AGIL-CM (grant P2013/ICE-3019, co-funded by EU Structural Funds FSE and FEDER).

Acknowledgments: The authors express their gratitude to Emotion Research Lab team for sharing their emotion recognition product for this research work. This work is supported by the Spanish Ministry of Economy and Competitiveness under the R&D projects SEMOLA (TEC2015-68284-R) and EmoSpaces (RTC-2016-5053-7), by the Regional Government of Madrid through the project MOSI-AGIL-CM (grant P2013/ICE-3019, co-funded by EU Structural Funds FSE and FEDER).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Miorandi, D.; Sicari, S.; Pellegrini, F.D.; Chlamtac, I. Internet of things: Vision, applications and research challenges. *Ad Hoc Netw.* **2012**, *10*, 1497–1516. [CrossRef]
2. Augusto, J.C. *Ambient Intelligence: The Confluence of Ubiquitous/Pervasive Computing and Artificial Intelligence*; Springer: London, UK; pp. 213–234.
3. Gutnik, L.A.; Hakimzada, A.F.; Yoskowitz, N.A.; Patel, V.L. The role of emotion in decision-making: A cognitive neuroeconomic approach towards understanding sexual risk behavior. *J. Biomed. Inform.* **2006**, *39*, 720–736. [CrossRef] [PubMed]

4. Kok, B.E.; Coffey, K.A.; Cohn, M.A.; Catalino, L.I.; Vacharkulksemsuk, T.; Algoe, S.B.; Brantley, M.; Fredrickson, B.L. How Positive Emotions Build Physical Health : Perceived Positive Social Connections Account for the Upward Spiral Between Positive Emotions and Vagal Tone. *Psychol. Sci.* **2013**, *24*, 1123–1132. [CrossRef] [PubMed]
5. Nguyen, V.T.; Longin, D.; Ho, T.V.; Gaudou, B. Integration of Emotion in Evacuation Simulation. In *Information Systems for Crisis Response and Management in Mediterranean Countries, Proceedings of the First International Conference, ISCRAM-med 2014, Toulouse, France, 15–17 October 2014*; Springer International Publishing: Cham, Switzerland, 2014; pp. 192–205.
6. Pervez, M.A. Impact of emotions on employee's job performance: An evidence from organizations of Pakistan. *OIDA Int. J. Sustain. Dev.* **2010**, *1*, 11–16.
7. Weiss, H.M. Introductory Comments: Antecedents of Emotional Experiences at Work. *Motiv. Emot.* **2002**, *26*, 1–2. [CrossRef]
8. Bhuyar, R.; Ansari, S. Design and Implementation of Smart Office Automation System. *Int. J. Comput. Appl.* **2016**, *151*, 37–42. [CrossRef]
9. Van der Valk, S.; Myers, T.; Atkinson, I.; Mohring, K. Sensor networks in workplaces: Correlating comfort and productivity. In Proceedings of the 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Singapore, 7–9 April 2015; pp. 1–6.
10. Zhou, J.; Yu, C.; Riekki, J.; Kärkkäinen, E. AmE framework: A model for emotion-aware ambient intelligence. In Proceedings of the second international conference on affective computing and intelligent interaction (ACII2007): Doctoral Consortium, Lisbon, Portugal, 12–14 September 2007; p. 45.
11. Acampora, G.; Loia, V.; Vitiello, A. Distributing emotional services in ambient intelligence through cognitive agents. *Serv. Oriented Comput. Appl.* **2011**, *5*, 17–35. [CrossRef]
12. Beer, W.; Christian, V.; Ferscha, A.; Mehrmann, L. Modeling Context-Aware Behavior by Interpreted ECA Rules. In *Euro-Par 2003 Parallel Processing*; Kosch, H., Böszörményi, L., Hellwagner, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 1064–1073.
13. Coronado, M.; Iglesias, C.A.; Serrano, E. Modelling rules for automating the Evented WEb by semantic technologies. *Expert Syst. Appl.* **2015**, *42*, 7979–7990. [CrossRef]
14. Muñoz, S.; Fernández, A.; Coronado, M.; Iglesias, C.A. Smart Office Automation based on Semantic Event-Driven Rules. In Proceedings of the Workshop on Smart Offices and Other Workplaces, Colocated with 12th International Conference on Intelligent Environments (IE'16), London, UK, 14–16 September 2016; Ambient Intelligence and Smart Environments; IOS Press: Clifton, VA, USA, 2016; Volume 21, pp. 33–42.
15. Inada, T.; Igaki, H.; Ikegami, K.; Matsumoto, S.; Nakamura, M.; Kusumoto, S. Detecting Service Chains and Feature Interactions in Sensor-Driven Home Network Services. *Sensors* **2012**, *12*, 8447–8464. [CrossRef] [PubMed]
16. Zhou, J.; Kallio, P. Ambient emotion intelligence: from business awareness to emotion awareness. In Proceedings of the 17th International Conference on Systems Research, Berlin, Germany, 15–17 April 2014; pp. 47–54.
17. Kanjo, E.; Al-Husain, L.; Chamberlain, A. Emotions in context: Examining pervasive affective sensing systems, applications, and analyses. *Pers. Ubiquitous Comput.* **2015**, *19*, 1197–1212. [CrossRef]
18. Kanjo, E.; El Mawass, N.; Craveiro, J. Social, disconnected or in between: Mobile data reveals urban mood. In Proceedings of the 3rd International Conference on the Analysis of Mobile Phone Datasets (NetMob'13), Cambridge, MA, USA, 1–3 May 2013.
19. Wagner, J.; André, E.; Jung, F. Smart sensor integration: A framework for multimodal emotion recognition in real-time. In Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam, The Netherlands, 10–12 September 2009; pp. 1–8.
20. Gay, G.; Pollak, J.; Adams, P.; Leonard, J.P. Pilot study of Aurora, a social, mobile-phone-based emotion sharing and recording system. *J. Diabetes Sci. Technol.* **2011**, *5*, 325–332. [CrossRef] [PubMed]
21. Gaggioli, A.; Pioggia, G.; Tartarisco, G.; Baldus, G.; Corda, D.; Cipresso, P.; Riva, G. A mobile data collection platform for mental health research. *Pers. Ubiquitous Comput.* **2013**, *17*, 241–251. [CrossRef]
22. Morris, M.E.; Kathawala, Q.; Leen, T.K.; Gorenstein, E.E.; Guilak, F.; Labhard, M.; Deleeuw, W. Mobile therapy: Case study evaluations of a cell phone application for emotional self-awareness. *J. Med. Internet Res.* **2010**, *12*, e10. [CrossRef] [PubMed]

23. Bergner, B.S.; Exner, J.P.; Zeile, P.; Rumberg, M. Sensing the city—How to identify recreational benefits of urban green areas with the help of sensor technology. In Proceedings of the REAL CORP 2012, Schwechat, Austria, 14–16 May 2012.
24. Fernández-Caballero, A.; Martínez-Rodrigo, A.; Pastor, J.M.; Castillo, J.C.; Lozano-Monasor, E.; López, M.T.; Zangróniz, R.; Latorre, J.M.; Fernández-Sotos, A. Smart environment architecture for emotion detection and regulation. *J. Biomed. Inform.* **2016**, *64*, 55–73. [CrossRef] [PubMed]
25. Jungum, N.V.; Laurent, E. Emotions in pervasive computing environments. *Int. J. Comput. Sci. Issues* **2009**, *6*, 8–22.
26. Bisio, I.; Delfino, A.; Lavagetto, F.; Marchese, M.; Sciarrone, A. Gender-driven emotion recognition through speech signals for ambient intelligence applications. *IEEE Trans. Emerg. Top. Comput.* **2013**, *1*, 244–257. [CrossRef]
27. Acampora, G.; Vitiello, A. Interoperable neuro-fuzzy services for emotion-aware ambient intelligence. *Neurocomputing* **2013**, *122*, 3–12. [CrossRef]
28. Marreiros, G.; Santos, R.; Novais, P.; Machado, J.; Ramos, C.; Neves, J.; Bula-Cruz, J. Argumentation-based decision making in ambient intelligence environments. In Proceedings of the Portuguese Conference on Artificial Intelligence, Guimarães, Portugal, 3–7 December 2007; pp. 309–322.
29. Hagras, H.; Callaghan, V.; Colley, M.; Clarke, G.; Pounds-Cornish, A.; Duman, H. Creating an ambient-intelligence environment using embedded agents. *IEEE Intell. Syst.* **2004**, *19*, 12–20. [CrossRef]
30. Mennicken, S.; Vermeulen, J.; Huang, E.M. From today's augmented houses to tomorrow's smart homes: New directions for home automation research. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 105–115.
31. Cook, D.; Das, S. *Smart Environments: Technology, Protocols and Applications (Wiley Series on Parallel and Distributed Computing)*; Wiley-Interscience: Seattle, WA, USA, 2004.
32. Marsa-maestre, I.; Lopez-carmona, M.A.; Velasco, J.R.; Navarro, A. Mobile Agents for Service Personalization in Smart Environments. *J. Netw.* **2008**, *3*. [CrossRef]
33. Furdik, K.; Lukac, G.; Sabol, T.; Kostelnik, P. The Network Architecture Designed for an Adaptable IoT-based Smart Office Solution. *Int. J. Comput. Netw. Commun. Secur.* **2013**, *1*, 216–224.
34. Shigeta, H.; Nakase, J.; Tsunematsu, Y.; Kiyokawa, K.; Hatanaka, M.; Hosoda, K.; Okada, M.; Ishihara, Y.; Ooshita, F.; Kakugawa, H.; et al. Implementation of a smart office system in an ambient environment. In Proceedings of the 2012 IEEE Virtual Reality Workshops (VRW), Costa Mesa, CA, USA, 4–8 March 2012; pp. 1–2.
35. Zenonos, A.; Khan, A.; Kalogridis, G.; Vatsikas, S.; Lewis, T.; Sooriyabandara, M. HealthyOffice: Mood recognition at work using smartphones and wearable sensors. In Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), Sydney, Australia, 14–18 March 2016; pp. 1–6.
36. Li, H. A novel design for a comprehensive smart automation system for the office environment. In Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA), Barcelona, Spain, 16–19 September 2014; pp. 1–4.
37. Jalal, A.; Kamal, S.; Kim, D. A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments. *Sensors* **2014**, *14*, 11735–11759. [CrossRef] [PubMed]
38. Kumar, V.; Fensel, A.; Fröhlich, P. Context Based Adaptation of Semantic Rules in Smart Buildings. In Proceedings of the International Conference on Information Integration and Web-based Applications & Services, Vienna, Austria, 2–4 December 2013; ACM: New York, NY, USA, 2013; pp. 719–728.
39. Alirezaie, M.; Renoux, J.; Köckemann, U.; Kristoffersson, A.; Karlsson, L.; Blomqvist, E.; Tsiftes, N.; Voigt, T.; Loutfi, A. An Ontology-based Context-aware System for Smart Homes: E-care@home. *Sensors* **2017**, *17*, 1586. [CrossRef] [PubMed]
40. Coronato, A.; Pietro, G.D.; Esposito, M. A Semantic Context Service for Smart Offices. In Proceedings of the 2006 International Conference on Hybrid Information Technology, Cheju Island, Korea, 9–11 November 2006; Volume 2, pp. 391–399.
41. Picard, R.W.; Healey, J. Affective wearables. *Pers. Technol.* **1997**, *1*, 231–240. [CrossRef]
42. Gyrard, A. A machine-to-machine architecture to merge semantic sensor measurements. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013.

43. Araque, O.; Corcuera-Platas, I.; Sánchez-Rada, J.F.; Iglesias, C.A. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Syst. Appl.* **2017**, *77*, 236–246. [CrossRef]
44. Pantic, M.; Bartlett, M. Machine Analysis of Facial Expressions. In *Face Recognition*; I-Tech Education and Publishing: Vienna, Austria, 2007; pp. 377–416.
45. Sebe, N.; Cohen, I.; Gevers, T.; Huang, T.S. Multimodal approaches for emotion recognition: A survey. In Proceedings of the Electronic Imaging 2005, San Jose, CA, USA, 17 January 2005; Volume 5670, p. 5670.
46. Mehta, D.; Siddiqui, M.F.H.; Javaid, A.Y. Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality. *Sensors* **2018**, *18*, 416. [CrossRef] [PubMed]
47. Anagnostopoulos, C.N.; Iliou, T.; Giannoukos, I. Features and Classifiers for Emotion Recognition from Speech: A Survey from 2000 to 2011. *Artif. Intell. Rev.* **2015**, *43*, 155–177. [CrossRef]
48. Ayadi, M.E.; Kamel, M.S.; Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognit.* **2011**, *44*, 572–587. [CrossRef]
49. Vinola, C.; Vimaladevi, K. A Survey on Human Emotion Recognition Approaches, Databases and Applications. *ELCVIA Electron. Lett. Comput. Vis. Image Anal.* **2015**, *14*, 24–44. [CrossRef]
50. Brouwer, A.M.; van Wouwe, N.; Mühl, C.; van Erp, J.; Toet, A. Perceiving blocks of emotional pictures and sounds: Effects on physiological variables. *Front. Hum. Neurosci.* **2013**, *7*, 1–10. [CrossRef] [PubMed]
51. Campos, J.J.; Frankel, C.B.; Camras, L. On the Nature of Emotion Regulation. *Child Dev.* **2004**, *75*, 377–394. [CrossRef] [PubMed]
52. Sokolova, M.V.; Fernández-Caballero, A.; Ros, L.; Latorre, J.M.; Serrano, J.P. Evaluation of Color Preference for Emotion Regulation. In Proceedings of the Artificial Computation in Biology and Medicine: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2015, Elche, Spain, 1–5 June 2015; Springer International Publishing: Cham, Switzerland, 2015; pp. 479–487.
53. Philippot, P.; Chapelle, G.; Blairy, S. Respiratory feedback in the generation of emotion. *Cogn. Emot.* **2002**, *16*, 605–627. [CrossRef]
54. Xin, J.H.; Cheng, K.M.; Taylor, G.; Sato, T.; Hansuebsai, A. Cross-regional comparison of colour emotions Part I: Quantitative analysis. *Color Res. Appl.* **2004**, *29*, 451–457. [CrossRef]
55. Xin, J.H.; Cheng, K.M.; Taylor, G.; Sato, T.; Hansuebsai, A. Cross-regional comparison of colour emotions Part II: Qualitative analysis. *Color Res. Appl.* **2004**, *29*, 458–466. [CrossRef]
56. Ortiz-García-Cervigón, V.; Sokolova, M.V.; García-Muñoz, R.M.; Fernández-Caballero, A. LED Strips for Color- and Illumination-Based Emotion Regulation at Home. In Proceedings of the 7th International Work-Conference, IWAAL 2015, ICT-Based Solutions in Real Life Situations, Puerto Varas, Chile, 1–4 December 2015; pp. 277–287.
57. Lingham, J.; Theorell, T. Self-selected "favourite" stimulative and sedative music listening—How does familiar and preferred music listening affect the body? *Nord. J. Music Ther.* **2009**, *18*, 150–166. [CrossRef]
58. Pannese, A. A gray matter of taste: Sound perception, music cognition, and Baumgarten's aesthetics. *Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci.* **2012**, *43*, 594–601. [CrossRef] [PubMed]
59. Van der Zwaag, M.D.; Dijksterhuis, C.; de Waard, D.; Mulder, B.L.; Westerink, J.H.; Brookhuis, K.A. The influence of music on mood and performance while driving. *Ergonomics* **2012**, *55*, 12–22. [CrossRef] [PubMed]
60. Uhlig, S.; Jaschke, A.; Scherder, E. Effects of Music on Emotion Regulation: A Systematic Literature Review. In Proceedings of the 3rd International Conference on Music and Emotion (ICME3), Yväskylä, Finland, 11–15 June 2013; pp. 11–15.
61. Freggens, M.J. The Effect of Music Type on Emotion Regulation: An Emotional Stroop Experiment. Ph.D. Thesis, Georgia State University, Atlanta, GA, USA, 2015.
62. Gangemi, A. Ontology design patterns for semantic web content. In Proceedings of the 4th International Semantic Web Conference, Galway, Ireland, 6–10 November 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 262–276.
63. Prud, E.; Seaborne, A. *SPARQL Query Language for RDF*; Technical Report; W3C: Cambridge, MA, USA, 2006.
64. Klyne, G.; Carroll, J.J. *Resource Description Framework (RDF): Concepts and Abstract Syntax*; Technical Report; W3C: Cambridge, MA, USA, 2006.
65. Sporny, M.; Longley, D.; Kellogg, G.; Lanthaler, M.; Lindström, N. *JSON-LD 1.0*; Technical Report; W3C: Cambridge, MA, USA, 2014.

66. Boley, H.; Paschke, A.; Shafiq, O. RuleML 1.0: The Overarching Specification of Web Rules. In Proceedings of the Semantic Web Rules: International Symposium, RuleML 2010, Washington, DC, USA, 21–23 October 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 162–178.
67. O'Connor, M.; Knublauch, H.; Tu, S.; Grosof, B.; Dean, M.; Grosso, W.; Musen, M. Supporting Rule System Interoperability on the Semantic Web with SWRL. In Proceedings of the Semantic Web–ISWC 2005: 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, 6–10 November 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 974–986.
68. Kifer, M. Rule Interchange Format: The Framework. In Proceedings of the Web Reasoning and Rule Systems: Second International Conference, RR 2008, Karlsruhe, Germany, 31 October–1 November 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–11.
69. Knublauch, H.; Hendler, J.A.; Idehen, K. *SPIN Overview and Motivation*; Technical Report; W3C: Cambridge, MA, USA, 2011.
70. Berners-Lee, T. *Notation3 Logic*; Technical Report; W3C: Cambridge, MA, USA, 2011.
71. Coronado, M.; Iglesias, C.A.; Serrano, E. Task Automation Services Study. 2015. Available online: http://www.gsi.dit.upm.es/ontologies/ewe/study/full-results.html (accessed on 18 April 2018).
72. Schröder, M.; Baggia, P.; Burkhardt, F.; Pelachaud, C.; Peter, C.; Zovato, E. EmotionML—An upcoming standard for representing emotions and related states. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction, Memphis, TN, USA, 9–12 October 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 316–325.
73. Sánchez-Rada, J.F.; Iglesias, C.A. Onyx: A Linked Data Approach to Emotion Representation. *Inf. Process. Manag.* **2016**, *52*, 99–114. [CrossRef]
74. Grassi, M. Developing HEO Human Emotions Ontology. In *Biometric ID Management and Multimodal Communication, Proceedings of the Joint COST 2101 and 2102 International Conference, BioID_MultiComm 2009, Madrid, Spain, 16–18 September 2009*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 244–251.
75. Lebo, T.; Sahoo, S.; McGuinness, D.; Belhajjame, K.; Cheney, J.; Corsar, D.; Garijo, D.; Soiland-Reyes, S.; Zednik, S.; Zhao, J. Prov-o: The prov ontology. In *W3C Recommendation, 30th April*; 00000 bibtex: Lebo2013; W3C: Cambridge, MA, USA, 2013.
76. Schröder, M.; Pelachaud, C.; Ashimura, K.; Baggia, P.; Burkhardt, F.; Oltramari, A.; Peter, C.; Zovato, E. Vocabularies for emotionml. In *W3C Working Group Note, World Wide Web Consortium*; W3C: Cambridge, MA, USA, 2011.
77. Sánchez-Rada, J.F.; Iglesias, C.A.; Gil, R. A linked data model for multimodal sentiment and emotion analysis. In Proceedings of the 4th Workshop on Linked Data in Linguistics: Resources and Applications, Beijing, China, 31 July 2015; pp. 11–19.
78. Sánchez-Rada, J.F.; Iglesias, C.A.; Sagha, H.; Schuller, B.; Wood, I.; Buitelaar, P. Multimodal multimodel emotion analysis as linked data. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), San Antonio, TX, USA, 23–26 October 2017; pp. 111–116.
79. Sánchez-Rada, J.F.; Schuller, B.; Patti, V.; Buitelaar, P.; Vulcu, G.; Bulkhardt, F.; Clavel, C.; Petychakis, M.; Iglesias, C.A. Towards a Common Linked Data Model for Sentiment and Emotion Analysis. In Proceedings of the LREC 2016 Workshop Emotion and Sentiment Analysis (ESA 2016), Portorož, Slovenia, 23 May 2016; Sánchez-Rada, J.F., Schuller, B., Eds.; 2016; pp. 48–54.
80. Khoozani, E.N.; Hadzic, M. Designing the human stress ontology: A formal framework to capture and represent knowledge about human stress. *Aust. Psychol.* **2010**, *45*, 258–273. [CrossRef]
81. Coronado, M.; Iglesias, C.A. Task Automation Services: Automation for the masses. *IEEE Internet Comput.* **2015**, *20*, 52–58. [CrossRef]
82. Verborgh, R.; Roo, J.D. Drawing Conclusions from Linked Data on the Web: The EYE Reasoner. *IEEE Softw.* **2015**, *32*, 23–27. [CrossRef]
83. Sánchez-Rada, J.F.; Iglesias, C.A.; Coronado, M. A modular architecture for intelligent agents in the evented web. *Web Intell.* **2017**, *15*, 19–33. [CrossRef]
84. Coronado Barrios, M. A Personal Agent Architecture for Task Automation in the Web of Data. Bringing Intelligence to Everyday Tasks. Ph.D. Thesis, Technical University of Madrid, Madrid, Spain, 2016.
85. Williams, K. The Technology Ecosystem: Fueling Google's Chromecast [WIE from Around the World]. *IEEE Women Eng. Mag.* **2014**, *8*, 30–32. [CrossRef]

86. Cortina, J.M. What is coefficient alpha? An examination of theory and applications. *J. Appl. Psychol.* **1993**, *78*, 98. [CrossRef]
87. Gaeta, M.; Loia, V.; Orciuoli, F.; Ritrovato, P. S-WOLF: Semantic workplace learning framework. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *45*, 56–72. [CrossRef]

CHAPTER 4

General Discussion, Conclusions and Future Research

This chapter aims to provide a comprehensive overview of the work presented in the preceding chapters. It begins by providing a general overview of the solutions proposed in the thesis and an analysis of the results obtained. It then presents a discussion of the main findings of the research describing their implications and significance. Finally, the chapter concludes with a description of possible lines for future research that can build upon the work presented in this thesis. Overall, the purpose of this chapter is to reflect upon the research conducted and to provide a summary of the contributions made to the field.

4.1 Overview

The first chapter outlined the main areas this thesis has contributed to, as well as its objectives. The main goal of this thesis is to advance the research on technology-assisted well-being promotion at work. The main issues involved in this undertaking were identified and gave rise to three contribution areas: automatic psychological awareness using non-obtrusive methods, analysis of well-being promotion strategies, and smart environment adaptation. As a response to these contribution areas, the following objectives were defined: (i) specification of an approach for stress prediction using non-obtrusive methods; (ii) design of an agent-based simulation model for understanding stress-related factors and regulation policies in the workplace; and, (iii) development of a reference architecture for an emotion-aware automation platform based on semantic technologies. The contributions made during the course of the thesis are gathered under these objectives. Figure 4.1 illustrates how the developed work is distributed along with the main objectives, showing what we contributed to each one.

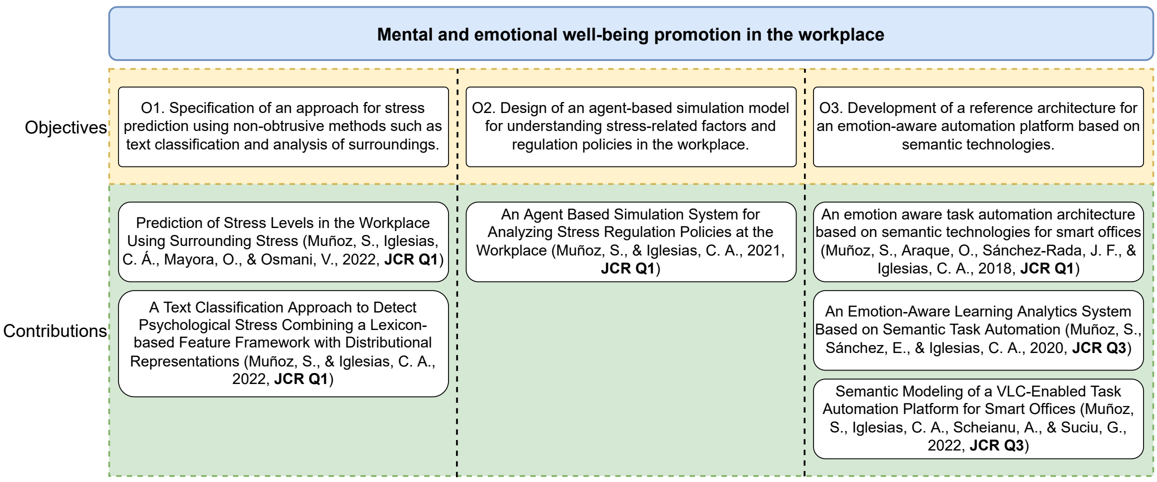


Figure 4.1: An overview of the contributions of this thesis, organized by objective.

We conclude the thesis by summarizing the contributions and considering their relevance in relation to the state of the art and contemporary works in the field.

One of the most relevant contributions of this thesis is the research on stress prediction using non-obtrusive methods. With this objective, two different approaches have been considered. One of the methods exploited is the use of textual data. This method benefits from the enormous quantity of information regarding people's emotions, feelings, and worries contained in social media services for recognizing stress. However, whereas textual data has been extensively studied for recognizing sentiments or emotions, its use for the recognition of stress was still in its early. The lack of well-validated solutions that demonstrate positive results in several corpora motivated the development of this work, described in **"A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations"** (Section 3.2.1). This work presents an approach to detect psychological stress in texts combining lexicon-based features with distributional representations. Compared with prior research, our work presents two noteworthy contributions: a lexicon-based feature framework that exploits affective, syntactic, social, and topic-related features; and a thorough evaluation of the proposed methods using several public English datasets. Several kinds of feature sets are explored through a lexicon-based feature framework, and three word embedding techniques are studied to exploit distributional representations. The proposed approach is evaluated on three English public datasets, providing a baseline for other researchers. Besides, the evaluation shows that the proposed approach improves the performance of previous work in all the datasets that were evaluated.

This thesis also contributes to the research on stress prediction using non-obtrusive methods with a novel approach to predicting stress from surrounding stress-related data. The proposed method uses the stress history of individuals and their close colleagues to predict future stress levels. Thus, this method can help to address the issue of data corruption or loss, which is common in real scenarios. The approach is presented in **"Prediction of stress levels in the workplace using surrounding stress"** (Section 3.2.2). This work also presents the validation of the approach, which is conducted through a public dataset and with a real-life study in which 30 workers were monitored over the course of 8 weeks. To the best of our knowledge, this is the first work on supervised stress prediction utilizing surrounding stress-related data gathered from individuals and their near colleagues. The findings of this study may have significant implications for the improvement of stress recognition systems, as it demonstrates that surrounding stress-related data can be used to predict stress. In this manner, a method exploiting surrounding stress data could be integrated with other techniques such as those utilizing physiological and behavioral data, thus decreasing the number of sensors and data required and increasing the effectiveness of stress detection methods in situations where data is limited.

Another of the contributions of this thesis falls within the research line of validating and evaluating well-being promotion strategies. Regarding this area, the design and development of well-validated models that correctly represent complex human characteristics is an open challenge for the research community. This thesis has contributed to this field by defining a model to calculate the stress and productivity of employees based on ambient and work conditions. The work is described in **"An agent based simulation system for analyzing**

stress regulation policies at the workplace" (Section 3.2.4). The stress model proposed is also implemented in an agent-based simulation system, empowering the analysis of how different stress regulation policies impact occupational stress and productivity. Four different stress regulation policies are analyzed thanks to this simulation system, providing insights into which strategies lead to the best results. Furthermore, to guarantee the reliability of the system and the results obtained, it was validated through a sensitivity analysis and through its application to a real scenario. In this manner, we contribute with a validated method for evaluating stress regulation policies in the workplace. Such method enables the simulation of a workplace to examine how different work and ambient conditions affect workers' stress and productivity over time.

To conclude, this thesis contributes with the development of an emotion-aware automation platform. This platform, based on event-driven semantic rules, enables the seamless adaptation of the environment to users' preferences, activities, and moods. Besides, the integration of semantic technologies helps to address interoperability and vendor lock-in issues, which are common in these kinds of systems. The architecture of this platform is presented in **"An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices"** (Section 3.2.6). This work also proposes a semantic model that covers the different components of the system: automation, emotions, and regulation techniques. The semantic model guarantees the smooth integration of multiple devices and services in the platform, as well as the portability of automations. Besides, the platform was validated through an experiment in a real-settings scenario. Thus, we contributed with a validated emotion-aware automation platform to regulate emotions free of interoperability or vendor lock-in issues. This line of research has undergone a series of improvements through the course of the thesis.

Firstly, it was applied to a different domain: e-learning. The rise of teleworking motivated the adaption of such a platform to online environments. Therefore, the automation platform was integrated with an e-learning service and provided with learning analytics capabilities. This work is described in **"An Emotion-Aware Learning Analytics System Based on Semantic Task Automation"** (Section 3.2.5). In addition to the online learning service, several sensors were incorporated into the platform to detect the user's level of attention and emotion while attending the lessons. Besides, the platform was further evaluated using partial least squares structural equation modeling (PLS-SEM) methodology. Therefore, through this work we contributed with the adaptation of the proposed platform to a different domain, with the integration of additional sensors, and with its thorough evaluation using PLS-SEM methodology.

Finally, we studied some communication improvements in the context of TETRAMAX VLP-Automation (Task Automation based on Visible Light Positioning and Blockchain) European project[1]. The communication between components in the platform was enhanced with the integration of Visible Light Communication (VLC). Also, this project was used as an opportunity to enhance the semantic model, making it compatible with relevant ontologies in the field of IoT: SSN and SOSA. This enables the representation of further information regarding devices, communication, etc. Thus, this thesis has contributed to this field by proposing a semantic model for a VLC-enabled automation platform. Such work is presented in **"Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices"** (Section 3.2.3). The work covers the design of an enhanced semantic model for the platform, the evaluation through a set of competency questions to verify its selection capabilities, and a prototype of architecture for a semantic VLC-enabled automation platform. The extension of the semantic vocabulary to make it compatible with SSN and SOSA and the exploration of additional communication technologies for the platform are the main contributions of this work.

4.2 Scientific results

The purpose of this section is to discuss the scientific results of this thesis, along with their relation to the primary objectives described in Section 1. Each objective is accompanied by a list of publications that have contributed to the achievement of that objective. The referenced pages provide complete details regarding each publication, including the abstract and full text.

[1] `https://short.upm.es/rmdft`

4.2.1 Objective 1: Specification of an approach for stress prediction using non-obtrusive methods

Table 4.1: Publications related to Objective 1

Page	Title	Impact	Contribution
35	A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations	JCR 2021 Q1 (7.466)	Definition of an approach to detect stress from text is proposed combining lexicon-based features with distributional representations. Proposal of a lexicon-based feature framework to explore several kinds of feature sets. Exploitation of distributional representations through three word embedding techniques. Establishment of a baseline for other researchers through the evaluation of the proposed approach on three public English datasets. (**Objective-1**)
55	Prediction of stress levels in the workplace using surrounding stress	JCR 2021 Q1 (7.466)	Definition of an approach for stress prediction from the stress history of individuals and their close colleagues. Validation of the approach on a public dataset and with data collected from a real-life study. (**Objective-1**)

Research on automatic stress recognition systems has progressed significantly in recent years in an effort to facilitate its early detection and prevent its associated adverse health and economic effects. A wide range of methods based on the analysis of physiological and behavioural data emerged, and research in the field has proven the reliability of these methods for the automatic detection of stress. However, the implementation of these techniques in real-life settings often presents further challenges that diminish their feasibility in real-world settings. The first objective of this thesis is motivated by these challenges. Intending to facilitate the detection of stress in real environments, we have worked on a solution to detect stress in a cost-effective and non-intrusive way. This solution has been approached from two different directions: the analysis of textual data and the exploitation of personal and social stress-related information.

Regarding the use of textual data, we aimed to provide a reliable method for detecting stress from text. The scientific community has effectively exploited textual data obtained from social media services in a broad range of mental health-related applications, including recognition of depression (William and Suhartono, 2021) or bipolar disorder (Sekulić, Gjurković, and Šnajder, 2018). However, existing solutions for the detection of stress have been limited to particular data sources, and there has been a lack of thoroughly validated methods that provide consistent results in several datasets. This made it hard to evaluate the robustness and effectiveness of this method to detect stress. Hence, we aimed to advance the state of the art by proposing an approach to stress detection in textual data and assessing its effectiveness using diverse public English datasets. The method put forward combines lexicon-based features with distributional representations in an effort to improve classification performance.

First, a lexicon-based feature framework is proposed to help organize and characterize features for stress detection in text. Such a framework also provides insight into which kinds of features and lexicons are more effective. The features included in the framework are derived from existing lexicons and analysis techniques that enable the extraction of information from textual data. A total of eleven distinct lexicons have been utilized to extract these features. Besides, some stylistic features intended to gauge the comprehensibility and readability of a text have been considered. This leads to a total of over 500 features of various types and natures.

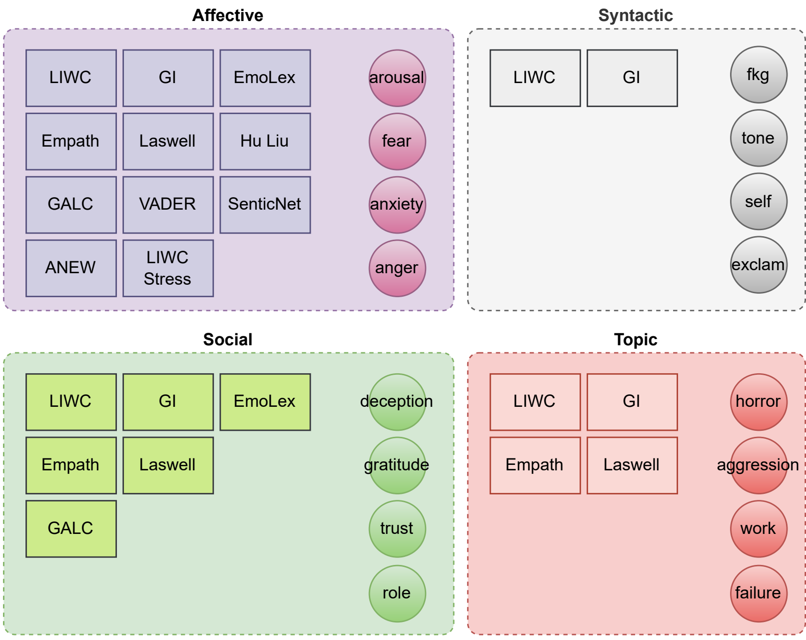
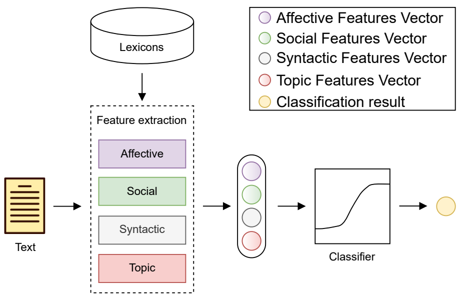


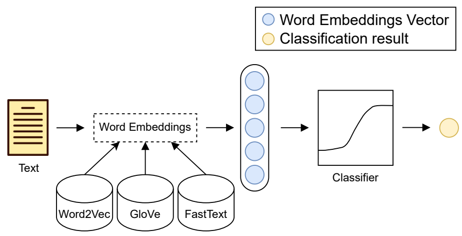
Figure 4.2: Overview of the proposed feature framework. The blocks represent the lexicons contributing to each set, while the circles represent example features of each set.

We divided the features into four categories in order to assess their influence depending on their nature. This led to four sets: affective, social, syntactic, and topic-related features. The set of affective features encompass those related to emotions (e.g., fear, sadness, joy), sentiment (positive, negative, or neutral), and moods (e.g., anxiety, attention, nervousness). This set is represented in all used lexicons, resulting in a total of 173 features. This set includes, for example, the following features: *fear*, *anxiety*, or *arousal*. Features related to social relations fall under the social set, which includes categories such as *trust*, *deception*, or *gratitude*. This set consists of 65 features derived from the lexicons Empath, Linguistic Inquiry and Word Count (LIWC), Geneva Affect Label Coder (GALC), General Inquirer (GI), Laswell, and EmoLex. The syntactic set consists of 75 features extracted from LIWC and GI, in addition to two stylistic features: Automated Readability Index (ARI) and Flesch-Kincaid Grade Level (FKG). Features in this set are related to the style and language used in the text. *Tone*, *self* and *fkg* are instances of features that belong to this set. Lastly, the topic set contains 272 features that refer to different topics existing in the text. Empath, LIWC, Laswell, and GI dictionaries include features of this kind, like *horror*, *aggression*, or *work*. Fig. 4.2 illustrates these feature sets, indicating the lexicon contributing to each set along with a few examples of features.

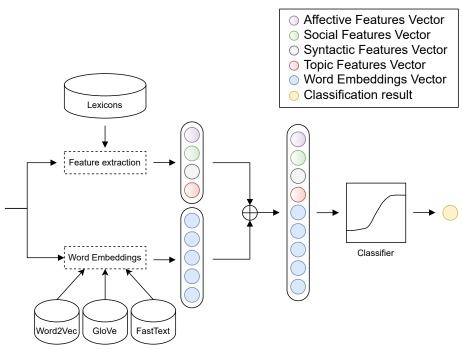
A preliminary analysis of this framework suggested that affective features are those which provide more information regarding the detection of stress. Besides, this framework was used as the basis for a machine learning model to develop a stress classifier. Then, two additional stress classification models were proposed to identify which method performs better in the detection of stress from text: a distributional representation model and an ensemble model that combines lexicon-based features with distributional representations.



(a) Lexicon-based features model



(b) Distributional representations model



(c) Ensemble model

Figure 4.3: General architecture representation of the proposed models.

The distributional representation model is founded on the use of word embedding techniques. Word embeddings are a method of distributed text representation that assigns similar representations to words that share similar meanings (Mikolov et al., 2013). This enables the transformation of text into vector representations that encode semantic and syntactic information. Some popular word embedding techniques are Word2Vec (Mikolov et al., 2013), Global Vectors for Word Representation (GloVe) (Pennington, Socher, and Manning, 2014) and FastText (Joulin et al., 2016). Each method presents a distinct way of learning the embedding matrix. Whereas how each technique affects the classification performance has been studied in many sentiment analysis-related problems, it has not been thoroughly analyzed for the detection of stress. Consequently, to evaluate the impact of each technique on the stress detection performance, the three outlined techniques were applied to the extraction of word vectors and evaluated separately.

Finally, to enhance the classification performance, an ensemble model that takes into account both distributional representations and lexicon-based features was proposed. This model combines the word embedding vector representation of each text instance with the lexicon-based feature vector extracted from it. Therefore, the information from word embeddings is combined with the affective and lexical information obtained from lexicon-based features. This combination of information may enhance the performance of a classifier algorithm learning from this combined set compared to one that only learns separately from word embedding or lexicon-based features data. The general architecture representation of the proposed models is shown in Figure 4.3.

These models were validated through an experimental study across three English public datasets from different sources (Reddit, Twitter, and personal interviews). In this manner, we contributed with the validation of the proposed method in different datasets, establishing a baseline for other researchers. Furthermore, the experimental study allowed us to assess the effectiveness of distinct lexicons or word embedding methods and to gain insight into which models and techniques are more effective for stress detection across different corpora.

According to the results, the combination of lexicon-based features with distributional representations is the best-performing approach. The Friedman test confirmed these results and demonstrated their statistical relevance. Regarding the different kinds of features, the obtained results confirmed that affective features yield better results in text classification of stress. However, incorporating syntactic and topic-related features can further improve classification performance. Out of the various lexicons used, LIWCand GI were found to be the most effective in terms of performance. These lexicons incorporate features of all kinds and are the only ones that furnish syntactic features. Finally, with regard to how each word embedding technique performs, the statistical results indicate that FastText embeddings are the most effective distributional representation technique for this task.

The experiments also showed that the proposed model demonstrates superior performance in comparison to earlier works on stress detection from text. Our approach makes two notable contributions when compared to prior works: the evaluation and validation across data from multiple corpora and the comprehensive study of different feature extraction techniques. Compared with other stress detection solutions (Panicker and Gayathri, 2019), stress detection from text attains comparable performances while involving less complex or expensive implementations. In addition, the proposed techniques can be incorporated into existing software solutions, enabling real-time stress detection and decreasing the costs of complex hardware implementations. For instance, these methods could be integrated into existing software solutions like a company communication platform or social messaging apps. This would allow managers to conveniently monitor their employees' stress levels and users to have a better understanding of their personal state.

The complete work is presented in the journal publication **"A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations"** (Section 3.2.1).

The other direction followed to tackle the challenge of stress detection in real environments is based on the exploitation of personal and social stress-related information. In this regard, we aimed at taking advantage of the contagious nature of stress and its proneness to be affected by surroundings (Dimitroff et al., 2017) for enabling the prediction of stress values in the presence of scarce data. We sought to investigate how utilizing stress-related data from past measurements, and near colleagues can aid in enhancing stress recognition performance in a non-intrusive and cost-effective manner. This could reduce the number of sensors and data required, as it enables the use of the data obtained from specific individuals to the prediction of the stress levels of other individuals. Additionally, it can also be employed to predict an individual's stress level at a specific time when data corruption or loss has occurred. Therefore, it can help to facilitate the task of automatic stress detection.

With this objective, we proposed a machine learning method that uses surrounding stress information to predict the current stress levels of an individual. We considered the surrounding stress data to be a compound of two factors: the personal component and the social component. The personal component refers to the individual's previous stress levels, and the social component refers to the stress levels of the individual's nearest colleagues. We define nearest colleagues as those who are physically near the individual, as determined by smartphone sensor data, including WiFi, cell, and GPS location. The self-reported stress level (low, moderate, high) is the target class to predict, while the surrounding stress data is used as input attributes.

Figure 4.4 depicts an overview of the proposed model. As depicted, the surrounding stress information goes through two distinct processing modules. These modules receive the previous stress levels from the individual and their nearest colleagues as input, generate a feature vector representing this information, and output this vector. The feature vectors are then concatenated and fed to a machine learning classifier, which generates a prediction founded on the provided information.

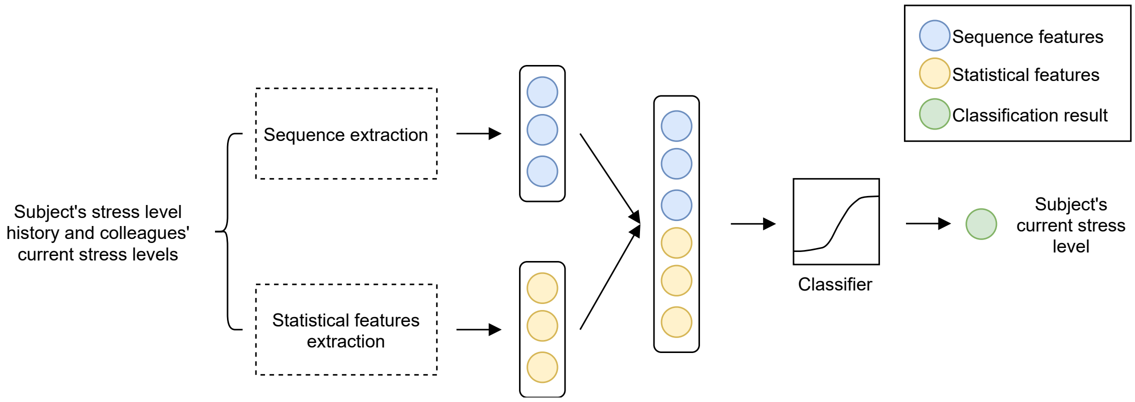


Figure 4.4: General architecture representation of the method to predict stress based on surroundings

With the aim of enhancing stress prediction performance, we explored different ways to combine the information. Consequently, three distinct models were defined: (i) using only sequential data (M_{SEQ}); (ii) using only statistical features (M_{SF}); and (iii) using a feature ensemble of sequential and statistical data (M_{FE}). The first model is focused on predicting stress using only the sequential data of the surrounding stress. This sequential data is comprised of the individual's stress levels ordered chronologically for a specific time frame. The second model exploits various statistical features extracted from the surrounding stress data to make predictions. The extracted features are shown in Table 4.2. Lastly, the third model merges both sequential and statistical features into one feature set, thereby leveraging the complementary information they provide to improve prediction performance.

In addition, to investigate the impact of the personal and social stress components on each model, three distinct methods were proposed for calculating the surrounding stress data. Each of these methods computes the surrounding data using different components: (i) only personal data; (ii) only social data; and (iii) both personal and social data. In this manner, our work contributed with a novel approach for stress prediction and provided a study on the techniques that led to the best performance.

Table 4.2: Statistical features extracted from the data.

Feature	Formula	Description
avg	$E_{avg} = \frac{1}{w} \sum_{i=t}^{t-w} e_i$	Average of all levels
std	$E_{std} = \sqrt{\frac{1}{w-1} \sum_{i=t}^{t-w} (e_i - \overline{e})^2}$	Standard deviation of all levels
max	$E_{max} = \max E(T)$	Maximum value between levels
min	$E_{min} = \min E(T)$	Minimum value between levels
last	$E_{last} = e_t$	Last value of all levels
Δ	$E_{\Delta} = e_t - e_{t-w}$	Total level increment
δ	$E_{\delta} = e_t - e_{t-1}$	Last level increment

An experimental study was conducted to comprehensively evaluate the performance of each model in multiclass stress prediction and determine the optimal resources and methods. The main objective of the experiment was to gain insight into whether surrounding stress information can support the prediction of stress and which model is better suited to this task. Three different scenarios were designed to isolate the effects of the personal stress component, the social stress component, and the combination of both. This allowed for a detailed analysis of the contribution of each component to the overall performance of the models.

To this aim, we used a previously gathered dataset comprising psychological self-assessment data obtained from standardized, validated questionnaires and unobtrusive sensor information collected from smartphones. These data were collected during an eight-week-long real-world study involving 30 participants. Furthermore, to establish the generalizability of our proposed method, we performed an external validation using an independent dataset known as StudentLife (R. Wang et al., 2014). This dataset contains real-world behavioral data collected from students over the course of an entire university course. This served to provide insight into the generalizability of our proposed method across different contexts and populations.

The experiments showed that our approach yielded promising performance. Hence, it is reasonable to assume that surrounding stress-related information can be relevant and support the prediction of stress. Also, the statistical analysis revealed that incorporating both personal and social stress information into the prediction model yielded the highest performance in terms of stress prediction accuracy. This finding highlights the importance of considering multiple sources of information in stress detection, as it leads to a significant enhancement in comparison to other methods. If we compare it with the use of only one kind of data, the analysis of the results indicates that utilizing personal stress data alone outperforms utilizing social stress data alone, thus indicating the importance of individual-specific information in the prediction of stress levels.

Finally, regarding how different feature extraction methods can improve stress prediction performance, the analysis of the results revealed that combining sequential and statistical features is the most effective method for improving stress prediction performance. This approach, which combines both types of features, was shown to yield the highest performance results across all experimental scenarios.

On the basis of these results, we concluded that the use of surrounding stress-related information could yield relevant performance in stress prediction. Our findings are consistent with current research on the phenomenon of affective state contagion in interpersonal social interactions, which highlights the significant impact that close individuals can have on an individual's affective state. (Dimitroff et al., 2017; Engert, Linz, and Grant, 2019; Jia and Cheng, 2021; Petitta et al., 2021). Additionally, our results align with existing research on the impact of historical affective information on future emotional states (Hollis et al., 2017; Goodday and Friend, 2019). The results of our study have significant implications for the development of stress prediction methods. The approach presented here could be integrated with other techniques, like those utilizing physiological and behavioral data, to enhance the accuracy of stress level predictions. Incorporating personal and social stress-related data into the prediction model could improve the efficiency of the system by reducing the need for multiple sensors and a large amount of data and increase the effectiveness of stress detection in situations where available data is limited.

In conclusion,this thesis has contributed with a contagion model for predicting occupational stress that has been validated in a real scenario. The findings of this article point to the importance of taking a holistic approach to tackling occupational stress, rather than focusing solely on individual employees, when addressing the well-being of the workforce. The described work on predicting stress based on surrounding stress-related data is presented in the paper **"Prediction of stress levels in the workplace using surrounding stress"** (Section 3.2.2). To the best of our knowledge, this is the first research into the use of surrounding stress-related information to forecast future stress levels. Our study of the relationship between surrounding stress and workers' stress has the potential to open new avenues of research to enhance future monitoring systems, providing a deeper understanding of occupational stress, its effect on worker health, and the management of human resources.

4.2.2 Objective 2: Design of an agent-based simulation model for understanding stress-related factors and regulation policies in the workplace

Table 4.3: Publications related to Objective 2

Page	Title	Impact	Contribution
89	An agent based simulation system for analyzing stress regulation policies at the workplace	JCR 2021 Q1 (3.817)	Development of a stress model that incorporates both work and ambient factors to determine the stress levels and productivity of employees. Development of an agent-based simulation system where to implement this model, enabling the examination of occupational stress and productivity under various stress regulation strategies. Analysis of distinct stress regulation policies using the developed system. Validation of the model with a sensitivity analysis and with a real experiment in a real-settings scenario. (**Objective-2**)

The vast impact that occupational stress has on society has boosted the interest and research on prevention strategies. One of the critical issues in this regard is identifying the foremost provoking factors (or stressors). Identifying stressors helps to prevent stress and enables the design and implementation of stress regulation policies that decrease its harmful effects. Nevertheless, the reconditeness of human behavior coupled with the economic costs associated with implementing and validating interventions are known obstacles to the effective deployment of these policies. Therefore, there is a need for solutions that enable the validation of the suitability and efficacy of promotion strategies before their implementation. This challenge shapes the second objective of this thesis. Our goal is to facilitate the design and evaluation of well-being promotion and stress regulation strategies. With this aim, we propose a solution based on agent-based simulation and modeling techniques.

Agent-based simulation techniques provide an alternative method to assess a wide range of behavioral dimensions or parameters without incurring the cost of implementing complex systems. Their distinctive characteristics have given them a well-established role in the examination of human behavior, including in the field of work-related stress (Page and Ashlock, 2015; Duggirala et al., 2016). Despite the advancements made, the design of reliable agent behavior models is still an open challenge because of the vast amount of different variables associated with human behavior (Elkosantini, 2015b). These variables are usually tough to quantify, calibrate, and even justify. Consequently, existing solutions often focused on specific factors, and there was a lack of validated broad models that considered multiple aspects of the work environment. This work aims to fill this gap with the development and evaluation of an agent-based simulation system that helps to evaluate occupational stress regulation strategies. The system integrates a stress model based on existing literature that considers work and ambient conditions.

Our first contribution in this field is the definition of a workplace stress model. The defined model is based on the Silverman's stress model (Silverman, 2001). This model proposes that stress is comprised of three fundamental elements: *Event Stress (ES)*, *Time Pressure (TP)*, and *Effective Fatigue (EF)*. In the occupational context, *Event Stress* can be attributed to a significant increase in work volume or demands, *Time Pressure* to the relation between the time required to complete pending tasks in an individual's task queue and the remaining work time available, and *Effective Fatigue* to the accumulated exhaustion stemming from performing work-related tasks and the effect of external factors. In this way, the stress level is given by Equation 4.1:

$$S = \frac{ES + TP + EF}{3} \tag{4.1}$$

where stress (S) and all the elements that compound it range between 0 and 1.0. A stress level of 0 indicates that the agent is over-confident and free of stress, 0.5 represents a neutral state, and 1.0 denotes the state of maximum stress.

The original model and existing works based on it (Duggirala et al., 2016; Hayatnagarkar et al., 2016) incorporate specific workload-related factors to calculate each stress component, such as the number of tasks and the available time to complete them. According to previous works, *Event Stress* is determined as the ratio between the number of tasks received in a day and the average number of tasks an individual is accustomed to handling. Thus, for an individual who typically handles 20 tasks per day, if they receive 30 tasks on one particular day, the contribution of the *Event Stress* component will be 0.75. Similarly, if that same individual receives ten tasks on another day, the contribution will be 0.25. Likewise, *Time Pressure* is computed based on the relationship between the time needed to complete a task accurately and the actual time available for the task. Hence, for an individual whose pending tasks ideally should take 30 minutes to complete but only has 20 minutes available, the stress impact of TP will be 0.6.

The third stress component, *Effective Fatigue*, is composed of contributions from multiple parameters known as Performance Moderator Functions (PMF), which can have a positive or negative contribution. Negative stressors are those that raise effective fatigue, whereas positive stressors are those which decrease it. EF is computed as the normalized sum of all

PMFs as depicted in Equation 4.2:

$$EF = \frac{1}{N}\sum_{i=1}^{N}(\overbrace{\frac{WPMF_i}{WPMF_i + FT_i}}^{\text{negative stressors}} - \underbrace{\frac{WPMF_i}{10}}_{\text{positive stressors}}) \tag{4.2}$$

where $WPFM_i$ represents the weight of the PMF, FT_i represents the fatigue tolerance of an individual to the PMF ($0 \leq FT_i \leq 1$), and N is the number of PMFs.

Based on this model, our research aimed to explore the effects of different ambient and work factors to further develop it. The main novelty in our model is the integration of ambient conditions and additional work conditions, for example, by analyzing the impact of working overtime hours or the interruptions caused by electronic communication (i.e., email and messaging apps). The illustration presented in Fig. 4.5 provides a general overview of the model design, depicting the three stress components and the factors that contribute to them.

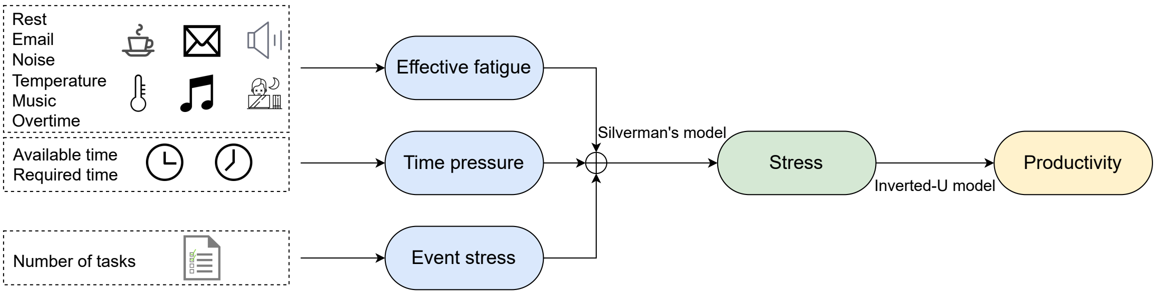


Figure 4.5: Stress model for agent-based simulation

As can be seen in the figure, we extended the model by including several factors that affect effective fatigue. With this aim, the literature regarding stress sources was exhaustively analyzed. Then, the contributions of each factor to occupational stress were estimated according to the findings of relevant works in the field to design the model as reliable as possible. Table 4.4 shows the factors that have been included along with their contribution to effective fatigue.

Besides, we also integrated the Yerkes-Dodson Inverted-U model of stimuli and performance (Yerkes and Dodson, 1908) to calculate productivity from the stress level. This model describes the relationship between stress and performance and establishes the need to have an optimal arousal level to achieve maximum performance. If the level of arousal exceeds or falls short of the optimal level, performance will be negatively impacted.

In addition to the definition of the model, we intended to contribute with the development

Feature	Contribution	Description	Source
Temperature and humidity	If WBGT $\leq$ 20 ºC or WBGT $\geq$ 25 ºC: $WPMF_t = 0.04 \cdot \lvert WBGT - 22 \rvert$	An increase in the difference between the current wet-bulb globe temperature (WBGT) and the ideal WBGT (22 ºC) will increase the effective fatigue	E. Somanathan et al. (Somanathan et al., 2021)
Overtime hours	$WPMF_o = 0.021 \cdot O_h$	Working overtime hours (O_h) will increase the effective fatigue	M. Singh et al. (Singh et al., 2016)
Rest time	$WPMF_r = 0.016 \cdot R_t$	Taking a break (R_t) will decrease effective fatigue	A. Dababneh et al. (Dababneh, Swanson, and Shell, 2001)
Noise	If $N_L \geq 65$ dB: $WPMF_n = 0.03 \cdot (N_L - 65)$	If the noise level (N_L) increases over a certain limit (65 dB), the effective fatigue will also increase	L. E. Maxwell (Maxwell, 2016)
Electronic communication	$WPMF_e = 0.0029 \cdot E_r$	The effective fatigue will increase with the number of emails or messages received (Er)	K. Kushlev et al. (Kushlev and Dunn, 2015)

Table 4.4: Contribution of Performance Moderator Functions to effective fatigue

of a simulation system that facilitates the analysis of stress regulation strategies. This system, which integrated the previously defined stress model, was designed with the primary objective of providing a reliable representation of the work environment. In this way, the system allows the user to effortlessly adjust several different parameters of the simulation using the graphical user interface. In addition, the graphical interface shows real-time data during the simulation run. After completion, it stores all the data for later examination and representation.

To give insight into the effectiveness of stress regulation strategies, a set of policies were evaluated through the developed system. These policies were based on the adaption of the workload and ambient conditions and included different levels of automation to conduct these adaptations. For each automation scenario, stress and productivity-related data, such as stress level and the number of completed tasks, were measured and analyzed. The results of the simulation suggested that stress regulation policies based on the adaption of workload conditions were successful in regulating stress levels and enhancing performance. Besides, the study showed that regulation policies based on ambient adaption had a lower overall impact on stress levels compared to those based on workload adaption but helped in preventing stress peaks.

Lastly, we conducted a validation of the proposed system in order to verify its reliability. This validation was performed from two points of view: through a sensitivity analysis and through a real experiment. A sensitivity analysis is a popular technique for the validation of agent-based simulation systems. It can reveal which components of a model have a more significant influence on the results and how each component interacts with each other to influence model outcomes (Borgonovo et al., 2022). The results showed that workload conditions have a greater impact on stress and validated that the influence of the different input variables on the output is consistent with what was expected based on the design of the model.

In addition, we validated the conclusions drawn from the simulation with a real experiment with 20 real participants. Participants were supposed to complete a set of programming tasks with different duration and difficulties, while their stress level was measured through biometrical signals and a questionnaire. In order to analyze the effect of the adaptation policies, the participants were split into two groups: the control group and the treatment group. Participants assigned to the control group conducted the experiment without the implementation of any form of adaptation or automation. Conversely, those assigned to the treatment group conducted the experiment utilizing the adaptation system.

Data from questionnaires revealed that users feel more inclined to use the adaptation system rather than perform the task without adaptation. Furthermore, the results obtained from biometric data were consistent with this idea, indicating that the participants' stress level was lower with the automatic adaptation enabled. To conclude, quantitative data extracted from the experiment itself, such as the number of tasks completed, supported the idea that the automation of the environment helped to enhance productivity. These outcomes are consistent with those obtained from the simulation, thus confirming the reliability of the proposed system.

The complete work is presented in **"An agent based simulation system for analyzing stress regulation policies at the workplace"** (Section 3.2.4). The work contributes with a stress model to analyze the influence of different factors on stress. Besides, a simulation system integrating this model was developed and validated. The findings in this work confirm the potential of agent-based simulation technologies to contribute to the prevention of stress, thanks to the advantages they offer in the identification of stress causes and the design and evaluation of regulation policies.

4.2.3 Objective 3: Development of a reference architecture for an emotion-aware automation platform based on semantic technologies

Table 4.5: Publications related to Objective 3

Page	Title	Impact	Contribution
128	An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices	JCR 2021 Q1 (3.847)	Design of an emotion aware automation platform architecture for smart offices. Definition of a semantic model covering the different components of the system: automation, emotions, and regulation techniques. Implementation and evaluation of the proposed architecture in a real scenario. (**Objective-3**)
103	An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices	JCR 2021 Q3 (2.690)	Integration of the emotion aware automation platform with an e-learning service. Design and implementation of an emotion-aware learning analytics architecture. Validation of the platform using partial least squares structural equation modeling (PLS-SEM) methodology. (**Objective-3**)
69	Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices	JCR 2021 Q3 (2.690)	Improvement of the communication between components in the automation platform with the integration of Visible Light Communication (VLC). Design of an enhanced semantic model for the platform that includes relationships with known ontologies in the IoT field. Definition of a prototype of architecture for a semantic VLC-enabled automation platform. Evaluation of the model through a set of competency questions to check its selection capabilities. (**Objective-3**)

The evolution of ambient intelligence has opened many new opportunities to improve people's well-being. One of the opportunities these technologies offer is their application to working environments to increase the comfort and productivity of employees. To this end, the environment must be adaptive to workers' moods and needs. Recognition and intervention processes that automatically detect and regulate the individual's psychological state must be seamlessly and continuously integrated to guarantee the efficacy and efficiency of well-being promotion strategies. However, the wide variety of components and technologies involved in these systems often entails a significant challenge: the interoperability between these components. The third objective of this thesis is motivated by this challenge. Hence, this thesis aims to advance state of the art in smart environment adaption with a reference architecture for an emotion-aware automation platform based on semantic technologies.

Over the last few years, there has been a boost in the research on pervasive environments that automatically adapt to users' needs. Despite the recognised importance of emotion awareness and semantic modelling in the context of intelligent automation, our review of the current literature suggests that there is a lack of research that addresses the integration of these concepts within a unified platform. Furthermore, the absence of a common representation and modeling for data and components in these systems raises the challenge of integrating additional devices or services and tailoring them to further scenarios. In this thesis, we aimed to advance the state of the art with a solution that fills this gap.

With this aim, we defined a semantic model for an emotion-aware automation platform. The core of the model is based on the EWE ontology (Coronado, Carlos A Iglesias, and Serrano, 2015), a vocabulary specifically devised to model the most relevant elements of TASs in a descriptive way. Besides, to model the platform's emotion-related components, the defined model was integrated with Onyx ontology (Sánchez-Rada and Carlos A. Iglesias, 2016), a vocabulary specifically designed for the representation of emotions and emotion detection processes. The main classes of these ontologies are described in Section 1.2.3.

```
emo:VideoCamera a owl:Class ;
    rdfs:label "A video camera able to detect emotions." ;
    rdfs:subClassOf emo:EmotionSensor .

emo:EmotionDetected a owl:Class ;
    rdfs:label "Event representing the detection of an emotion";
    rdfs:subClassOf ewe:Event ;
    rdfs:domain emo:EmotionSensor .

emo:SmartLight a owl:Class ;
    rdfs:label "Smart Light" ;
    rdfs:subClassOf emo:EmotionRegulator .

emo:ChangeAmbientColor a owl:Class ;
    rdfs:label "Change ambient color";
    rdfs:subClassOf ewe:Action ;
    rdfs:domain emo:SmartLight .
```

Listing 4.1: Semantic representation of an emotion sensor and an emotion regulator channel written in Notation3

The class *ewe:Channel* was sub-classed in order to provide a class representing a device or service able to provide emotion-related capabilities. In this manner, a class *emo:Channel* was created in order to provide an emotional channel class that handles the generation and trigger of events and actions associated with emotions. The channels *emo:EmotionSensor* and *emo:EmotionRegulator* were derived from this class. The former generates events associated with the recognition of emotions, while the latter provides actions that are intended to regulate these emotions. These classes enable the representation of any device or service able to detect or regulate emotions. Events and actions provided by these emotional channels may include parameters. To enable the representation of emotion-related parameters, a new class was defined sub-classing *ewe:Parameter* and *onyx:Emotion*. In this manner, emotions can be modeled as parameters of events or actions while exploiting the valuable properties and relationships provided by the Onyx ontology.

The proposed model can be used for the modeling of all components in the platform. For example, let us consider a scenario consisting of a smart office with emotion detection and regulation capabilities. The environment includes video cameras to detect the emotions of users and smart lights whose color, temperature, and brightness can be adjusted to empower emotion regulation. These devices can be easily represented using the *emo:EmotionSensor* and *emo:EmotionRegulator* classes. Listing 4.1 shows the modeling of this example written in Notation3.

An instance of events and actions with specific parameters, based on the concepts defined in the previous listing, is depicted in Listing 4.2. This listing shows the modeling of the event of sadness detection and the action of changing ambient color to regulate the emotion.

```
:sad-emotion-detected a emo:EmotionDetected ;
    ewe:hasEmotion onyx:sadness .

:play-music a emo:PlayRelaxingMusic ;
    ewe:hasSong "the title of the song to be played".

:change-ambient-color-green a emo:ChangeAmbientColor ;
    ewe:hasColor dbpedia:Green .
```

Listing 4.2: Event and action instances

Likewise, automation rules can be represented by utilizing the punning mechanism to link classes to properties of Rule instances. An example of this is depicted in Listing 4.3, where we can see the definition of a rule instance representing a rule that is triggered by the event of *sad emotion detection* and fires the action of *changing ambient color to green* (both defined in Listing 4.2).

```
:regulate-stress a ewe:Rule ;
    dcterms:title "Stress regulation rule "^xsd:string ;
    ewe:triggeredByEvent    :sad-emotion-detected ;
    ewe:firesAction         :change-ambient-color-green .
```

Listing 4.3: Rule instance

Besides, this thesis also contributed with a reference architecture for an emotion-aware semantic automation platform. The proposed platform integrates the defined semantic model and is based on the reference architecture for TAS (Coronado and Carlos A Iglesias, 2015). However, it was expanded to enable emotional awareness. The proposed system consists of two primary modules: an emotional context recognizer and an emotion-aware task automation server. The former is responsible for the collection of users' emotions and context-related information and their delivery to the automation platform. This platform receives this information and evaluates it to trigger the corresponding actions. The received events are evaluated along with a set of rules that can be configured and managed through the platform. To perform the evaluation, the module also integrates the semantic rule engine EYE (Verborgh and De Roo, 2015), which handles the logical reasoning of the received events and configured rules and generates the corresponding actions that must be triggered. An overview of this architecture is depicted in Figure 4.6.

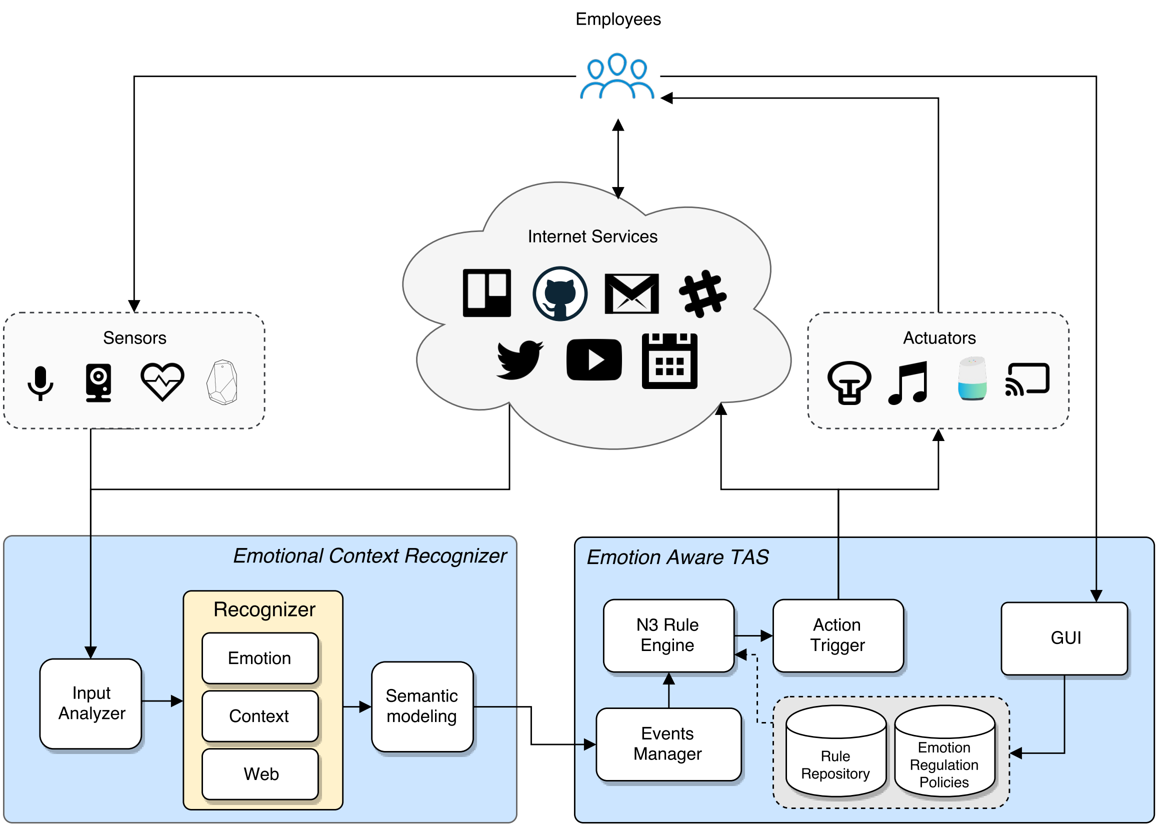


Figure 4.6: Emotion Aware Automation Platform Architecture

A set of devices and services has been integrated into the system, including several sensors and actuators that serve as an information source or as action executors to adapt the environment. Thereby, the platform empowers users to effortlessly define and configure their own automation rules to tailor their environment with the aim of improving their comfort and productivity. The effectiveness of the proposed platform was evaluated through an experiment with 28 participants in a real-world scenario.

The results obtained from the experiment demonstrated the effectiveness of the proposed system, concluding that users positively evaluate the use of the adaptation system and prefer it to perform the task without adaptation. Furthermore, users experience greater satisfaction with an adapted work environment, and as a result, they consider the implementation of an adaptation system for their workspace. Besides, the results suggest that the automation system can contribute to enhancing the performance of the user in a particular task.

This work is described in the journal article **"An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices"** (Section 3.2.6).

Given the positive results obtained with the implementation of the automation platform in a work-based scenario, we worked on adapting it to a different context: the e-learning context, since it allows us to test it seamlessly. E-learning has become an integral part of the academic landscape due to the wide range of opportunities it provides for the learning experience. Nonetheless, these platforms often lead to an exacerbation of difficulties in communication between teachers and students, resulting in a decrease in emphasis on the human aspect of learning. Also, whereas recognizing students' emotions seems essential to stimulate their learning process, there are few real implementations of these platforms that include emotion recognition. Therefore, we have worked to fill this gap with the integration of the semantic automation platform in an e-learning environment.

With this purpose, we proposed an emotion-aware e-learning platform architecture that automatically monitors students' emotions, engagement, and attention, aiming to enhance their academic performance. Such a platform gives educators the capability to consider this information and tailor their materials and methods as needed. Also, the platform integrates ambient intelligence technologies to adapt the environment to the students' moods, taking advantage of the growing prevalence of intelligent devices.

The existing e-learning literature was reviewed to identify the main requirements that such an architecture should address (Horton, 2011; Bentley, Johnston, and Baggo, 2002; Keller and Suzuki, 2004). Then, a reference architecture meeting these requirements was designed and developed.

The usefulness of the platform was validated using PLS-SEM methodology (Fornell, 1994). An experiment involving 30 participants was designed to verify that the use of the platform enhances user satisfaction and helps students to achieve better academic results. During the course, the e-learning system received information gathered from sensors that continuously monitored the emotional and attention states of participants. The emotional and attention states of participants triggered specific adaptation rules, including adjustments to light color or music. Furthermore, participants were able to visualize their emotions through the developed dashboard. Once the lesson concluded, a questionnaire was given to the participants with questions related to the participant's opinions about the platform.

The analysis of the results showed that emotional information contributes to improving user satisfaction and benefits. Additionally, we also noted that the system quality also leads to an enhancement of user satisfaction and benefits. The PLS-SEM analysis also validated the structural model, confirming the validity of the findings obtained from the experiment.

This work was published in the journal article **"An Emotion-Aware Learning Analytics System Based on Semantic Task Automation"** (Section 3.2.5). The findings of this work promote the utilization and enhancement of this kind of e-learning system. Cording to the results, they offer users a set of advantages, including the assessment of their affective state or concentration level and the adjustment of the environment to their feelings and needs. In this way, this work contributes to the research on mental and emotional well-being promotion in distance educational environments with an architecture for an emotion aware e-learning platform based on semantic technologies. Such architecture provides the students with attention and mood regulation capabilities and helps to decrease communication barriers between teachers and students in distance learning environments. In addition, the application of a semantic layer enables data interoperability and portability of components. This is a significant characteristic compared with previous works, as the integration of semantic technologies and emotion regulation techniques in a unified e-learning platform had not been thoroughly explored.

Finally, an enhancement was conducted regarding the communication technologies involved in the automation platform. In the context of the TETRAMAX VLP-Automation European project, the platform was extended to support Visible Light Communication (VLC). VLC enables the exploitation of the visible light bandwidth to transmit information using the lights present in an environment (e.g., Light Emitting Diode (LED) lights). For this work, we cooperated with partners who have extensive experience in developing VLC systems.

On the occasion of the extension of the platform, we also took the opportunity to extend the semantic model and make it compatible with one of the preeminent ontologies in the IoT field: SSN (Compton et al., 2012). Hence, a vocabulary with the relevant associations between ontologies was specified. Such vocabulary facilitates the semantic representation of all entities in the smart office environment. It is founded on two ontologies: SSN (and its core, SOSA) and EWE, and its selection capabilities were evaluated through a set of Competency Questions (CQs).

Compared with other automation or intelligent systems (Sahil Gupta and Swati Gupta, 2021; Bhatia and Sood, 2019; Ullo and Sinha, 2020), the proposed automation platform presents some advantages. Some of these advantages are related to the use of VLC as communication technology: alleviation of the load of the RF spectrum, prevention of the electromagnetic interference, and reduction in operating costs and carbon footprint. However, the main advantages derive from the use of a semantic model for representation and organization of all the components within the intelligent environment. This approach not only facilitates interoperability and scalability of the system but also enables easy access to linked data related to the components thanks to its straightforward and common representation. By leveraging this feature, users are able to access complementary information from external sources regarding the different elements of the platform. The integration of this information from external sources allows the users to expand the scope of their queries, thereby providing a more comprehensive understanding of the platform and its components. Overall, the use of a semantic model in this platform offers a number of benefits that contribute to its effectiveness and usability.

The proposed semantic model for a VLC-enabled task automation platform was published as a journal article, **"Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices"** (Section 3.2.3). Therefore, the main contributions of this work are the exploration of additional communication technologies for the platform, and the extension of the semantic vocabulary. This extension enables the representation of all the components in the platform using the most prominent IoT ontology (SSN), expanding the information provision and reasoning capabilities of the platform.

4.3 Applications

The findings of this thesis have advanced the understanding of well-being promotion in ways that go beyond the results presented in the preceding section. The list of contributions includes the translation of the scientific outcomes of this thesis into practical applications in industry. In this regard, results from this thesis have been used in research projects at European and national levels:

- The national project COGNOS (Agentes Cognitivos para la Interacción Personal Segura y Emotiva en Entornos de Movilidad) [2] (Jan 2019 - Dec 2022) aims to investigate

[2]COGNOS project: `https://short.upm.es/jh7mb`

the application of affective computing and agent technologies for achieving empathetic social interactions in a secure and trustworthy interaction environment. To this aim, the project proposes to advance research on different techniques to analyze the human-computer interactions in the system in several modalities, including the use of different sources such as text and biometric signals. The COGNOS project exploits results from two areas of this thesis: it uses the approaches to recognize stress for doting the agents with the capability to detect the psychological state of the users, and it uses the work on agent-based simulation techniques for stress regulation in the design of emotion regulation techniques. These approaches are results from **Objective 1** and **Objective 2**.

- The European project TETRAMAX VLP-Automation (Task Automation based on Visible Light Positioning and Blockchain) European project[3] (Sep 2021 - May 2022) aims to design and implement a task automation and access management platform based on blockchain and Visible Light Communication technologies. The developed platform is able to detect context-related information, such as the location of objects or people using the LED lighting infrastructure and the automation of tasks based on this information. The platform used in this project has been developed as part of the **Objective 3** of this thesis.

- The international project EmoSpaces (Enhanced Affective Wellbeing based on Emotion Technologies for adapting IoT spaces) [4] (May 2016 - Sep 2019) aimed to develop an IoT platform that provides context-awareness capabilities with a focus on emotion detection and ambient adaptation. The goal was to achieve emotion-aware task automation, allowing for the adaptation of various parameters of the environment as a response to the context and emotions of the inhabitants. The project aimed to provide technologies for analyzing the emotion of users in intelligent spaces and adapting these spaces to enhance the user experience. These technologies were designed to be applied in a variety of settings, including retail, home, educational, office, and tourism environments. The automation platform used in this project was a direct outcome of this thesis (**Objective 3**).

- The national project SEMOLA (Tecnologías de Análisis de Sentimientos y emociones para agentes sociales empáticos en inteligencia ambiental) [5] (May 2016 - Dec 2018) is focused on researching models, techniques, and tools for developing emotion-aware personal agents. This research is intended to facilitate the management of user rela-

[3]TETRAMAX VLP-Automation project: `https://short.upm.es/rmdft`
[4]EmoSpaces project: `https://short.upm.es/1sykz`
[5]SEMOLA project: `https://short.upm.es/weip1`

tionships with smart social environments based on ambient intelligence. To achieve these goals, this research project aims to explore the use of various recognition techniques and models for the analysis of sentiments and emotions in the context of social networks and intelligent environments. To this end, the project will utilize Semantic Web and Linked Data technologies, natural language processing, and social simulation. Additionally, the project aims to research customization services based on the context provided by ambient intelligence devices in intelligent environments. Two main areas of this thesis are exploited by the SEMOLA project: it uses the automation platform to automate tasks in the intelligent environment (outcome from **Objective 3**), and it uses the agent-based simulation system to analyze emotion regulation techniques (outcome from **Objective 2**).

- Use by industry and community. The proposed semantic automation platform has been used by industry partners such as Spanish (e.g., Experis, Answare, Taiger), and European (e.g., BEIA) SMEs. These companies have used the platform mainly to automate tasks and to dote an environment with adaption capabilities.

The software tools developed in this thesis (e.g., the automation platform and the agent-based simulation system) have been published with open source licenses. These tools are listed in Table 4.6 and provide a practical application of the research conducted in this thesis.

Table 4.6: Software tools developed in this thesis.

Name	**URL**
EweTasker Server	`https://github.com/gsi-upm/ewetasker-server`
EweTasker Web Client	`https://github.com/gsi-upm/ewetasker-webclient`
EweTasker Mobile	`https://github.com/gsi-upm/ewetasker-mobile`
Agent-based Simulation Stress	`https://github.com/gsi-upm/agent-based-stress`

4.4 Conclusions

The first group of contributions of this thesis included an approach to detect psychological stress based on non-obtrusive methods. This approach has been tackled from two different perspectives: through the exploitation of textual data and through the analysis of personal and social stress-related information.

Regarding the analysis of textual data, in this thesis, we have studied the effectiveness of surface and deep features with traditional machine learning models. With this aim, we have proposed a lexicon-based feature framework that enables the categorization and organization of features relevant to stress detection in text. The features that compose this framework have been extracted from a total of eleven distinct lexicons. Such a framework also offers insight into the effectiveness of various types of features and lexicons. The features have been divided into four distinct categories depending on their nature: affective, social, topic, and syntactic. We have also evaluated the effectiveness of these features, studying which features perform better depending on their nature or lexicon. Besides, we also combined these features with distributional representations exploiting three word embedding techniques. Our experiments demonstrated that this combination is very effective and can yield high overall performance. The validation of this approach has been conducted through an experiment with three English public datasets from diverse sources and nature. This validation provides a baseline for other researchers. The advancements in automatic stress detection from text presented here have the potential to facilitate early stress detection, which is crucial for expediting its diagnosis and minimizing its growing impact. Applied to the workplace, such an approach could benefit from textual data present in existing communication platforms or forums of a company to assess the stress levels of the employees.

Regarding the analysis of personal and social stress-related data, we have presented in this thesis a novel method to predict stress using these data. This method leverages surrounding stress-related data. These data include past levels from the individuals and their nearest fellows. As far as we know, this study is the first to examine the potential of using this kind of data to detect stress. Our findings support the feasibility of predicting occupational stress through the use of surrounding stress-related data. Furthermore, we analyzed three distinct forms to calculate the surrounding stress data: (i) including only personal data, (ii) including only social data, and (iii) including both personal and social data. This allowed us to gain insights into the separate impact of the personal and social stress components. The results support the idea that combining personal and social data leads to higher performance and demonstrate the efficacy of combining various sources of stress information utilized by these methods. The outcomes show that this combination leads to a significant improvement compared to the other methods. To increase this performance, we also exploited three different manners to combine the information: (i) using only sequential data, (ii) using only statistical features, and (iii) using a feature ensemble of sequential and statistical data. The results demonstrate that combining sequential and statistical features is the most effective method in this regard. The model consistently produces the best results across all scenarios.

Our second group of contributions is on the use of agent-based simulation to analyze well-being promotion techniques. In this regard, this thesis proposes a stress model to analyze stress regulation policies in the workplace. The proposed model is integrated into an agent-based simulation system also proposed in this thesis. The model incorporates several aspects related to occupational stress grouped into workload and ambient conditions. The system proposed enables the simulation of a work environment, facilitating the analysis of the progression of employees' stress and performance over time and allowing users to easily configure and evaluate tailored adaptation strategies. In addition, a set of experiments have been conducted with this system to analyze four stress regulation strategies, evaluating how diverse adaptation levels impact stress and productivity. The results indicate that ambient conditions adaptation, by itself, is not particularly effective at decreasing stress in the workplace. However, when combined with workload adaptation, it results in a 10% reduction in employees' stress. Furthermore, the sensibility analysis conducted confirms this finding and demonstrates that the workload condition is the input parameter that most impacts stress and productivity. Lastly, these results are further validated through an experiment carried out in a real scenario. The responses to the self-reported questionnaires show the positive impact of automatic adaptation on stress regulation. Furthermore, the analysis of the biometric data indicates that stress levels are lower in the presence of adaptation, which is consistent with the conclusions drawn. Therefore, the system offers relevant insight into exploiting agent-based simulation technologies to study well-being promotion strategies.

As a result of this work, an open-source tool is provided for surveying and analyzing diverse stress regulation strategies. The tool enables the assessment of the effectiveness of each strategy in the regulation of workers' stress and the enhancement of their productivity. Four different stress regulation policies have been evaluated with the tool, and the obtained results are also presented. The conclusions drawn from this experiment evidence that implementing an automation platform in the workplace is worth it for the objective of lowering stress. Automating ambient and workload conditions is the strategy that best manages to regulate stress. Nevertheless, seeing that the automation of workload conditions has a substantially higher impact than ambient automation, a solution that implements only the former could be preferred depending on the financial constraints.

Lastly, the third group of contributions contemplates the implementation of automatic systems for regulating and promoting welfare. This thesis proposes an architecture for a task automation platform that integrates through semantic technologies sensors and actuators intended to enable the detection and regulation of emotions. The automation platform is characterized by event-driven semantic rules that can be configured and customized by users in a straightforward manner. These rules facilitate the automated adjustment of the workplaces to the needs and emotions of employees to help them cope with these emotions and increase their well-being and productivity. To achieve this goal, a prototype has been developed that includes various sensors and actuators for assessing and regulating different personal and ambient conditions.

Given that the architecture is based on semantic rule-based automation, this thesis also discloses the modeling of all system components. This helps to promote data interoperability and portability of automations. The proposed model has been defined on the basis of popular ontologies related to rule-based automation (EWE), affects (Onyx), and smart environments (SSN).

The platform has been implemented in a real scenario and evaluated experimentally. Through this evaluation, we have verified that using the proposed automation platform helps to improve users' mental and emotional well-being. Furthermore, the obtained results have also demonstrated that the proposed system enhances users' performance and satisfaction when implemented in a work environment. Such outcomes demonstrate the potential of automation-based systems to promote well-being, given the advantages they provide for users (e.g., regulation of stress or emotions and personalized workspaces). Actually, the reference implementation has already been utilized internally (e.g., bachelor and master theses) and by third parties (e.g., Emospaces and Tetramax projects).

Besides, the proposed platform has been applied to an e-learning scenario, enabling the analysis of emotions in study places and their automated adaption to the students' needs. In this manner, the platform helps to decrease communication barriers among teachers and students in remote or e-learning settings and empowers students with emotion regulation capabilities. Finally, the platform has been extended and provided with visible light communication capabilities.

In conclusion, this thesis advances research on the use of technology to promote mental and emotional well-being in the workplace. It encompasses such advancement following an interdisciplinary approach that entails using several technologies and fosters research in three different areas. Besides, as a result of this thesis, the following hypotheses have been supported:

Hypothesis 1 (H_1). Social, environmental, and personal-related signals can help to predict work-related mental disorders using non-obtrusive techniques.
This hypothesis has been supported by the proposal of two solutions to detect stress using non-obtrusive methods. The first solution consisted of detecting stress from text using a combination of lexicon-based features and word embeddings, as described in the article **"A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations"** (Section 3.2.1). The

second solution was based on the use of surrounding stress-related data, that is, the previous stress level of an individual and the stress levels of their near colleagues. This solution was described in the article **"Prediction of stress levels in the workplace using surrounding stress"** (Section 3.2.2). The positive results obtained with the experimental validation of both approaches demonstrate the validity of this hypothesis.

Hypothesis 2 (H_2). The development of mental disorder models can help understand each factor's effects for proposing suitable regulation techniques.
Such as hypothesis has been supported by the findings of our work described in **"An agent based simulation system for analyzing stress regulation policies at the workplace"** (Section 3.2.4). The work demonstrates the usefulness of a stress model to identify the influence of different factors on occupational stress and provide further insight into their foundations and regulation.

Hypothesis 3 (H_3). Mental and emotional well-being in the workplace can benefit from the seamless integration of social and personal sensors and actuators.
The last hypothesis has been validated with the development of an automation platform described in **"An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices"** (Section 3.2.6). The experimental validation of the platform demonstrated its usefulness for regulating emotions in the workplace and improving users' well-being. Besides, this hypothesis is further confirmed by the outcomes obtained with subsequent enhancements described in the works **"An Emotion-Aware Learning Analytics System Based on Semantic Task Automation"** (Section 3.2.5), and **"Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices"** (Section 3.2.3).

4.5 Future research

Over the course of this thesis, we have identified a number of possible directions in which our research could be broadened or continued. This section discusses these directions and providers a summary of possible lines for future research.

First of all, we believe our findings on the impact of surrounding stress on workers provide new avenues for research that could contribute to enhancing future detection solutions. These enhancements may facilitate a better comprehension of occupational stress, its impact on work-related health, and human resources management. Although the use of surrounding stress has shown promising results for predicting stress, its application as a complementary method together with another approach has yet to be fully exploited. The combination of our approach with a method based on physiological or behaviour signals has the potential to enhance stress prediction performance while decreasing the quantity of data needed. However, further research is required to assess and evaluate the extent of this enhancement.

Besides, the proposed approach for predicting stress from surrounding data poses certain challenges that could be addressed by future research. The first of these challenges is the expansion of the dataset with additional experiments conducted in real-world settings. Increasing the volume of data may allow the system to take advantage of more sophisticated neural network models, which are usually more effective at handling sequential data. Also, the extension of the dataset can enable long-term prediction with longer sequences and the evaluation that sequence length has on prediction performance.

Future work on the proposed approach for predicting stress could also focus on understanding the role of weekends in stress levels. In addition to allowing predictions to be made between different weeks, this would help to identify factors related to weekends, such as rest or leisure activities and other parameters that may contribute to reducing work-related stress. Additionally, investigating patterns or factors associated explicitly with remote work could provide valuable insights into how to mitigate stress in this context.

Regarding the detection of stress in text, one of the critical strengths of the proposed approach is its feasibility for deployment in real-world environments. Its good performance and low computational resource requirements make it a practical and cost-effective option for predicting stress. This is particularly beneficial for organizations that may have limited resources or need to deploy the solution in multiple locations. Nevertheless, further exploration of how to increase the reliability, robustness, and performance of the proposed approach would be very interesting. In this line, pre-trained models like MentalBERT could be trained to take into consideration our feature framework. Such a combination of pre-trained models with the proposed framework may yield significant performance in stress detection. Also, to further exploring the use of word embeddings to study semantic similarity among texts may be a promising approach for improving the accuracy and potential of stress detection systems.

Another potential area of future work is to extend the domain of the proposed approach to additional languages or even to diverse paradigms. For example, adapting the method to other languages could enable its use in multilingual environments or in countries where English is not the dominant language. Additionally, exploring the applicability of the method to different paradigms, such as different domains or contexts, could provide insights into the generalizability and adaptability of the approach. This includes using the proposed approach to detect related mental issues like anxiety or depression.

Regarding the use of agent-based simulation for analyzing regulation policies, our approach is the first to include work and ambient conditions in modeling occupational stress. However, there are still important factors that could be included to increase the reliability and accuracy of the model. In this line, a possible direction is the integration of the personality factor. Providing agents with different personalities allows researchers in the field to analyze the influence of personal characteristics that lead to different ways of coping with the workload and ambient conditions. Another possible factor to include is the individual's emotional state to analyze how different emotions or moods influence the stress level.

Another potential direction for future work is the incorporation of room offices and locations into the model. This would allow us to analyze the influence of the office layout on workers' stress. Besides, applying a contagion model would allow the system to evaluate the stress contagion among workers working close. In this regard, we are currently working on the integration of the SIR model, described in Section 1.2, into our stress model. Lastly, further validation of the proposed system would help to demonstrate its reliability. In this regard, extending the validation method would be interesting for future work. Such extension may consider the conduction of additional experiments, increasing the number of participants, and including a pairwise comparison method. This could provide a more robust and comprehensive evaluation of the proposed approach, helping to identify additional strengths and weaknesses and leading to its further enhancement.

Lastly, our work in the area of environmental adaptation also opens several lines for future research. One potential area for future work is to further explore the context of mental and emotional regulation. This could involve the integration of more sophisticated ways for smart environments to adapt to emotions, which could be a key factor in improving the effectiveness of these systems. In this line, further research on the use of music and color to automatically regulate emotions with environment adaptation is an interesting approach. Also, additional regulation methods based on mindfulness or emotional intelligence could be included. This would allow users to choose between various regulation techniques depending on their preferences or personal experiences.

In addition to the workplace, the proposed system could potentially be applied to a wide range of other scenarios. The high scalability of the system, facilitated by the developed architecture and tools, makes it relatively easy to extend to new scenarios, potentially increasing its overall utility and impact. For example, its application would be exciting in healthcare or home scenarios, particularly to enhance the health and well-being of elderly or vulnerable people.

To conclude, a promising potential application of the data generated by the platform is the use of machine learning techniques to learn from the behavior of users over time. This would also allow the system to improve the efficacy of adaptations or even the recommendation of automation rules between users.

Bibliography

Abdelhamid, Sherif Elmeligy et al. (2016). "Agent-based modeling and simulation of depression and its impact on student success and academic retention". In: *2016 ASEE Annual Conference & Exposition*.

Acampora, Giovanni, Diane J. Cook, et al. (2013). "A survey on ambient intelligence in healthcare". In: *Proceedings of the IEEE* 101.12, pp. 2470–2494.

Acampora, Giovanni and Autilia Vitiello (2013). "Interoperable neuro-fuzzy services for emotion-aware ambient intelligence". In: *Neurocomputing* 122, pp. 3–12.

Alberdi, Ane, Asier Aztiria, and Adrian Basarab (2016). "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review". In: *Journal of Biomedical Informatics* 59, pp. 49–75. ISSN: 1532-0464.

Alberdi, Ane, Asier Aztiria, Adrian Basarab, and Diane J Cook (2018). "Using smart offices to predict occupational stress". In: *International Journal of Industrial Ergonomics* 67, pp. 13–26.

Almatarneh, Sattam and Pablo Gamallo (2018). "A lexicon based method to search for extreme opinions". In: *PLoS ONE* 13.5, e0197816.

Alpaydin, Ethem (2020). *Introduction to machine learning*. MIT press.

Altieri, Alex, Silvia Ceccacci, and Maura Mengoni (2019). "Emotion-aware ambient intelligence: changing smart environment interaction paradigms through affective computing". In: *International Conference on Human-Computer Interaction*. Springer, pp. 258–270.

Appelbaum, Eileen et al. (2000). *Manufacturing advantage: Why high-performance work systems pay off*. Cornell University Press.

Areàn, Patricia A., Kien Hoa Ly, and Gerhard Andersson (2022). "Mobile technology for mental health assessment". In: *Dialogues in Clinical Neuroscience* 18.2, pp. 163–169.

Augusto, Juan Carlos et al. (2008). "Management of uncertainty and spatio-temporal aspects for monitoring and diagnosis in a smart home". In: *International Journal of Computational Intelligence Systems* 1.4, pp. 361–378.

Badham, Jennifer et al. (2018). "Developing agent-based models of complex health behaviour". In: *Health & place* 54, pp. 170–177.

Bandhakavi, Anil et al. (2017). "Lexicon based feature extraction for emotion text classification". In: *Pattern recognition letters* 93, pp. 133–142.

Banerjee, Satyaki and Nuzhat F. Shaikh (2021). "A Survey on Mental Health Monitoring System Via Social Media Data Using Deep Learning Framework". In: *Techno-Societal 2020*. Cham: Springer International Publishing, pp. 879–887.

Barry, Margaret M et al. (2019). *Implementing mental health promotion*. Springer.

Bazzan, Ana L. C. and Franziska Klügl (2014). "A review on agent-based technology for traffic and transportation". In: *The Knowledge Engineering Review* 29.3, pp. 375–403.

Bentley, Todd, Lorraine Johnston, and Karola von Baggo (2002). "Putting some emotion into requirements engineering". In: *Proceedings of the 7th Australian workshop on requirements engineering*. Citeseer, pp. 227–244.

Berners-Lee, Tim and Dan Connolly (2008). *Notation3 (N3): A readable RDF syntax*. Tech. rep. W3C. URL: http://www.w3.org/TeamSubmission/n3/.

Bettis, Alexandra H. et al. (2022). "Digital technologies for emotion-regulation assessment and intervention: A conceptual review". In: *Clinical Psychological Science* 10.1, pp. 3–26.

Bhatia, Munish and Sandeep K Sood (2019). "Exploring temporal analytics in fog-cloud architecture for smart office healthcare". In: *Mobile Networks and Applications* 24.4, pp. 1392–1410.

Bikakis, Antonis et al. (2007). "A survey of semantics-based approaches for context reasoning in ambient intelligence". In: *European Conference on Ambient Intelligence*. Springer, pp. 14–23.

Boley, Harold, Adrian Paschke, and Omair Shafiq (2010). "RuleML 1.0: The Overarching Specification of Web Rules". In: *Semantic Web Rules: International Symposium, RuleML 2010, Washington, DC, USA, October 21-23, 2010. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 162–178. ISBN: 978-3-642-16289-3.

Bonabeau, Eric (2002). "Agent-based modeling: Methods and techniques for simulating human systems". In: *Proceedings of the national academy of sciences* 99.suppl_3, pp. 7280–7287.

Borgonovo, Emanuele et al. (2022). "Sensitivity analysis of agent-based models: a new protocol". In: *Computational and Mathematical Organization Theory* 28.1, pp. 52–94.

Brouwer, Anne-Marie et al. (2013). "Perceiving blocks of emotional pictures and sounds: effects on physiological variables". In: *Frontiers in human neuroscience* 7, p. 295.

Campos, Joseph J., Carl B. Frankel, and Linda Camras (2004). "On the nature of emotion regulation". In: *Child development* 75.2, pp. 377–394.

Can, Yekta Said, Bert Arnrich, and Cem Ersoy (2019). "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey". In: *Journal of biomedical informatics* 92, p. 103139.

Cao, Lei et al. (2021). "Category-Aware Chronic Stress Detection on Microblogs". In: *IEEE Journal of Biomedical and Health Informatics* 26.2, pp. 852–864.

Castaldo, Rossana et al. (2019). “Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life”. In: *BMC medical informatics and decision making* 19.1, pp. 1–13.

Castillo, José Carlos et al. (2016). “Software architecture for smart emotion recognition and regulation of the ageing adult”. In: *Cognitive Computation* 8.2, pp. 357–367.

Chalabi, Zaid and Theo Lorenc (2013). “Using agent-based models to inform evaluation of complex interventions: examples from the built environment”. In: *Preventive medicine* 5.57, pp. 434–435.

Chancellor, Stevie and Munmun De Choudhury (2020). “Methods in predictive techniques for mental health status on social media: a critical review”. In: *NPJ digital medicine* 3.1, pp. 1–11.

Chen, Deng-Neng et al. (2008). “An agent-based model for consumer-to-business electronic commerce”. In: *Expert Systems with Applications* 34.1, pp. 469–481.

Cohen, Sheldon, William J. Doyle, et al. (1995). “State and trait negative affect as predictors of objective and subjective symptoms of respiratory viral infections.” In: *Journal of personality and Social Psychology* 68.1, p. 159.

Cohen, Sheldon, Tom Kamarck, Robin Mermelstein, et al. (1994). “Perceived stress scale”. In: *Measuring stress: A guide for health and social scientists* 10.2, pp. 1–2.

Compton, Michael et al. (2012). “The SSN ontology of the W3C semantic sensor network incubator group”. In: *Journal of Web Semantics* 17, pp. 25–32.

Coronado, Miguel and Carlos A Iglesias (2015). “Task automation services: automation for the masses”. In: *IEEE Internet Computing* 20.1, pp. 52–58.

Coronado, Miguel, Carlos A Iglesias, and Emilio Serrano (2015). “Modelling rules for automating the Evented WEb by semantic technologies”. In: *Expert Systems with Applications* 42.21, pp. 7979–7990.

Coronato, A., G. De Pietro, and M. Esposito (2006). “A semantic context service for smart offices”. In: *2006 International Conference on Hybrid Information Technology*. Vol. 2. IEEE, pp. 391–399.

Dababneh, Awwad J., Naomi Swanson, and Richard L. Shell (2001). “Impact of added rest breaks on the productivity and well being of workers”. In: *Ergonomics* 44.2, pp. 164–174.

Dacunhasilva, Dennis Rodrigo, Zelun Wang, and Ricardo Gutierrez-Osuna (2021). “Towards Participant-Independent Stress Detection Using Instrumented Peripherals”. In: *IEEE Transactions on Affective Computing*.

De Simone, Stefania (2014). “Conceptualizing Wellbeing in the Workplace”. In: *International Journal of Business and Social Science* 5.12.

Decker, Stefan et al. (2000). “The semantic web: The roles of XML and RDF”. In: *IEEE Internet computing* 4.5, pp. 63–73.

Dimitroff, Stephanie J. et al. (2017). “Physiological dynamics of stress contagion”. In: *Scientific reports* 7.1, pp. 1–8.

Duggirala, Mayuri et al. (2016). “Understanding Impact of Stress on Workplace Outcomes Using an Agent Based Simulation”. In: *Proceedings of the Summer Computer Simulation Conference*. SCSC ’16. Montreal, Quebec, Canada: Society for Computer Simulation International, 35:1–35:10. ISBN: 978-1-5108-2424-9.

Elkosantini, Sabeur (2015a). “Toward a new generic behavior model for human centered system simulation”. In: *Simulation Modelling Practice and Theory* 52, pp. 108–122.

— (2015b). “Toward a new generic behavior model for human centered system simulation”. In: *Simulation Modelling Practice and Theory* 52, pp. 108–122. ISSN: 1569-190X.

Engert, Veronika, Roman Linz, and Joshua A. Grant (2019). “Embodied stress: The physiological resonance of psychosocial stress”. In: *Psychoneuroendocrinology* 105. Festschrift for Dirk Hellhammer, pp. 138–146. ISSN: 0306-4530.

Ferdous, Raihana, Venet Osmani, and Oscar Mayora (2015). “Smartphone app usage as a predictor of perceived stress levels at workplace”. In: *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, pp. 225–228.

Fernández-Caballero, Antonio et al. (2016). “Smart environment architecture for emotion detection and regulation”. In: *Journal of biomedical informatics* 64, pp. 55–73.

Fornell, Claes (1994). “Partial least squares”. In: *Advanced methods of marketing research*.

Fu, Xianghua et al. (2018). “Lexicon-enhanced LSTM with attention for general sentiment analysis”. In: *IEEE Access* 6, pp. 71884–71891.

Gams, Matjaz et al. (2019). “Artificial intelligence and ambient intelligence”. In: *Journal of Ambient Intelligence and Smart Environments* 11.1, pp. 71–86.

García-Magariño, Iván and Inmaculada Plaza (2017). “ABSEM: An agent-based simulator of emotions in mindfulness programs”. In: *Expert Systems with Applications* 84, pp. 49–57.

Giakoumis, Dimitris et al. (Sept. 2012). “Using Activity-Related Behavioural Features towards More Effective Automatic Stress Detection”. In: *PLoS ONE* 7.9, pp. 1–16.

Giannakakis, G et al. (2017). “Stress and anxiety detection using facial cues from videos”. In: *Biomedical Signal Processing and Control* 31, pp. 89–101.

Giatsoglou, Maria et al. (2017). “Sentiment analysis leveraging emotions and word embeddings”. In: *Expert Systems with Applications* 69, pp. 214–224.

Gilbody, Simon M, Allan O House, and Trevor A Sheldon (2001). “Routinely administered questionnaires for depression and anxiety: systematic review”. In: *Bmj* 322.7283, pp. 406–409.

Giuntini, Felipe T. et al. (2020). “A review on recognizing depression in social networks: challenges and opportunities”. In: *Journal of Ambient Intelligence and Humanized Computing* 11.11, pp. 4713–4729.

Goodday, Sarah M. and Stephen Friend (2019). “Unlocking stress and forecasting its consequences with digital technology”. In: *NPJ digital medicine* 2.1, pp. 1–5.

Gordon, Amie M and Wendy Berry Mendes (2021). “A large-scale study of stress, emotions, and blood pressure in daily life using a digital platform”. In: *Proceedings of the National Academy of Sciences* 118.31.

Graham, Sarah et al. (2019). “Artificial intelligence for mental health and mental illnesses: an overview”. In: *Current psychiatry reports* 21.11, pp. 1–18.

Grassi, Marco (2009). “Developing HEO human emotions ontology”. In: *European workshop on biometrics and identity management*. Springer, pp. 244–251.

Greene, Shalom, Himanshu Thapliyal, and Allison Caban-Holt (2016). “A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health”. In: *IEEE Consumer Electronics Magazine* 5.4, pp. 44–56.

Gross, James J. (2013). *Handbook of emotion regulation*. Guilford publications.

— (2015). “Emotion regulation: Current status and future prospects”. In: *Psychological inquiry* 26.1, pp. 1–26.

Gruber, Thomas R. (1993). “A translation approach to portable ontology specifications”. In: *Knowledge acquisition* 5.2, pp. 199–220.

Gupta, Sahil and Swati Gupta (2021). “Analysis and Design of Smart Office Automation System”. In: *Computer Networks and Inventive Communication Technologies*. Ed. by S. Smys et al. Singapore: Springer Singapore, pp. 963–973.

Haeringen, E. S. van, C. Gerritsen, and K. V. Hindriks (2023). “Emotion contagion in agent-based simulations of crowds: a systematic review”. In: *Autonomous Agents and Multi-Agent Systems* 37.1, pp. 1–41.

Hafner, Marco et al. (2015). *Health, wellbeing and productivity in the workplace: A Britain’s Healthiest Company summary report*. Santa Monica, CA: RAND Corporation.

Haller, Armin et al. (2019). “The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation”. In: *Semantic Web* 10.1, pp. 9–32.

Han, Hee Jeong et al. (2020). “Objective stress monitoring based on wearable sensors in everyday settings”. In: *Journal of Medical Engineering & Technology* 44.4, pp. 177–189.

Handouzi, Wahida et al. (2014). “Objective model assessment for short-term anxiety recognition from blood volume pulse signal”. In: *Biomedical Signal Processing and Control* 14, pp. 217–227.

Harnois, Gaston et al. (2000). *Mental health and work: impact, issues and good practices*. Tech. rep. URL: https://hdl.handle.net/1813/76479.

Hartmann, Jochen et al. (2019). "Comparing automated text classification methods". In: *International Journal of Research in Marketing* 36.1, pp. 20–38.

Hawe, Glenn I. et al. (2012). "Agent-based simulation for large-scale emergency response: A survey of usage and implementation". In: *ACM Computing Surveys (CSUR)* 45.1, pp. 1–51.

Hayatnagarkar, Harshal et al. (2016). "Can a buffering strategy reduce workload related stress?" In: *Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems*. Springer, pp. 378–388.

Heath, Brian, Raymond Hill, and Frank Ciarallo (2009). "A survey of agent-based modeling practices (January 1998 to July 2008)". In: *Journal of Artificial Societies and Social Simulation* 12.4, p. 9.

Hollis, Victoria et al. (2017). "What does all this data mean for my future mood? Actionable analytics and targeted reflection for emotional well-being". In: *Human–Computer Interaction* 32.5-6, pp. 208–267.

Horton, William (2011). *E-learning by design*. John Wiley & Sons.

Jager, Wander (2017). "Enhancing the realism of simulation (EROS): On implementing and developing psychological theory in social simulation". In: *Journal of Artificial Societies and Social Simulation* 20.3.

Janowicz, Krzysztof et al. (2019). "SOSA: A lightweight ontology for sensors, observations, samples, and actuators". In: *Journal of Web Semantics* 56, pp. 1–10.

Jia, Moyi and Jiuqing Cheng (2021). "Emotional experiences in the Workplace: Biological Sex, Supervisor Nonverbal Behaviors, and Subordinate Susceptibility to Emotional Contagion". In: *Psychological Reports* 124.4, pp. 1687–1714.

Joulin, Armand et al. (2016). "Fasttext. zip: Compressing text classification models". In: *arXiv preprint arXiv:1612.03651*.

Kalton, Alan et al. (2016). "Multi-agent-based simulation of a complex ecosystem of mental health care". In: *Journal of medical systems* 40.2, pp. 1–8.

Keller, John and Katsuaki Suzuki (2004). "Learner motivation and e-learning design: A multinationally validated process". In: *Journal of educational Media* 29.3, pp. 229–239.

Khoo, Christopher SG. and Sathik Basha Johnkhan (2018). "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons". In: *Journal of Information Science* 44.4, pp. 491–511.

Kifer, Michael (2008). "Rule Interchange Format: The Framework". In: *Web Reasoning and Rule Systems: Second International Conference, RR 2008, Karlsruhe, Germany, October*

31-November 1, 2008. Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–11. ISBN: 978-3-540-88737-9.

Koolagudi, Shashidhar G. and K. Sreenivasa Rao (2012). "Emotion recognition from speech: a review". In: *International journal of speech technology* 15.2, pp. 99–117.

Krekel, Christian, George Ward, and Jan-Emmanuel De Neve (2019). "Employee wellbeing, productivity, and firm performance". In: *Saıd Business School WP* 4.

Kushlev, Kostadin and Elizabeth W. Dunn (2015). "Checking email less frequently reduces stress". In: *Computers in Human Behavior* 43, pp. 220–228.

Leong, Chui Yew, Abdul Rahman Ramli, and Thinagaran Perumal (2009). "A rule-based framework for heterogeneous subsystems management in smart home environment". In: *IEEE Transactions on Consumer Electronics* 55.3, pp. 1208–1213.

Lin, Huijie et al. (2016). "What Does Social Media Say about Your Stress?." In: *IJCAI*, pp. 3775–3781.

Liu, Zhenyu et al. (2015). "Detection of depression in speech". In: *2015 international conference on affective computing and intelligent interaction (ACII)*. IEEE, pp. 743–747.

Loo Gee, Brendan, Kathleen M. Griffiths, and Amelia Gulliver (2016). "Effectiveness of mobile technologies delivering Ecological Momentary Interventions for stress and anxiety: a systematic review". In: *Journal of the American Medical Informatics Association* 23.1, pp. 221–229.

Lyubomirsky, Sonja, Laura King, and Ed Diener (2005). "The benefits of frequent positive affect: Does happiness lead to success?" In: *Psychological bulletin* 131.6, p. 803.

Maxwell, Lorraine E. (2016). "Noise in the office workplace". In: *Facility Planning and Management Notes* 1.11.

McDaid, David, Claire Curran, and Martin Knapp (2005). "Promoting mental well-being in the workplace: a European policy perspective". In: *International review of psychiatry* 17.5, pp. 365–373.

Mikolov, Tomas et al. (2013). "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*, pp. 3111–3119.

Moore, Kimberly Sena (2013). "A systematic review on the neural effects of music on emotion regulation: implications for music therapy practice". In: *Journal of music therapy* 50.3, pp. 198–242.

Mouromtsev, Dmitry (2021). "Semantic Reference Model for Individualization of Information Processes in IoT Heterogeneous Environment". In: *Electronics* 10.20. ISSN: 2079-9292.

Novais, Paulo and Davide Carneiro (2016). "The role of non-intrusive approaches in the development of people-aware systems". In: *Progress in Artificial Intelligence* 5.3, pp. 215–220.

O'Connor, Martin et al. (2005). "Supporting Rule System Interoperability on the Semantic Web with SWRL". In: *The Semantic Web – ISWC 2005: 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 974–986. ISBN: 978-3-540-32082-1.

Ortiz-García-Cervigón, Vicente et al. (2015). "LED strips for color-and illumination-based emotion regulation at home". In: *International Work-Conference on Ambient Assisted Living.* Springer, pp. 277–287.

Page, Matthew and Daniel Ashlock (2015). "Stress and productivity performance in the workforce modelled with binary decision automata". In: *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2015 IEEE Conference on.* IEEE, pp. 1–8.

Pakarinen, Tomppa, Julia Pietilä, and Hannu Nieminen (2019). "Prediction of Self-Perceived Stress and Arousal Based on Electrodermal Activity". In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* IEEE, pp. 2191–2195.

Palavalli, Amarnath, Durgaprasad Karri, and Swarnalatha Pasupuleti (2016). "Semantic internet of things". In: *2016 IEEE Tenth International Conference on Semantic Computing (ICSC).* IEEE, pp. 91–95.

Pang, Jianhui et al. (2019). "Fast supervised topic models for short text emotion detection". In: *IEEE Transactions on Cybernetics* 51.2, pp. 815–828.

Panicker, Suja Sreeith and Prakasam Gayathri (2019). "A survey of machine learning techniques in physiology based mental stress detection systems". In: *Biocybernetics and Biomedical Engineering* 39.2, pp. 444–469.

Pedrini, Laura et al. (2022). "School-based interventions to improve emotional regulation skills in adolescent students: A systematic review". In: *Journal of Adolescence.*

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Peters, Dorian, Rafael A. Calvo, and Richard M. Ryan (2018). "Designing for motivation, engagement and wellbeing in digital experience". In: *Frontiers in psychology* 9, p. 797.

Petitta, Laura et al. (2021). "The impact of emotional contagion on workplace safety: Investigating the roles of sleep, health, and production pressure". In: *Current Psychology*, pp. 1–15.

Pham, Minh et al. (2021). "Negative emotion management using a smart shirt and a robot assistant". In: *IEEE Robotics and Automation Letters* 6.2, pp. 4040–4047.

Picard, Rosalind W. (2000). *Affective computing.* MIT press.

Picard, Rosalind W. and Jennifer Healey (1997). "Affective wearables". In: *Personal technologies* 1.4, pp. 231–240.

Prasetyo, T. F., D. Zaliluddin, and M. Iqbal (2018). "Prototype of smart office system using based security system". In: *Journal of Physics: Conference Series*. Vol. 1013. IOP Publishing, p. 012189.

Rahim, Aqsa et al. (2019). "Emotion charting using real-time monitoring of physiological signals". In: *2019 International Conference on Robotics and Automation in Industry (ICRAI)*. IEEE, pp. 1–5.

Rahman, Hafizur and Md. Iftekhar Hussain (2020). "A comprehensive survey on semantic interoperability for Internet of Things: State-of-the-art and research challenges". In: *Transactions on Emerging Telecommunications Technologies* 31.12, e3902.

Šabanović, Selma et al. (2013). "PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia". In: *2013 IEEE 13th international conference on rehabilitation robotics (ICORR)*. IEEE, pp. 1–6.

Sadat-Mohammadi, Milad et al. (2021). "Non-invasive physical demand assessment using wearable respiration sensor and random forest classifier". In: *Journal of Building Engineering* 44, p. 103279.

Sadri, Fariba (2011). "Ambient intelligence: A survey". In: *ACM Computing Surveys (CSUR)* 43.4, pp. 1–66.

Saltelli, Andrea and Paola Annoni (2010). "How to avoid a perfunctory sensitivity analysis". In: *Environmental Modelling & Software* 25.12, pp. 1508–1517.

Sánchez-Rada, J. Fernando and Carlos A. Iglesias (2016). "Onyx: A linked data approach to emotion representation". In: *Information Processing & Management* 52.1, pp. 99–114.

Santarossa, Sara et al. (2018). "Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice?" In: *Journal of medical Internet research* 20.4, e8480.

Saxena, Anvita, Ashish Khanna, and Deepak Gupta (2020). "Emotion recognition and detection methods: A comprehensive survey". In: *Journal of Artificial Intelligence and Systems* 2.1, pp. 53–79.

Scheuermann, Constantin et al. (2020). "Digital twin: A machine learning approach to predict individual stress levels in extreme environments". In: *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pp. 657–664.

Sekulić, Ivan, Matej Gjurković, and Jan Šnajder (2018). "Not Just Depressed: Bipolar Disorder Prediction on Reddit". In: *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 72–78.

Shamil, Md et al. (2021). "An agent-based modeling of COVID-19: validation, analysis, and recommendations". In: *Cognitive Computation*, pp. 1–12.

Shi, Feifei et al. (2018). "A survey of data semantization in internet of things". In: *Sensors* 18.1, p. 313.

Shu, Lin et al. (2018). "A review of emotion recognition using physiological signals". In: *Sensors* 18.7, p. 2074.

Siettos, Constantinos I. and Lucia Russo (2013). "Mathematical modeling of infectious disease dynamics". In: *Virulence* 4.4, pp. 295–306.

Silverman, Barry G. (2001). "More realistic human behavior models for agents in virtual worlds: Emotion, stress, and value ontologies". In: *Center for Human Modeling and Simulation*, p. 34.

Singh, Meghendra et al. (2016). "Multi-agent model of workgroup behavior in an enterprise using a compositional approach". In: *2nd Modeling Symposium (ModSym 2016). ISEC.*

Skaramagkas, Vasileios et al. (2021). "Review of eye tracking metrics involved in emotional and cognitive processes". In: *IEEE Reviews in Biomedical Engineering.*

Slade, Mike (2009). *Personal recovery and mental illness: A guide for mental health professionals.* Cambridge University Press.

Sobkowicz, Pawel (2016). "Agent based model of effects of task allocation strategies in flat organizations". In: *Physica A: Statistical Mechanics and its Applications* 458, pp. 17–30.

Sokolova, Marina V. et al. (2015). "Evaluation of color preference for emotion regulation". In: *International Work-Conference on the Interplay Between Natural and Artificial Computation.* Springer, pp. 479–487.

Somanathan, Eswaran et al. (2021). "The impact of temperature on productivity and labor supply: Evidence from Indian manufacturing". In: *Journal of Political Economy* 129.6, pp. 1797–1827.

Song, Tengfei et al. (2018). "EEG emotion recognition using dynamical graph convolutional neural networks". In: *IEEE Transactions on Affective Computing* 11.3, pp. 532–541.

Sovacool, Benjamin K. and Dylan D. Furszyfer Del Rio (2020). "Smart home technologies in Europe: a critical review of concepts, benefits, risks and policies". In: *Renewable and sustainable energy reviews* 120, p. 109663.

Spitzer, Robert L et al. (1999). "Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study". In: *Journal of the American Medical Association* 282.18, pp. 1737–1744.

Sroykham, Watchara, J. Wongsathikun, and Yodchanan Wongsawat (2014). "The effects of perceiving color in living environment on QEEG, oxygen saturation, pulse rate, and emotion regulation in humans". In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, pp. 6226–6229.

Szeto, Andrew C. H. and Keith S. Dobson (2010). “Reducing the stigma of mental disorders at work: a review of current workplace anti-stigma intervention programs”. In: *Applied and Preventive Psychology* 14.1-4, pp. 41–56.

Tracy, Melissa, Magdalena Cerdá, and Katherine M Keyes (2018). “Agent-based modeling in public health: current applications and future directions”. In: *Annual review of public health* 39, p. 77.

Ullo, Silvia Liberata and G. R. Sinha (2020). “Advances in Smart Environment Monitoring Systems Using IoT and Sensors”. In: *Sensors* 20.11. ISSN: 1424-8220.

Verborgh, Ruben and Jos De Roo (2015). “Drawing conclusions from linked data on the web: The EYE reasoner”. In: *IEEE Software* 32.3, pp. 23–27.

Wang, Rui et al. (2014). “StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones”. In: *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pp. 3–14.

Wang, Shirui, Wenan Zhou, and Chao Jiang (2020). “A survey of word embeddings based on deep learning”. In: *Computing* 102.3, pp. 717–740.

Wasil, Akash R. et al. (2019). “A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content”. In: *Behaviour research and therapy* 123, p. 103498.

Weitzel, Leila, Ronaldo Cristiano Prati, and Raul Freire Aguiar (2016). “The comprehension of figurative language: What is the influence of irony and sarcasm on NLP techniques?” In: *Sentiment analysis and ontology engineering*. Springer, pp. 49–74.

William, David and Derwin Suhartono (2021). “Text-based depression detection on social media posts: A systematic literature review”. In: *Procedia Computer Science* 179, pp. 582–589.

Wisniewski, Dawid et al. (2019). “Analysis of Ontology Competency Questions and their formalizations in SPARQL-OWL”. In: *Journal of Web Semantics* 59, p. 100534.

Yang, Kangning et al. (2021). “Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition”. In: *IEEE Transactions on Affective Computing*, pp. 1–1.

Yang, Shanwen et al. (2020). “A review on crowd simulation and modeling”. In: *Graphical Models* 111, p. 101081.

Yerkes, Robert Mearns and John D. Dodson (1908). “The relation of strength of stimulus to rapidity of habit formation”. In: *Journal of Comparative Neurology and Psychology* 18, pp. 459–482.

Yue, Lin et al. (2019). “A survey of sentiment analysis in social media”. In: *Knowledge and Information Systems* 60.2, pp. 617–663.

Zapf, Dieter et al. (2021). “Emotion Work: A Work Psychology Perspective”. In: *Annual Review of Organizational Psychology and Organizational Behavior* 8.1, pp. 139–172.

Zhang, Jin et al. (2019). “Detecting negative emotional stress based on facial expression in real time”. In: *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. IEEE, pp. 430–434.

Zhang, Ting et al. (2021). “Interaction of emotional regulation and outdoor thermal perception: A pilot study in a cold region of China”. In: *Building and Environment* 198, p. 107870.

Zhou, Jiehan and Päivi Kallio (2005). “Ambient emotion intelligence: from business awareness to emotion awareness”. In: *The 17th International Conference on Systems Research*, pp. 47–54.

List of Figures

List of Tables

Glossary

AmI Ambient Intelligence
ARI Automated Readability Index
CQ Competency Question
ECA Event-Condition-Action
EF Effective Fatigue
ES Event Stress
EWE Evented Web
EYE Euler Yet another proof Engine
FKG Flesch-Kincaid Grade Level
GALC Geneva Affect Label Coder
GI General Inquirer
GloVe Global Vectors for Word Representation
IoT Internet of Things
LED Light Emitting Diode
LIWC Linguistic Inquiry and Word Count
N3 Notation 3
RDF Resource Description Framework
SOSA Sensor, Observation, Sample, and Actuator
SSN Semantic Sensor Network
PMF Performance Moderator Functions
TAS Task Automation Service
VLC Visible Light Communication
TP Time Pressure
WBGT wet-bulb globe temperature

APPENDIX A

Publications

A.1 Summary of all publications

The table below provides an overview of all the publications of this thesis. The following color scheme has been used: gray for conference papers, and blue for journal papers.

Table A.1: Summary of all publications made throughout this thesis

Page	Title	Year	Venue	Ranking
35	A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations	2022	Information Processing & Management	JCR 2021 Q1 (7.466)
55	Prediction of stress levels in the workplace using surrounding stress	2022	Information Processing & Management	JCR 2021 Q1 (7.466)
69	Semantic Modeling of a VLC-Enabled Task Automation Platform for Smart Offices	2022	Electronics	JCR 2021 Q3 (2.690)

Table A.1: Summary of all publications made throughout this thesis

89	An agent based simulation system for analyzing stress regulation policies at the workplace	2021	Journal of Computational Science	JCR 2021 Q1 (3.817)
203	JAICOB: A Data Science Chatbot	2020	IEEE Access	JCR 2019 Q1 (3.745)
103	An Emotion-Aware Learning Analytics System Based on Semantic Task Automation	2020	Electronics	JCR 2021 Q3 (2.690)
213	A Cognitive Agent for Mining Bugs Reports, Feature Suggestions and Sentiment in a Mobile Application Store	2018	2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)	
128	An Emotion Aware Task Automation Architecture Based on Semantic Technologies for Smart Offices	2018	Sensors	JCR 2021 Q1 (3.847)

A.2 Publications not directly related to the core of the thesis.

A.2.1 JAICOB: A Data Science Chatbot

Title	JAICOB: A Data Science Chatbot
Authors	Carlander-Reuterfelt, Daniel and Carrera, Alvaro and Iglesias, Carlos A. and Araque, Oscar and Sanchez Rada, Juan Fernando Sanchez and Munoz, Sergio
Journal	IEEE Access
Impact factor	JCR 2019 Q1 (3.745)
ISSN	2169-3536
Publisher	Institute of Electrical and Electronics Engineers
Volume	8
Year	2020
Keywords	Cognitive informatics, educational technology, human-computer interaction, machine learning, natural language processing
Pages	180672–180680
Online	https://ieeexplore.ieee.org/document/9200315/
Abstract	The application of natural language to improve students' interaction with information systems is demonstrated to be beneficial. In particular, advances in cognitive computing enable a new way of interaction that accelerates insight from existing information sources, thereby contributing to the process of learning. This work aims at researching the application of cognitive computing in blended learning environments. We propose a modular cognitive agent architecture for pedagogical question answering, featuring social dialogue (small talk), improved for a specific knowledge domain. This system has been implemented as a personal agent to assist students in learning Data Science and Machine Learning techniques. Its implementation includes the training of machine learning models and natural language understanding algorithms in a humanlike interface. The effectiveness of the system has been validated through an experiment.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 10, 2020, accepted August 17, 2020, date of publication September 18, 2020, date of current version October 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3024795

JAICOB: A Data Science Chatbot

DANIEL CARLANDER-REUTERFELT, ÁLVARO CARRERA, CARLOS A. IGLESIAS, ÓSCAR ARAQUE, JUAN FERNANDO SÁNCHEZ RADA, AND SERGIO MUÑOZ

Grupo de Sistemas Inteligentes, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Corresponding author: Álvaro Carrera (a.carrera@upm.es)

This work was supported by the Educational Innovation Programme of Universidad Politécnica de Madrid under Grant IE1819.0908.

ABSTRACT The application of natural language to improve students' interaction with information systems is demonstrated to be beneficial. In particular, advances in cognitive computing enable a new way of interaction that accelerates insight from existing information sources, thereby contributing to the process of learning. This work aims at researching the application of cognitive computing in blended learning environments. We propose a modular cognitive agent architecture for pedagogical question answering, featuring social dialogue (small talk), improved for a specific knowledge domain. This system has been implemented as a personal agent to assist students in learning Data Science and Machine Learning techniques. Its implementation includes the training of machine learning models and natural language understanding algorithms in a human-like interface. The effectiveness of the system has been validated through an experiment.

INDEX TERMS Cognitive informatics, educational technology, human-computer interaction, machine learning, natural language processing.

I. INTRODUCTION

Cognitive computing has grown in the last few years, increasing the research and commercial interest in the topic [1]. Conversational agents have evolved from simple pattern-based programs into rather complex systems, including Natural Language Understanding and Machine Learning Techniques, which have allowed them to be more flexible in maintaining a conversation. Every day more businesses include chatbots as a way to interact with consumers to answer requests and FAQs. Natural Language Interface (NLI) increases user satisfaction and can help to find the information needed in a more comfortable way than other less sophisticated and time-consuming search interfaces [2].

Like humans, cognitive systems can use their knowledge to deduce data meaning based on context [3]. By having the advantage of computational power, a system like this can be even more successful than a human in this kind of task. Though they do not understand the meaning as humans do, the insights these systems provide can be beneficial. As they grow in time, it is expected that they gain abilities such as sensing and awareness [4].

Some of the benefits of the application of cognitive computing in the development of learning applications are: (1) They can actively enhance students' performances [5], especially in computer science classes [6]; (2) studying cognitive computing behavior can lead to significant results in educational applications, especially in AI-related studies [6]; (3) *using a cognitive computing layer for digital interactions with students can enhance their performances and ease the teachers' job in managing classes and learning materials [6];* and (4) chatbots are excellent analysis tools, as students feel more inclined to send more messages to chatbots than real people [2].

Compared to other traditional e-learning training, chatbots generate a more positive response from the users [7]. Moreover, there are advantages in this type of learning, such as interaction, active learning, and sociability [8].

Despite these reasons, these technologies have not been widely adopted yet in education, and the ones that have are usually very rule-based and, therefore, less practical and functional. This article presents a modular architecture chatbot named Jaicob, adapted to the learning of Data Science techniques that aims to take advantage of all the benefits for education previously described. It is designed in a modular way that allows its adaptation to other areas of knowledge. It includes a flexible conversation workflow and is easy to maintain. This contribution has been evaluated with real users for a specific use case in a Data Science class.

The associate editor coordinating the review of this manuscript and approving it for publication was Tony Thomas.

The rest of the paper is organized as follows. Section II analyses related works about chatbots and the techniques applied in their development. Section III describes the different modules of the architecture and how they are interconnected. Section IV describes the evaluation process and results. Finally, Section V summarizes the learnings of this article with conclusions and defining future works.

II. RELATED WORK

A comprehensive systematic review of the use of chatbots in education is provided in this recent survey [9]. The authors identify several perspectives for analyzing current research following the theoretical model of Technology-Mediated Learning (TML) [10]: structure (input), learning process (process), and learning outcome (output). Regarding the input perspective, several dimensions have been identified [9]: student profile, educational settings, and chatbot technology. Learning outputs depend on individual student characteristics such as personality traits, technological skills as well as educational and social background [9].

Some research works claim that chatbot technology is so disruptive that *it will eliminate the need for websites and apps* [11]. Chatbots have been used in different educational settings, such as language learning [12], health-related coaching agents [13], chatbots designed to provide feedback to students [14], programming language learning [15], administrative support [16] or increase students' motivation [17]. These are examples if we don't take into account open-domain solutions such as Amazon or Google's [18], which aim to answer any kind of question, instead of a specific area of knowledge. While these types of chatbots are astoundingly ambitious and function with a near-human precision, sometimes, they come at a very high cost. Closed-domain question answering systems benefit from the ability to respond with more profound and specific knowledge [19], and also can achieve high quality at a lower complexity cost.

Design aspects of chatbots can influence the learning process. Flow-based chatbots, like [20]–[22], also called rule-based, can require an extensive database of questions and answers and need to have a clear flow of conversation that, if the user decides not to follow, can result in a bad experience. A study on chatbots of this type [23] concludes that they are quite limited to human direction and control. These can be built with frameworks like *Landbot.ai*,[1] or with simple coding abilities, but require great sophistication to work correctly. There is where its limitations lie. An extension to this kind of bots is button-based, like *HelloFreshus*,[2] that avoids the possibility of exiting the pre-planned flow. These can work well but can be very limited in scope and depth.

On the other hand, artificial intelligent based chatbots can better understand student intents. Even the most simple non-rule-based natural language understanding methods significantly outperform the most carefully crafted rule-based systems [24]. The reason is that they can achieve a more profound understanding of the intent and the requested information, thanks to machine learning techniques [25]. The most usual and effective approach [26], which is explained in greater detail in Section III, is based on intent-entity and Knowledge Base (KB).

Another aspect to take into account is if they are text or voice-based. Users tend to use longer sentences with voice-based chatbots and prefer reading expanded answers in a text-like manner. However, there is no significant difference in perceived effectiveness, learnability, and humanness between text-based and voice-based chatbots [27].

III. PROPOSED ARCHITECTURE FOR THE COGNITIVE BOT

The first step to design the proposed architecture was to identify the way students learn and the types of questions. Different types of requirements for different types of learning (inductive and deductive) [28] were identified due to the nature of students' curiosity, and the specifics of the topic. The following **pedagogical solutions** were identified:

1) A **definition** of a concept is a consequence of the usual teaching style, which is deductive, starting from the main concepts and developing towards the applications. It is part of the process of learning, but cannot be the whole process. In the Oliver model [29], definitions provide learning content.
2) As stated in [30], the learning of programming techniques can be enhanced by using **examples** of code using analogy [31] and induction. Also, learning is significantly facilitated by examples in initial coding attempts. Furthermore, surveys suggest that engineering students usually view themselves as inductive learners [28]. In the Oliver model [29], examples can provide learner support.
3) Lastly, the human need for **small-talk**, such as joking and asking for the weather, must be satisfied to provide a more significant communication source [2].

With that in mind, the architecture was designed, having identified the pedagogical needs of the student. There are several steps involved in the process and are explained below and represented in Figure 1.

A Knowledge Base (KB) was populated with pertinent information regarding the topic at hand, to satisfy the requests for definitions and examples. The Question Answering (QA) module is designed to extract meaning from all the data with the pedagogical requirements in mind to make sense of that information.

To analyze the students' question, we use the Speech Act Classifier. It selects the module where the question must be delegated. The way it works will be explained in greater detail in Section III-B. If small talk is detected, it is passed onto the Small Talk module or into the QA Module if a question regarding Data Science is detected.

[1] https://landbot.io
[2] https://chatfuel.com/bot/HelloFreshus

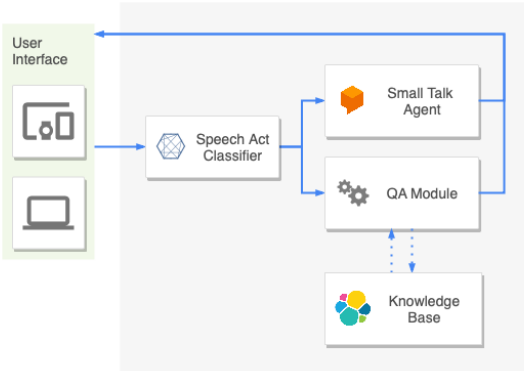


FIGURE 1. System architecture.

Afterward, the modules generate an answer to satisfy the student request. The answer is sent back to the student, and feedback is collected to evaluate and improve the model.

A. KNOWLEDGE BASE

The KB is the place where all the information used by the chatbot is stored. It has been populated from several online academic resources. Its selection was based on the previously identified student interactions with the chatbot. Glossaries and Frequently Asked Questions (FAQs) of the topic have been mined using web scraping techniques to provide concept definitions. Regarding searching code snippets, technical documentation has also been mined.

This approach benefits from providing students a curated list of pedagogical sources that are credible and useful. According to some studies [32], undergraduates tend to use Google for searching for information, and the usage of academic resources is low. Thus, our system increases the use of curated academic resources since the bot can enhance their familiarity [33].

The adaptation of the bot to other domains could be made replicating the same approach.

According to the categories previously described, the sites that fit the necessities of the definition answering are: (1) **Big Data glossary**[3] with a list of terms regarding big data. (2) **Machine Learning glossary**[4] with a complete glossary of machine learning and statistics terms and definitions.

The documentation sites used to populate the KB for answering with examples are *Pandas Documentation*. The use of the Python Pandas library is widely used when developing machine learning models. It is beneficial to have examples available for standard implementations of data handling. This documentation is structured with brief descriptions with code examples; and *Scikit-Learn Documentation*, being the library used widely for Machine Learning purposes, Scikit

[3]Big Data glossary: https://bigdata-madesimple.com/big-data-a-to-zz-a-glossary-of-big-data-terminology/

[4]Machine Learning glossary: https://www.analyticsvidhya.com/glossary-of-common-statistics-and-machine-learning-terms/

TABLE 1. Post classification examples.

Classification	Example	Next Module
Wh-Question	What is an elasticsearch database?	QA
Yes Answer	yes I do, lol	Small Talk
Yes/No Question	is overfitting a real problem	QA
Emotion	lol	Small Talk

examples of implementations is an obvious use case for the chatbot, and therefore an nearly important part of the KB.

For more complex questions, the use of FAQs solves the problem. The *Machine Learning Mastery* site[5] used for this purpose is structured as a list of questions with the answers associated. It was selected because of the rich and adequate answer for the project.

B. SPEECH ACT CLASSIFIER

The speech act classification task involves classifying a specific sentence into a set of predefined speech act categories. This classification is relevant to the project because it is indispensable to know the student's intention [34] and answer accordingly.

The dataset [35] used to train the classifier consists of 10567 posts from five different age-oriented chat rooms at an internet chat site. It is sanitized to protect user privacy. The posts were tagged using 15 post categories (*Accept, Bye, Clarify, Continuer, Emphasis, Greet, No Answer, Other, Reject, Statement, System, Wh-question, Yes Answer, Yes/No question, Emotion*). Examples of these classes are shown in Table 1.

Since this is a chatbot system that requires a fast response time, the preprocessing has been simplified to improve the model's time complexity while not sacrificing relevant performance. Each phrase is processed into machine-understandable information using a raw pipeline. The overall process is (1) Simple tokenization because n-grams did not present a significant improvement in accuracy, (2) Stemming, and (3) Feature extraction by vectorization.

By training and evaluating some of the most popular classification algorithms, the best one is selected based on the score achieved by a K-Fold. This process can be automated through a grid search that finds the best parameters optimally.

The scores following are calculated, saving a fourth of the dataset for testing afterward and using the rest to get these results.

The scores shown in Table 2 are obtained by performing a 5-Fold and calculating the mean of the scores. Support Vector Machines, such as the SVC algorithm, achieves higher performance.

Using as the training data 3/4 of the dataset and the rest as testing data, we obtain with the SVC a final accuracy score of 0.799.

[5]Site for complex questions: https://machinelearningmastery.com/faq/

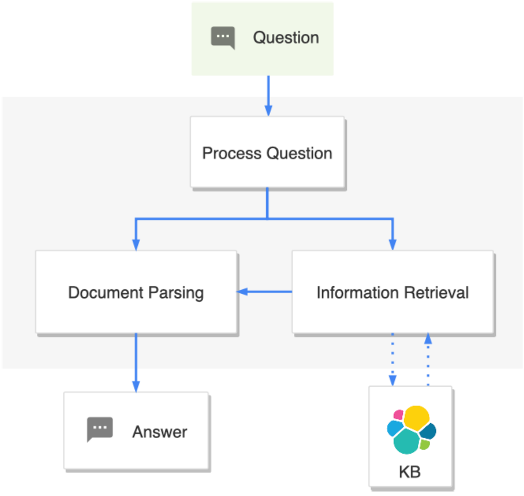


FIGURE 2. QA architecture.

TABLE 2. Evaluation scores.

Model	Accuracy	Precision	Recall	F1
Multinomial NB	0.685 (+/- 0.012)	0.75	0.71	0.66
Decision Tree	0.711 (+/- 0.005)	0.76	0.76	0.75
Random Forest	0.745 (+/- 0.005)	0.78	0.77	0.76
SVC	0.768 (+/- 0.007)	0.79	0.8	0.77

C. QUESTION ANSWERING MODULE

The Question Answering module comes into place when the user asks for a specific piece of information. These can range from doubt, a consultation, or documentation clarifications. It must be able to understand what the user is asking for to retrieve the information effectively.

Using natural language processing techniques, it answers the question in near real-time. This general-purpose model is enhanced to attend specific cases to the task at hand, such as code examples.

The general view of the architecture is defined in Figure 2. The modules involved in the process are the following.

The **Process Question** module extracts the relevant information and intention of the question. The output contains a type of question, a type of answer, and a vector with the relevant ideas.

The **Information Retrieval** This module receives the `question vector` and the `answer type` from the question processor as an input. The question vector is, in essence, a list of keywords ordered by importance. An ElasticSearch query is generated to retrieve relevant documents and pieces of information that match the keywords, using this valuable information.

The **Document Parsing** module receives and parses the retrieved information, so it matches the questions intented to generate an answer.

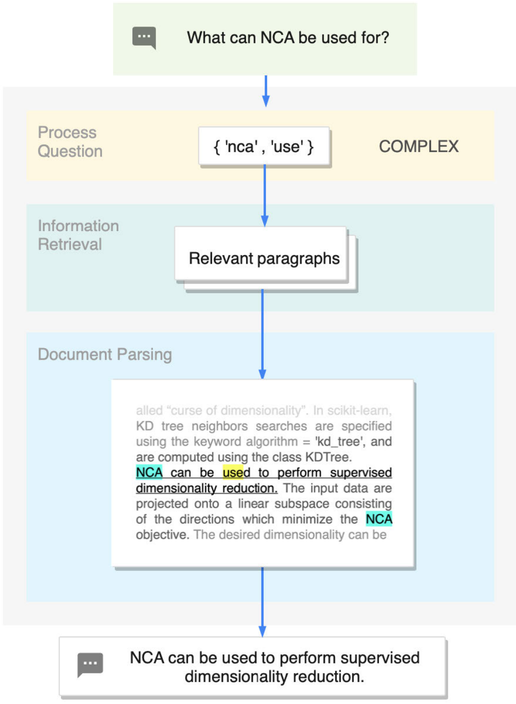


FIGURE 3. QA example.

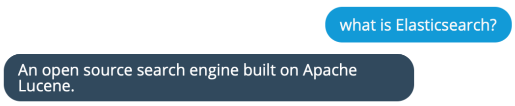


FIGURE 4. Definition intent use case.

1) DEFINITION ANSWERING

When the `Answer Type` is of the *definition* type, the module searches in the Knowledge Base's Glossary index. It searches for a match with the terms in the index. When a match is found, the corresponding definition is sent as an answer. Common questions of this type are:

- What is a neural network?
- Can you give me a definition of overfitting?

This module is implemented as a DialogFlow agent, with an intent to recognize that the user wants a definition. The intent is trained with multiple training phrases that can be used to ask for a definition. It extracts a term as the slot. These slots are recognized thanks to an entity[6] defined as all the terms available in the Knowledge Base. An example can be seen in Figure 4.

[6] An entity is the definition of a type of slot. It can be defined as a list with all the possible terms for that slot.

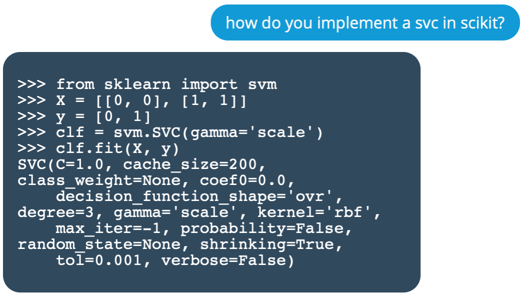


FIGURE 5. Example intent use case.

2) EXAMPLE ANSWERING

When the `Answer Type` is of the *example* type, we need a more complex type of search. There is a search across the documentation text to match the keywords of the query. When a match is found, the corresponding code snippet is sent to respond with the appropriate format. Examples of these type of questions are:

- How is a dataframe defined in Pandas?
- How can I implement a k-fold using scikit?

This module is implemented as a DialogFlow agent with and intent trained to detect example queries. The slot, in this case, is more open, so there is no Entity defined. The example can be of any kind. The result can be seen in Figure 5.

D. SMALL TALK MODULE

According to [36], the users' satisfaction with a certain chatbot is influenced by various factors. By testing which of these factors were more influential, the results revealed that the bot's human-likeness was significantly correlated with the users' satisfaction.

It was stated [2] that people were *inclined to send more than twice as many messages to chatbots with a human-like interaction compared to other people, contrary to our expectations and disconfirming the notion that people feel less confident or comfortable communicating with chatbots.*

Including a module to handle small-talk improves the bot's human-likeness and makes it more fun and engaging. Instead of answering with the fallback answer, if the question is not about the topic, it triggers the small-talk module to simulate human interaction and cleverness. Some examples of the behavior that the bot can answer are collected in Figure 6.

1) IMPLEMENTATION

This module is implemented with Google's DialogFlow technology. There is a specially trained agent to provide the desired output. This agent can detect more than 100 different intents.

Among these intents are some of the provided with the default Small Talk module and some custom ones. The intents are defined to fit the purpose of this project. For example, when asked what it can do, it responds with directions to ask questions about Data Science.

FIGURE 6. Small talk examples.

TABLE 3. Intent distribution.

Intent	Queries Rate
Example Intent	36.5%
Definition Intent	25.7%
Complex	21.1%
Small Talk	26.7%

E. GRAPHICAL USER INTERFACE

The bot needs an identifier to generate a more personal relationship [37]. Being a Cognitive bot and an intelligent one, it was decided to be called *Just an Artificial Intelligence Cognitive Bot* (JAICOB).

In contrast with Jaicob, a general-purpose bot would gain quality from a text-to-speech transformer, giving it a more human appearance. This feature is not the case of Jaicob because it is centered on answering documentation and programming related questions. The frequent use of acronyms and code examples in the answers would not make for a pleasant listening experience. Instead, the use of text is the best option in this case.

IV. EVALUATION

A. EVALUATION TECHNIQUES

The evaluation method for Jaicob chatbot is a Partial Least Squares (PLS) analysis. A detailed example [38] is followed to perform PLS methods. The tool being used is Smart PLS.[7]

The method is based on a questionnaire and requires the definition of latent variables to be evaluated, which are abstract variables that are connected to directly measurable variables. These variables' values are scored by the

[7]https://www.smartpls.com/documentation/pls-sem-compared-with-cb-sem

TABLE 4. Outer loadings.

Constructs and Items	Outer loading	CR	AVE
Social Handling (original scale, 4 items)		0.9132	0.7246
SH1: I felt a sense of human contact in the agent.	0.8608		
SH2: I felt a sense of personality in the agent.	0.8902		
SH3: I felt a sense of human warmth in the agent.	0.8327		
SH4: I felt a sense of sociability in the agent.	0.8196		
Behavioral intentions (original scale, 3 items)		0.9551	0.8764
R1: I would be willing to talk positively about JAICOB to my friends.	0.9319		
R2: I would be willing to recommend JAICOB to my friends.	0.9313		
R3: When visiting forums, I would be willing to say positive things about JAICOB to other people.	0.9451		
Satisfaction (original scale, 4 items)		0.9287	0.7655
S1: After visiting the website, I am satisfied.	0.8103		
S2: After visiting the website, I am happy.	0.8977		
S3: I found the visit on the website pleasant.	0.8614		
S4: Compared to other things I could have done, the time spent talking with JAICOB was enjoyable.	0.9261		
Utilitarian value (original scale, 3 items)		0.8891	0.7285
U1: Using JAICOB would make my life easier.	0.7812		
U2: I think of this Web site as an expert.	0.8815		
U3: If I want to solve doubts, the information JAICOB is able to provide would be what I would look for.	0.8933		
Answer Accuracy (original scale, 4 items)		0.9147	0.7288
A1: JAICOB responded effectively to my questions regarding definitions of terms.	0.8569		
A2: JAICOB was able to answer my doubts about code implementations correctly.	0.7866		
A3: The answer JAICOB gave me was what I was looking for.	0.9116		
A4: When he did not have the answer he told me so in a touchful way.	0.855		

responses of the questionnaires. These latent variables can also have relations, and these can be hypothesized, as shown in Section IV-C.

Being a conversational interface, the way to test it is with real users who answer the questionnaire after using the chatbot. The number of observations (number of questionnaires answered by users) should be at least ten times the number of relations between latent variables.

B. PARTICIPANTS

The experiment was done with 50 participants, all of them with technical backgrounds. All of them were unaware of the inner workings of Jaicob. They were asked to use the chatbot as a tool to answer any questions or doubts that may arise in understanding Data Science related topics or writing the corresponding code.

The median of the ages of the participants is 22 years. A 51% of them were studying a Telecommunication Engineering Grade and the rest a Master or superior studies.

About their technological background, 54% of the participants had developed and implemented some machine learning programs. The rest had some basic knowledge in the field.

C. EXPERIMENT DESIGN

As explained in Section III-D, small talk is an essential part of the architecture of the chatbot. Therefore, before making the measurements, it is taken into account.

Five latent variables were defined to evaluate the conversational bot:

- **Social Handling (SH)** refers to the personality and human-likeness of the bot.

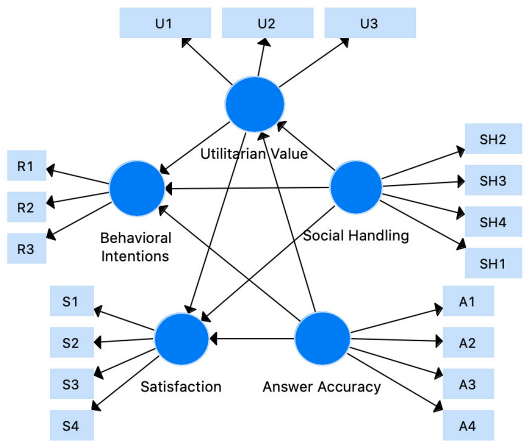


FIGURE 7. Structural paths in the applied PLS model.

- **Behavioral intentions (BI)** refers to the recommendation of users to others to use the bot.
- **Satisfaction (SS)** refers to the feeling after using the bot.
- **Utilitarian value (UV)** refers to the value it provides to the task you are looking to complete.
- **Answer Accuracy (AA)** refers to the performance in the task it was programmed to do.

These latent variables are not independent, as represented in Figure 7. They present relations between latent variables, which are hypothesized and tested. Moreover, the relations between latent variables and questions, summarized in Table 4, are shown in the structural paths of the applied PLS Model.

Research [39] suggests that the quality of information on an e-commerce website has a positive impact on perceived value. Reference [40] suggests that accurate information can help users make better decisions, thus improving both utilitarian values. According to [41], the utilitarian values increases when the interaction with the process improves. These hypotheses are proposed:

H1. Perceptions of a better answer accuracy improve utilitarian value.

User satisfaction is influenced by the human-likeness of the chatbot [36]. Also, [2] state that people are more inclined to send messages to a chatbot that handles this type of small-talk well. A website's social dimension is another important antecedent of perceived value [42]. Research [43], [44] reveals that there is a direct link between perceived sociability and satisfaction.

H2. The social handling of the bot improves the overall satisfaction of the user.

H3. The social handling of the bot improves the utilitarian value a user perceives.

H4. Good social handling improves the behavioral intentions of users after using the bot.

Utilitarian value is central to user satisfaction and behavioral intentions. If the perceived value is low, the user probably switches to other sources [39].

H5. A higher perceived answer accuracy value increases positive behavioral intentions.

H6. A higher perceived utilitarian value increases positive behavioral intentions.

Perceived utilitarian value also enhances satisfaction [40]. Research [45] demonstrates that utilitarian value can improve the final user satisfaction:

H7. Perceived utilitarian value has a positive effect on user satisfaction.

H8. Perceived answer accuracy has a positive effect on user satisfaction.

D. RESULTS

The testers made an average of 15.86 queries per session. The intent that matched most of the queries was related to code example requests, which means that users used the bot for what it was intended. After that, there is the Definition intent and then the complex intent. Also, 26.7% of the queries resulted in small talk handling. The distribution can be seen in Table 3.

The results extracted from the PLS modeling, having used *SmartPLS 3.0* [46] meet the requirements, being the sample size ten times the largest number of structural paths directed at a particular construct in the structural model. There are three paths directed to Behavioral Intentions and Satisfaction in this model, so the minimum sample size should be 30, and the sample size is above this minimum.

To test the experiment's internal coherence, and therefore, reliability, we look at the outer loadings. These coefficients need to meet a threshold for every measure that points to the latent variables. All the measures met this reliability index, as shown in Table 4. The PLS analysis also provides us with the Composite Reliability of each latent variable. This index surpassed the minimum acceptable value of .70 in all variables, being all over .85.

TABLE 5. **Discriminant validity.**

	AA	BI	SS	SH	UV
Answer Accuracy	**0.8537**				
Behavioral Intentions	0.7746	**0.9361**			
Satisfaction	0.8276	0.8009	**0.8749**		
Social Handling	0.5328	0.4089	0.6326	**0.8513**	
Utilitarian Value	0.7152	0.7315	0.7126	0.5002	**0.8535**

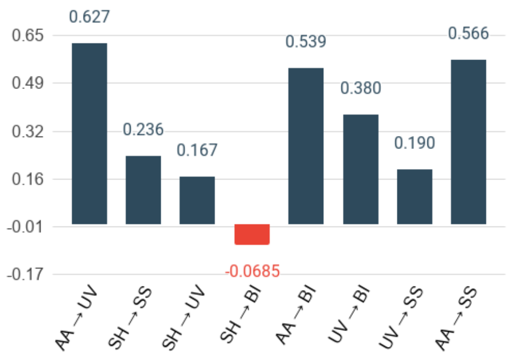


FIGURE 8. **Path coefficients.**

The average variance extracted (AVE) for each variable must surpass a threshold of .50 [43], [47], and provide a square root that is much larger than the correlation of the specific construct with any other construct in the model. All the latent variables surpass a .70 AVE, as shown in Table 4. Table 5 shows that the square roots of the AVE (on the diagonal) are higher than any other values, in support of the discriminant validity of the measurement scales [38].

Then, discriminant validity is tested, which indicates the extent to which a given construct (variable) differs from other latent constructs. The validity of these variables requires that each measurement item correlates weakly with all constructs, except for which it is theoretically associated. The results in Table 5 support the validity of the measurement scales.

All the direct hypotheses received support, except for H4, as shown in Figure 8. From these results, we can extract some insights, such as the impact that Answering Accuracy has on all the other variables. Therefore, the quality of the system and its ability to respond effectively is what makes the difference for overall user Satisfaction, Utilitarian Value, and Behavioral Intentions (H1, H5, H8). Also, the perceived Utilitarian Value has a positive effect on Behavioral Intentions and Satisfaction (H6, H7). Surprisingly, Social Handling was not significant in positive behavioral intentions (H4), contrasting with the Utilitarian Value and Satisfaction (H2, H3).

V. CONCLUSION

The use of chatbots has become prevalent in the last years in shopping, customer support, general assistance, and, though less developed, education. The use of chatbots as a form of e-learning brings lots of opportunities.

This article identified the advantages of cognitive assistants in education and the corresponding challenges in implementation. A result is a tool for students with a comfortable and usable interface and a human experience. It can provide insights and solve doubts about Data Science. The main contribution is the adaptation of students' real pedagogic needs to the design of the architecture and being flexible in maintaining a conversation.

Teachers can also use it as a tool to identify gaps in the knowledge of their students. They can also outsource to Jaicob the answering of all the questions. The pedagogue is also an excellent asset to select the most valuable sources of information from which Jaicob feeds from, thus providing a curated source of information instead of a regular Google Search.

The project was evaluated with a sample of students, achieving very favorable results in usability and originality. The experiment confirms that the system can answer effectively, that the answer accuracy affects the satisfaction, utilitarian value, and behavioral intentions of the user, and that proper social handling is significant in satisfaction and utilitarian value but not in behavioral intentions.

As these technologies evolve, more and more people will study these subjects. Therefore, the future impact of the project is promising, and the affected groups will increase. In future work, to achieve a broader reach in the areas of knowledge, it is straightforward to place additional information in the Knowledge Base and the corresponding Dialogflow intents.

REFERENCES

[1] A. Soffer, D. Konopnicki, and H. Roitman, "When watson went to work: Leveraging cognitive computing in the real world," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2016, pp. 455–456.

[2] J. Hill, W. Randolph Ford, and I. G. Farreras, "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations," *Comput. Hum. Behav.*, vol. 49, p. 245–250, Aug. 2015.

[3] Y. Chen, J. E. Argentinis, and G. Weber, "IBM Watson: How cognitive computing can be applied to big data challenges in life sciences research," *Clin. Therapeutics*, vol. 38, no. 4, p. 688–701, Apr. 2016.

[4] R. High, "The era of cognitive systems: An inside look at IBM Watson and how it works," IBM Corp., Redbooks, North Castle, NY, USA, Tech. Rep., 2012.

[5] M. Coccoli, A. Guercio, P. Maresca, and L. Stanganelli, "Smarter universities: A vision for the fast changing digital era," *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 1003–1011, Dec. 2014.

[6] M. Coccoli, P. Maresca, and L. Stanganelli, "Cognitive computing in education," *Big Data*, vol. 12, no. 2, p. 15, 2016.

[7] S. Kowalski, K. Pavlovska, and M. Goldstein, "Two case studies in using chatbots for security training," in *Proc. IFIP World Conf. Inf. Secur. Educ.* Cham, Switzerland: Springer, 2009, pp. 265–272.

[8] P. Bii, "Chatbot technology: A possible means of unlocking student potential to learn how to learn," *Educ. Res.*, vol. 4, no. 2, pp. 218–221, 2013.

[9] R. Winkler and M. Söllner, "Unleashing the potential of chatbots in education: A state-of-the-art analysis," in *Proc. Acad. Manage. Annu. Meeting*, 2018, pp. 1–40. [Online]. Available: https://www.alexandria.unisg.ch/254848/

[10] R. P. Bostrom, "Technology-mediated learning: A comprehensive theoretical model," *J. Assoc. Inf. Syst.*, vol. 10, no. 9, pp. 686–714, Sep. 2009.

[11] R. Dale, "The return of the chatbots," *Natural Lang. Eng.*, vol. 22, no. 5, pp. 811–817, Sep. 2016.

[12] X. L. Pham, T. Pham, Q. M. Nguyen, T. H. Nguyen, and T. T. H. Cao, "Chatbot as an intelligent personal assistant for mobile language learning," in *Proc. 2nd Int. Conf. Educ. E-Learn. (ICEEL)*, 2018, pp. 16–21.

[13] J. Beaudry, A. Consigli, C. Clark, and K. J. Robinson, "Getting ready for adult healthcare: Designing a chatbot to coach adolescents with special health needs through the transitions of care," *J. Pediatric Nursing*, vol. 49, pp. 85–91, Nov. 2019.

[14] M. Awais Hassan, U. Habiba, H. Khalid, M. Shoaib, and S. Arshad, "An adaptive feedback system to improve student performance based on collaborative behavior," *IEEE Access*, vol. 7, pp. 107171–107178, 2019.

[15] M. Coronado, C. A. Iglesias, Á. Carrera, and A. Mardomingo, "A cognitive assistant for learning java featuring social dialogue," *Int. J. Hum.-Comput. Stud.*, vol. 117, pp. 55–67, Sep. 2018.

[16] H. T. Hien, P.-N. Cuong, L. N. H. Nam, H. L. T. K. Nhung, and L. D. Thang, "Intelligent assistants in higher-education environments: The FIT-EBot, a chatbot for administrative and learning support," in *Proc. 9th Int. Symp. Inf. Commun. Technol. (SoICT)*, New York, NY, USA, 2018, pp. 69–76, doi: 10.1145/3287921.3287937.

[17] L. Benotti, M. C. Martínez, and F. Schapachnik, "Engaging high school students using chatbots," in *Proc. Conf. Innov. Technol. Comput. Sci. Educ. (ITiCSE)*, 2014, pp. 63–68.

[18] D. Adiwardana, M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, and Q. V. Le, "Towards a human-like open-domain chatbot," 2020, *arXiv:2001.09977*. [Online]. Available: http://arxiv.org/abs/2001.09977

[19] H. Doan-Nguyen and L. Kosseim, "Improving the precision of a closed-domain question-answering system with semantic information," in *Proc. Coupling Approaches, Coupling Media Coupling Lang. Inf. Retr.*, Le Centre De Hautes Etudes Internationales D'Informatique Documentaire, 2004, pp. 850–859.

[20] B. Heller, M. Proctor, D. Mah, L. Jewell, and B. Cheung, "Freudbot: An investigation of chatbot technology in distance education," in *Proc. EdMedia+ Innovate Learn. Assoc. Adv. Comput. Educ. (AACE)*, 2005, pp. 3913–3918.

[21] S. Hussain and G. Athula, "Extending a conventional chatbot knowledge base to external knowledge source and introducing user based sessions for diabetes education," in *Proc. 32nd Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, May 2018, pp. 698–703.

[22] J. Jia and W. Chen, "Motivate the learners to practice English through playing with Chatbot CSIEC," in *Proc. Int. Conf. Technol. E-Learn. Digit. Entertainment*. Cham, Switzerland: Springer, 2008, pp. 180–191.

[23] S. Roos, "Chatbots in education: A passing trend or a valuable pedagogical tool?" M.S. thesis, Dept. Inform. Media, Social Sci., Disciplinary Domain Humanities Social Sci., Uppsala Univ., Uppsala, Sweden, 2018, pp. 1–58.

[24] D. Suendermann, K. Evanini, J. Liscombe, P. Hunter, K. Dayanidhi, and R. Pieraccini, "From rule-based to statistical grammars: Continuous improvement of large-scale spoken dialog systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 4713–4716.

[25] S. Paliwal, V. Bharti, and A. K. Mishra, "Ai chatbots: Transforming the digital world," in *Recent Trends and Advances in Artificial Intelligence and Internet of Things*. Cham, Switzerland: Springer, 2020, pp. 455–482.

[26] H. T. Hien, P.-N. Cuong, L. N. H. Nam, H. L. T. K. Nhung, and L. D. Thang, "Intelligent assistants in higher-education environments: The FIT-EBot, a chatbot for administrative and learning support," in *Proc. 9th Int. Symp. Inf. Commun. Technol. (SoICT)*, 2018, pp. 69–76.

[27] O. Mangkang and J. Y. Yun, "A comparative study of self-diagnosis user interfaces for depression: Focusing on usability of paper-based, text-based and voice-based conversational interface," in *Proc. Korean HCI Conf.*, 2019, pp. 262–267.

[28] R. M. Felder and L. K. Silverman, "Learning and teaching styles in engineering education," *Eng. Educ.*, vol. 78, no. 7, pp. 674–681, 1988.

[29] R. Oliver, "Exploring strategies for online teaching and learning," *Distance Educ.*, vol. 20, no. 2, pp. 240–254, Jan. 1999.

[30] P. Brusilovsky, "Webex: Learning from examples in a programming course," in *Proc. WebNet*, vol. 1, 2001, pp. 124–129.

[31] P. L. Pirolli and J. R. Anderson, "The role of learning from examples in the acquisition of recursive programming skills," *Can. J. Psychol./Revue Canadienne de Psychologie*, vol. 39, no. 2, p. 240, 1985.

[32] J. R. Griffiths and P. Brophy, "Student searching behavior and the Web: Use of academic resources and Google," *Trends*, vol. 53, no. 4, p. 539, 2005.

[33] J. Y. Lee, W. Paik, and S. Joo, "Information resource selection of undergraduate students in academic search tasks," *Inf. Res., Int. Electron. J.*, vol. 17, no. 1, p. n1, 2012.
[34] B. Samei, H. Li, F. Keshtkar, V. Rus, and A. C. Graesser, "Context-based speech act classification in intelligent tutoring systems," in *Proc. Int. Conf. Intell. Tutoring Syst.* Cham, Switzerland: Springer, 2014, pp. 236–241.
[35] E. N. Forsythand and C. H. Martell, "Lexical and discourse analysis of online chat dialog," in *Proc. Int. Conf. Semantic Comput. (ICSC)*, Sep. 2007, p. 19.
[36] M. Xuetao, F. Bouchet, and J.-P. Sansonnet, "Impact of agent's answers variability on its believability and human-likeness and consequent chatbot improvements," in *Proc. AISB*, 2009, pp. 31–36.
[37] J. Finch, "Naming names: Kinship, individuality and personal names," *Sociology*, vol. 42, no. 4, pp. 709–725, Aug. 2008.
[38] M. S. Ben Mimoun and I. Poncin, "A valued agent: How ECAs affect website customers' satisfaction and behaviors," *J. Retailing Consum. Services*, vol. 26, pp. 70–82, Sep. 2015.
[39] M. Magni, M. Susan Taylor, and V. Venkatesh, "'To play or not to play': A cross-temporal investigation using hedonic and instrumental perspectives to explain user intentions to explore a technology," *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 9, pp. 572–588, Sep. 2010.
[40] C. Kim, R. D. Galliers, N. Shin, J.-H. Ryoo, and J. Kim, "Factors influencing Internet shopping value and customer repurchase intention," *Electron. Commerce Res. Appl.*, vol. 11, no. 4, pp. 374–387, Jul. 2012.
[41] T. Ahn, S. Ryu, and I. Han, "The impact of Web quality and playfulness on user acceptance of online retailing," *Inf. Manage.*, vol. 44, no. 3, pp. 263–275, Apr. 2007.
[42] J. Lawler, P. Vandepeutte, and A. Joseph, "An exploratory study of apparel dress model technology on European Web sites," *J. Inf., Inf. Technol., Organizations*, vol. 2, no. 244, pp. 31–46, Nov. 2007.
[43] S. S. Al-Gahtani, G. S. Hubona, and J. Wang, "Information technology (IT) in Saudi Arabia: Culture and the acceptance and use of IT," *Inf. Manage.*, vol. 44, no. 8, pp. 681–691, Dec. 2007.
[44] O. Turel, A. Serenko, and N. Bontis, "User acceptance of hedonic digital artifacts: A theory of consumption values perspective," *Inf. Manage.*, vol. 47, no. 1, pp. 53–59, Jan. 2010.
[45] T. W. Traphagan, Y.-H.-V. Chiang, H. M. Chang, B. Wattanawaha, H. Lee, M. C. Mayrath, J. Woo, H.-J. Yoon, M. J. Jee, and P. E. Resta, "Cognitive, social and teaching presence in a virtual world and a text chat," *Comput. Educ.*, vol. 55, no. 3, pp. 923–936, Nov. 2010.
[46] C. Ringle, S. Wende, and J.-M. Becker, "Smartpls 3," SmartPLS GmbH, Bönningstedt, Germany, Tech. Rep., Jan. 2015.
[47] C. Fornell and D. F. Larcker, "Structural equation models with unobservable variables and measurement error: Algebra and statistics," *J. Marketing Res.*, vol. 18, no. 3, pp. 382–388, 1981, doi: 10.1177/002224378101800313.

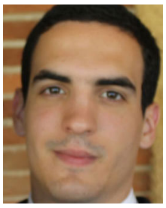
DANIEL CARLANDER-REUTERFELT is currently pursuing the master's degree in telecommunications engineering with the Universidad Politécnica de Madrid (UPM). He was awarded an honorary mention for the bachelor's thesis (Development of a Cognitive Bot for Data Science Tutoring based on a Big Data Natural Language Analytics Platform). He has been a part of the Intelligent Systems Group, UPM, since 2019. His research interests include intelligent agents, natural language processing and understanding, and time series prediction.

ÁLVARO CARRERA received the Telecommunication Engineering degree and the Ph.D. degree in telecommunications engineering from the Universidad Politécnica de Madrid (UPM), Spain, in 2010 and 2016, respectively. He is currently an Assistant Professor with the School of Telecommunications Engineering, UPM. His research interests include applying intelligent agents in telecommunication networks for critical tasks, such as fault diagnosis or intrusion detection in real-time, and in the application of simulation techniques for complex systems in dynamic environments, such as telecommunication networks and social networks.

CARLOS A. IGLESIAS received the Telecommunications Engineering degree and the Ph.D. degree in telecommunications from the Universidad Politécnica de Madrid (UPM), Spain, in 1993 and 1998, respectively. He is currently an Associate Professor with the Telecommunications Engineering School, UPM. He has been leading the Intelligent Systems Group, UPM, since 2014. He has been the Principal Investigator on numerous research grants and contracts in the field of advanced social and the IoT systems, funded by the regional, national, and European bodies. His main research interests include social computing, multiagent systems, information retrieval, sentiment and emotion analysis, linked data, and web engineering.

ÓSCAR ARAQUE received the graduate and master's degrees in telecommunication engineering from the Technical University of Madrid (Universidad Politécnica de Madrid), Spain, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree. He is currently a Teaching Assistant with the Universidad Politécnica de Madrid. His research interest includes the application of machine learning techniques for natural language processing. The main topic of his thesis is the introduction of specific domain knowledge into machine learning systems in order to enhance sentiment and emotion analysis techniques.

JUAN FERNANDO SÁNCHEZ RADA received the Ph.D. degree from the Universidad Politécnica de Madrid (UPM), Spain, in 2020. He is currently a Researcher with the Intelligent Systems Group, Universidad Politécnica de Madrid. His research interests include natural language processing (sentiment and emotion analysis), social network analysis (social context and graph embedding), the web and distributed systems (interoperability and federation), and semantic technologies (linked data, ontologies, and knowledge graphs).



SERGIO MUÑOZ received the graduate and master's degrees in telecommunication engineering from the Technical University of Madrid (Universidad Politécnica de Madrid), Spain, in 2016 and 2017, respectively, where he is currently pursuing the Ph.D. degree. He is currently a Teaching Assistant with the Technical University of Madrid. His research interests include ambient intelligence and agent-based simulation. The main topic of his thesis is the adaptation of smart environments to users' emotions, to enhance well-being and performance.

A.2.2 A Cognitive Agent for Mining Bugs Reports, Feature Suggestions and Sentiment in a Mobile Application Store

Title	A Cognitive Agent for Mining Bugs Reports, Feature Suggestions and Sentiment in a Mobile Application Store
Authors	Munoz, Sergio and Araque, Oscar and Llamas, Antonio F. and Iglesias, Carlos A.
Proceedings	2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)
ISBN	978-1-5386-7793-3
Year	2018
Keywords	cognitive agent, review, app, bug, suggestions, sentiment analysis
Pages	17–24
Online	`https://ieeexplore.ieee.org/document/8500065/`
Abstract	Over the last years, mobile applications and their corresponding distribution platforms have gained momentum. Applications stores allow users to write reviews and ratings about the apps, giving feedback to developers. User ratings and reviews may help to improve software quality, solve bugs and develop new features. However, this data is hard to be handled by an individual due to the ever growing amount of textual reviews. This paper proposes the use of cognitive computing technologies for addressing this challenge, by developing a smart agent able to mine bugs reports, feature suggestions and sentiment expressed in mobile app reviews. The main contributions of this paper are: the design of a cognitive agent for assisting developers in managing their interaction with their users, the application of machine learning algorithms for bug and feature request detection, and the agent implementation in a real scenario.

2018 4th International Conference on Big Data Innovations and Applications

A Cognitive Agent for Mining Bugs Reports, Feature Suggestions and Sentiment in a Mobile Application Store

Sergio Muñoz, Oscar Araque, Antonio F. Llamas and Carlos A. Iglesias
Intelligent Systems Group
Universidad Politécnica de Madrid
{sergio.munoz, o.araque, carlosangel.iglesias}@upm.es, {a.fllamas}@alumnos.upm.es

***Abstract*—Over the last years, mobile applications and their corresponding distribution platforms have gained momentum. Applications stores allow users to write reviews and ratings about the apps, giving feedback to developers. User ratings and reviews may help to improve software quality, solve bugs and develop new features. However, this data is hard to be handled by an individual due to the ever growing amount of textual reviews. This paper proposes the use of cognitive computing technologies for addressing this challenge, by developing a smart agent able to mine bugs reports, feature suggestions and sentiment expressed in mobile app reviews. The main contributions of this paper are: the design of a cognitive agent for assisting developers in managing their interaction with their users, the application of machine learning algorithms for bug and feature request detection, and the agent implementation in a real scenario.**

***Index Terms*—cognitive agent, review, app, bugs, suggestions, sentiment analysis**

I. Introduction

Nowadays, app development for the smart phone ecosystem is quite faster and easier than ever [1]. At the time of writing this paper, Google Play Store offers over 3.000.000 mobile apps, mostly developed by third-party companies, organizations and individual developers [2]. For mobile product developers, to maintain their apps in the top of the store rankings results crucial, and this is only possible by implementing new features and solving bugs regularly [3].

However, iOS and Android market platforms only provide customer feedback with an average rating of 1 up to 5, and sometimes including a short review with the user experience. In most cases, this information is not enough to identify the reason of a bad acceptance for an app, and consequently a bad positioning in the store [4]. This lack of data provided by marketplaces about how users feel using an app results in a problem for developers [5]. Furthermore, for translating bad ratings into good ones, companies should interact with their users writing a well-formed response [6]. This response must be personalized for better results, and this becomes a problem for high number of reviews [7].

ORCID info:
S. Muñoz: http://orcid.org/0000-0002-2070-8976
O. Araque: https://orcid.org/0000-0003-3224-0001
C. A. Iglesias: https://orcid.org/0000-0002-1755-2712

As the number of reviews increases, reading them and getting insight becomes a bigger challenge. This has boosted the interest of researchers in applying mining techniques to application stores reviews, in order to extract and analyze the expressed user opinions and sentiments [8].

Mining apps reviews and performing a sentiment or emotion analysis enables the obtention of more detailed information about the app. This information is useful for identifying how users feel. In addition, comparing reviews from a given user in similar apps enables the extraction of information about users' preferences in related apps, allowing developers to make recommendations based on this data.

This paper proposes the use of cognitive computing [9] to address the above challenges. The main contribution of the paper is the design and development of a cognitive agent for Android devices able to: (i) perform sentiment and emotion analysis for Play Store[1] apps; (ii) interact with the user through voice or text sentences, which are converted into specific actions using Natural Language Understanding; (iii) write automatic custom replies to Play Store reviews; (iv) identify if users mention bugs or propose feature requests; and (v) manage and analyze the status of the market of mobile applications in a given domain. In addition, this paper also describes the implementation and evaluation of the developed agent in a real scenario.

The rest of the paper is organized as follows. Firstly, and overview about cognitive computing is given in Sect. II. Sect. III describes the architecture of the developed system, describing the main components and modules; and the features and bug classification is explained in Sect. IV. The implementation of the system in a real scenario and its evaluation is described in Sect. V and, finally, the drawn conclusions are presented in Sect. VI.

II. Background

A. Cognitive computing

Cognitive computing *refers to smart systems that learn at scale, reason with purpose, and interact with humans and*

[1] https://play.google.com/store/

978-1-5386-7793-3/18/$31.00 ©2018 IEEE
DOI 10.1109/Innovate-Data.2018.00010

IEEE computer society

other smart systems naturally [9]. Cognitive systems are able to learn from incoming data and from their interactions with humans, opening new possibilities to produce better products taking advantage of the combination of computers' analytic capability and encyclopedic knowledge and humans' creativity and expertise [10].

Over the last few years, the research and commercial interest in cognitive computing has considerably grown [11]. The use of Natural Language Interface (NLI) [12] has arrived to a number of commercial products based on these technologies, such as Amazon's Alexa[2], Google's DialogFlow[3], Microsoft's Luis[4], IBM's Watson[5], Facebook's Wit[6] and Apple's SiriKit[7].

These systems rely on two concepts for performing Natural Language Understanding (NLU) operations: intent and entity. An intent represents a mapping between what a user says and what action should be taken by the agent. An entity, instead, is a tool for extracting parameter values from natural language inputs [13].

B. Dialog Flow

DialogFlow, previously known as api.ai, is a NLU cloud platform owned and maintained by Google. It is a free to use conversational platform that supports various languages, different programming languages, and has a series of built-in integration with other chatbot-based platforms (e.g., Telegram, Google Assistant, Amazon Alexa) [13].

There are four key concepts involved in any DialogFlow implementation: Agents, Entities, Intents and Contexts. **Agents** can be described as NLU modules for applications. Their purpose is to transform natural user language into actionable data. This transformation occurs when a user input matches one of the intents or domains. **Entities** represent concepts and serve as powerful tool for extracting parameter values from natural language inputs. The entities that are used in a particular agent will depend on the parameter values that are expected to be returned as a result of agent functioning. In other words, a developer does not need to create entities for every concept mentioned in the agent, being necessary only for those that require actionable data. **Intents** represent a mapping between what user says and what action should be taken by your software. An intent is composed by several modules. First of all, what user says in natural language is required. Then it is necessary to set up the corresponding action, and the response, which is provided by the external application service. Finally, **contexts** are designed for passing on information from previous conversations or external sources, such as user profile or device information. Also, they can be used to manage conversation flow.

In addition, DialogFlow provides machine learning capabilities, a tool that allows agents to understand user inputs in natural language and convert them into structured data, extracting relevant parameters. In the DialogFlow terminology, the developed agent uses machine learning algorithms to match user requests to specific intents and uses entities to extract relevant data from them. The agent *learns* both from the data is provided in it and from the language models developed by DialogFlow. Based on this data, it builds a model for making decisions on which intent should be triggered by a user input and what data needs to be extracted. The model is unique per agent.

[2]Amazon's Alexa (https://developer.amazon.com/alexa)
[3]Google's DialogFlow (https://developers.google.com/actions/dialogflow)
[4]Microsoft's Luis (https://www.luis.ai)
[5]IBM's Watson (https://www.ibm.com/watson)
[6]Facebook's Wit (https://wit.ai)
[7]Apple's SiriKit (https://developer.apple.com/sirikit)

III. Architecture

The proposed architecture can be divided into three groups: the server side; the DialogFlow agent; and, lastly the mobile smart agent. The complete global architecture of the system is shown in Fig. 1.

A. Server

Server side section is composed by several modules. At first, we have the controller class from where every request made to the server is handled. This controller adapts the input parameters to the final components which will carry out single functionalities.

All the requests are received through the API webhook, that is linked to the DialogFlow agent. An API REST has been defined and implemented in Flask. It provides methods for extending easily the functionality of cognitive agent. This API acts as a controller that redirects to the corresponding module depending on the application workflow, as shown in Table I.

The controller interacts with three submodules. The Play Store module is formed by all the procedures required for app info extraction from Google Play website. Scrapping and filter tasks are performed inside this component. Afterwards, the Senpy module is responsible for connecting with the sentiments and emotions analysis. In this class we can find some functions to adapt and process the response received from Senpy service, an online sentiment and emotion analysis service [14] described in [14]. Finally, the last module is composed by bug and features classifiers which are previously trained. Each of these components interacts with external applications, like Google Play, Senpy or Slack by extracting relevant information from their website, or using other API REST services.

Some of the information obtained inside these modules is saved in a database, such as recent analysis carried out or the trained classifiers pickles, in order to cache most requested information and speed up the communication process between the smart agent and the server.

The DialogFlow module acts as an intermediary between the logic hosted by the Google server and the smart mobile agent. All this behaviour is orchestrated by an Agent remotely configured using the DialogFlow web interface. This component is responsible for interpreting the request made by the smartphone, translating the voice clip into a text sentence, identifying the meaning of the query and extracting

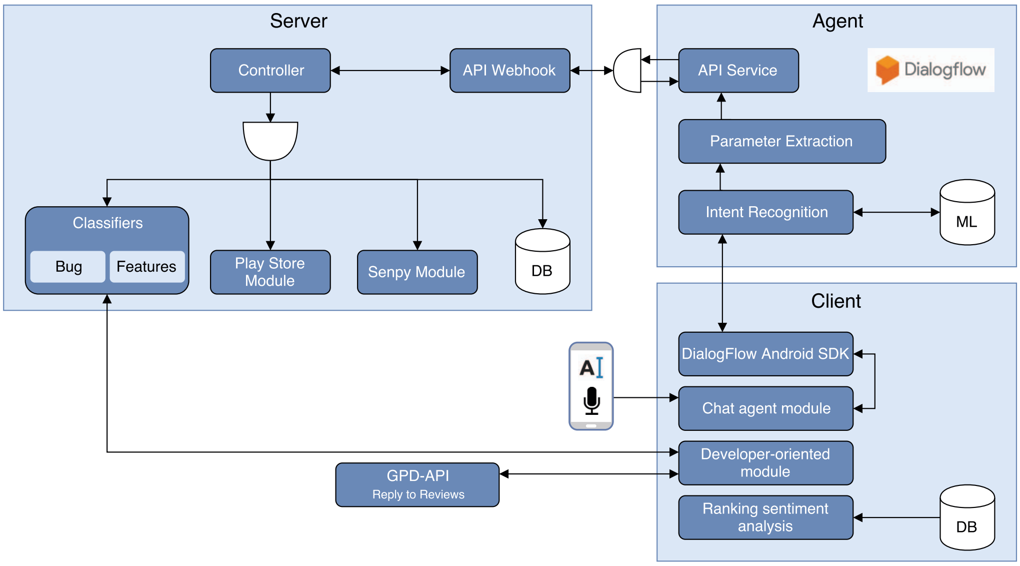


Fig. 1: Global architecture of the system developed

Method	Route	Params	Description
GET	/getAppInfo	appName	Retrieve complete information about the app requested. This information is scraped from the Play Store website and inserted in a JSON file.
POST	/analyze	appName analysisType maxReviews	Perform an **analysis** for specified application, obtaining the information form Play Store if necessary. Analysis types are sentiments and emotions.
POST	/classify	appName classificationType maxReviews	Perform a **classification** for specified application, obtaining the information form Play Store if necessary. Classification types are bugs and features.
GET	/checkQueue	taskId	Check a task status inside the server queue in case it requires a long execution time and the DialogFlow agent throws a request timeout exception.

TABLE I: API REST

the relevant parameters with pattern recognition and carry out the consequent action. This action might be as simple as replying with a simple text string, or more complex and require to communicate with the API webhook to use other modules (e.g. detect users' reviews mentioning a bug).

B. DialogFlow agent

Then, we are going to introduce the structure of the DialogFlow agent, explaining the intents and events implemented to carry out the client requests. This aids to understand the basic model of the agent and its architecture.

Intents refer to actions that our agent will execute. These intents could have dependencies from previous interactions with the agent, so it's necessary to contextualize each one of them. The intents handled by the agent are:

- **Analyze**: Reflects the analysis action. The agent receives the analysis type and the application desired to be analyzed as input parameters, offering the Senpy [14] analysis result as output. This intent uses the webhook option enabled because it needs to interact with the developed server.
- **Classify**: it represents the bug or feature classifier call, being necessary to communicate with the remote server. The input parameters are the classification type and the application in case it can't be extracted from the conversational context.

Every intent is prepared for being summoned with missing parameters, training the DialogFlow agent to request those that are mandatory, or package them inside the conversation context, from where the system can extract them. For example, if the *classify* intent is called like *Would you kindly perform a sentiments analysis for WhatsApp application*, the system collects the app name parameter and the analysis type for a while, in case future requests have these values empty. Afterwards, if we ask the agent *Could you perform a bug classifier*, The bot will extract the missing app name parameter from the context, referring to the last app used in the conversational process.

This humanizes the natural language understanding and accelerates the conversation, interacting with the bot through real natural sentences instead of sending simple commands or instructions to a robot.

Entities refer to the element which represents concepts involved in the intent event triggering. The entities defined for the agent are:

- **App**: refers to the application that wants to be processed, the name of any application available in the Google Play marketplace. This entity remains at session context since it can be different for every user session.
- **Analysis Type**: refers to the analysis types that can be performed. The value is bounded, being possible to run a sentiments or emotions analysis exclusively. The scalability offered by our architecture based on the defined API enables us to add more analysis types in the future.
- **Classifier Type**: specify which type of trained classifier would you like to call. The available values for this attribute are also limited to the detection of mentions about bugs or feature requests.

The agent has also enabled the *Small Talk* bundle, that includes predefined phrases to the most popular requests, what makes the bot to look more like a human [10].

C. Client

The client module is developed through an Android application that receives user requests and redirects them directly to the DialogFlow agent. This behaviour is implemented through the DialogFlow extension library, which provides us all the needed methods to talk with the DialogFlow service.

The application is structured in several modules.The main module manages the voice and text inputs from users in a chat bot interface, and is linked to DialogFlow, representing the response obtained from the DalogFlow server. The second module provides support to developers, and allows them to define automatic responses for the applications they have published in the play Store. Finally, there is an analysis module that review the top free apps in the Play Store marketplace This analysis shows an evolution of the app sentiment over time. This provides users insight about which product trends for succeeding in the app market

IV. Feature and Bug Classifier

The idea for the proposed architecture is to train a model that can then be used for prediction. In this way, the server controller can insert review inputs and extract the classification result without the need of creating, training and testing the model at each iteration. This enables the interaction with the classification module in a fast and efficient manner.

In order to perform the feature and bug classification, four machine learning algorithms have been used, trained for binary prediction: Logistic Regression [15], Naive Bayes [16], and both Linear and Gaussian SVM [17], [18]. This selection is done in order to compare the performance of these frequently used models.

The implemented classifiers have been trained and tested with a mobile application review dataset [19], obtained from [20]. This dataset is composed by reviews previously tagged as *bug* or *not_bug*, and *feature* or *not_feature*. Said data collection offers a supervised dataset composed by multiple reviews posted inside the iOS app store. However, this data has been extrapolated, using them for Google Play marketplace context.

Processed data is formed by 3,117 instances corresponding to bug reviews, and 1,924 reviews related to feature requests. Due to original data distributions, the labelling is not balanced in both bugs and features categories. Table II shows the dataset statistics, including the percentage of positive and negative classes.

Category	Size	#instances	%
Bug	3,117	2,740	54%
Not Bug		377	7%
Feature request	1,914	1,619	32%
Not Feature request		295	5%

TABLE II: Dataset sizes for supervised learning.

Information included in this dataset can be divided in two types, of which we use in the experiment the following. Firstly, the text of the review, which includes a preprocessed version with stopwords removed and lemmatization, used verbal tenses -present, past and future-, and number of words. Secondly, review-oriented metadata; the rating provided by the user. Consequently, we feed the learning models with Bag-of-Words features from the preprocessed text, and a normalized representation of the rest of features.

The training and testing methodology is based in a 10-fold cross validation using the aforementioned dataset. This strategy has also been used to tune model hyper-parameters. Bug and feature requests are treated separately, since as stated in previous work [20], our experimental results validate that this separation greatly enhances the final performance. Besides, in order to gain insight into each model, we use precision, recall and micro-averaged f-score as metrics.

Experiment results are summarized in Table III. It can be seen that, although there is no model that yields the best performance on all metric and the two categories, gaussian SVM does not reach as good numbers as the rest of the

Algorithm	Bugs			Feature requests		
	Precision	**Recall**	**F1 micro**	**Precision**	**Recall**	**F1 micro**
Logistic Regression	**95.06**	92.01	89.22	**93.91**	83.61	81.59
Naive Bayes	93.65	**95.52**	90.19	79.71	86.08	78.71
Linear SVM	94.28	95.00	**90.70**	91.51	**88.68**	**83.43**
Gaussian SVM	94.23	94.89	90.60	93.76	83.74	81.54

TABLE III: Performance metrics for the four learning models. In bold, the best value for each metric in each data category.

classifiers. This could be explained attending to the fact that gaussian SVM is the most powerful model, which can also lead to overfitting when training will a small number of data samples, as it is the case for this experiment (Table II). Also, results seem to indicate that linear SVM yields the better performance, if attending to all metrics.

In order to gain further insight into the classification process, we have extracted most relevant words from both categories as computed by the model. When considering bugs, those are: *annoy*, *broken*, *hopefully*, *responsive*, *implementation*. For feature requests, the model considers *ask*, *reinstall*, *game*, *closes*, *rethink*.

Given that linear SVM has the higher F1 micro in both categories, we finally select this algorithm to be implemented in the developed system architecture. As commented above, this model is deployed in the system through *pickle* serialization.

In relation to scalability, the developed system has been deployed in a 64-CPU high-availability cluster that ensures the system is able to exploit parallelization. The bottleneck of the system is the machine learning processing. Also, the implementation is done using scikit-learn [21], which is a Python library that enables multiprocessing and includes several complexity and computational performance optimizations, as seen in the official documentation [8].

V. CASE STUDY

The proposed scenario consists in a ***start-up*** whose business activity focuses on developing mobile applications for third companies, or by their own initiative. This new company is trying to explore the mobile application market, so they decided to use the smart agent designed in this work so they can study using sentiment analysis techniques how users feel about their own applications, and similar apps that directly compete with their products.

As discussed previously, it's desirable to process automatically the feedback obtained from their users and extract valuable data from it, such as bugs and crashes encountered, or features and improvements that could increase their audience opinion. Moreover, it will be great to redirect these results to a team management platforms like *Trello* or *Slack*, so the developer team could stack those reviews and solve it as soon as they can.

In the following, three different scenarios are presented for showing how the smart agent can be applied, and including some screen captures with the result obtained inside the Android app.

A. Market Explorer

The *start-up* is looking for a project management application in order to improve their inter-department communications and track their external project status from their smartphones. Therefore, the CIO has decided to explore the market looking for applications that are able to solve this upcoming idea. For this, he will perform a sentiment and emotion analysis for popular team management apps, more concretely he will focus on *Trello*, *Slack*, *Basecamp*, *Todoist* and *Evernote* apps. The main goal is to observe the sentiment and emotion on these apps and choose one to incorporate it within the activity of the company. To carry out this procedure, he will use the chat smart agent. In concrete, the CIO will perform a sentiment and emotion analysis for each published app and later compare them depending on the results.

Using the chat module, the user can insert the queries through text or voice input format. Figure 2 shows an example of the conversation the CIO has inside the proposed scenario. The agent accepts multiple input format for queries, so it could exist multiple possible dialog flows to obtain the same analysis output. The user interface given by the agent is really simple, acting as a task assistant answering any user input thanks to the DialogFlow platform. After the analysis process, final results are grouped in Table IV. The application enables to examine more in detail the results obtained by clicking the card, being able to extract even the sentiment or emotion of a single review, as shown in Figure 3.

To conclude, observing the results obtained from both analysis the CIO has to determine which third-party platform is better considered by the Play Store user community. Most of the applications have present the same emotion, so generally are good applications. On the other hand, the sentiment varies depending on the app, probably due to a recent updates that users don't like at all.

B. Feature mining

This section describes a scenario where the *start-up* receives a project plan offer from a well-recognized digital newspaper, which requests a second version of their Android mobile app named *The Guardian*. Due to the low budget the company had to face the project in the first version, they decided to implement only basic functionalities to reach a higher audience developing a simple application for smart-phones.

[8] http://scikit-learn.org/stable/modules/computational_performance.html

Application	Reviews	Sentiment	Emotion
Trello	15	😊	😊
Slack	15	🙁	😊
Basecamp	15	😐	😊
Todoist	15	😊	😊
Evernote	15	🙁	😊

TABLE IV: Results of sentiments and emotions analysis in project management applications

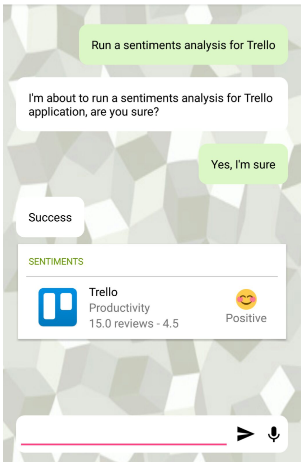


Fig. 2: Sentiment analysis for *Trello*

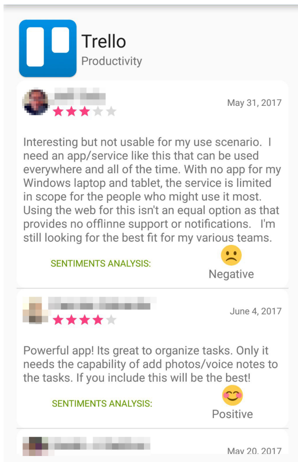


Fig. 3: Detailed sentiment analysis view

Now, the company wants to order a second version with advanced functionalities to our scenario company, and the *start-up* developer team members are not able to identify interesting improvements or features for the application. For this reason, they decided to *obtain* them from the current published version, in order to implement those mostly demanded in the Play Store reviews section.

To carry out this operation, they use the proposed smart agent, performing a feature classification over the *The Guardian* application, so they can detect new features and improvements without having to read every single comment and extract interesting values. So a feature classification is run for the app. The feature analysis is executed for a total of 30 reviews, considering that this number is enough to extract a significant amount of relevant features. The result is represented in Fig. 4

Some of the reviews tagged as features can be checked in Table V. It shows the classification result and the value obtained from a natural language comprehension, being able to detect if the classification output matches with a human understands after read that user review.

As shown, most of the improvements extracted match with the classification result. After this process, the developer team meet together and brainstorm about the most common features

Review	Classification Result	Value
...by default, it does tend to send quite lot of notifications...	Feature	Customize notification subscription system
... however, your news source doesn't update like CNN, you stay on the same old news headline...	Feature	Update more frequently the news title, in case an important new come up suddenly.
Very good, lots of interesting articles, good podcasts, easy to read posts and videos...	Not Feature	None

TABLE V: Results

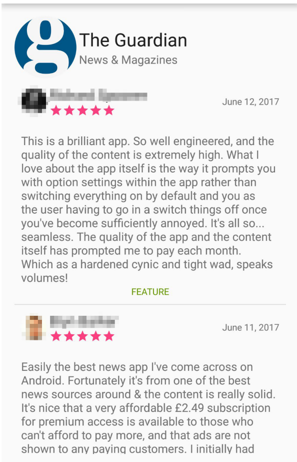


Fig. 4: Features classification details

demanded by the users, prioritizing those that seems to be more interesting for the client. Without usage, the developer team would have had to fetch every single review, filter those that refer to bugs, user experience, opinion, etc.; and finally evaluate their importance in the second version.

C. Automated reply for bugs detection

In this scenario, the *start-up* has just published the second version for the digital newspaper introduced in Sect. V-B, and the next step is to carry out a maintenance process during the first month since launch. The objective is to analyze the user experience over the app, extracting relevant comments from app reviews and trying to solve the errors detected as soon as possible. It seems really tedious to read each review and manually filter those that refers to bugs. For this reason, the development team think that the developer-oriented section included in the proposed smart agent could speed up this task.

The main objective is to collect the application feedback posted in Play Store, to classify it using the bug classifier, and finally to answer accordingly to that bug.

The developer accesses to the developer-oriented tab and visualizes which is represented in Fig. 5. We suppose that the agent has previously been linked to the Google Play developer company account. The input data must be the application name and the version that is currently published in production. The package name is also mandatory in order to identify the application.

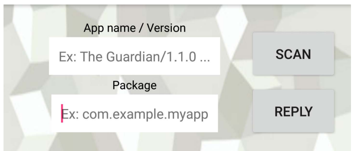


Fig. 5: Input parameters requested for developer-oriented scanning.

After pressing the *SCAN* button, the smart agent communicates with the *Reply to Reviews API*, and in case the app belongs to the authenticated account, it will obtain all the review comments posted in last 15 days. After obtaining them, it is necessary to evaluate them calling the remote bug classifier. If any review is tagged as a *bug*, the system will randomly extract a friendly response from a bug collection replies. The result is shown in Fig. 6. As can be seen, the user is complaining about a crash that happens after the app log in process, so it is clearly a bug. The system has extracted the following recommended answer: *We apologize for inconvenience, we will try to solve it in the next version. Thanks for your patience*.

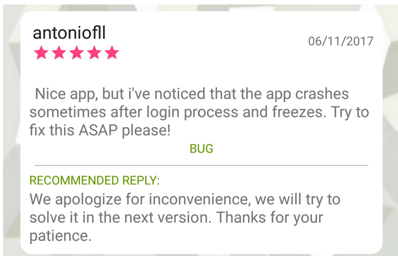


Fig. 6: Result obtained after retrieving recent app reviews, classify them as a *bug* or not, and finally recommend a suitable reply.

When the *REPLY* button is pressed, all the recommended reviews for reviews classified as bugs will be posted to the Play Store endpoint. The reply will also be sent to the author as an email, but that process acts regardless of the agent. Anyway, the publishing process can take a while until it appears in the website. Afterwards, the output will looks like the Fig. 7.

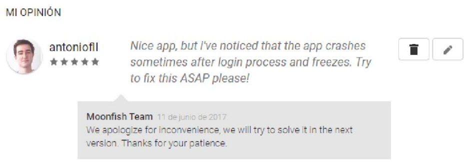


Fig. 7: Automated reply visualization inside Play Store website.

VI. Conclusions

In this paper, a smart agent for Android devices based on DialogFlow has been developed. The agent offers a great possibility to evaluate and analyze the feedback obtained from the Play Store market place using sentiment analysis techniques and binary classifiers, revealing interesting data about user experience. It also offers an innovative design for smart-phone apps, with an intuitive interface that handles voice commands to interact with all system components.

The proposed architecture follows a modular approach based on a REST interface of a controller module for extending the agent capabilities, and exploit the benefits of cognitive computing technologies for integrating natural language conversations.

The designed system can be used by freelance developers in order to analyze the application market status and track the feedback obtained, being able to extract a real meaning from user opinions and redirect those feedback directly to improvements or error detection tasks. This kind of tool isn't common nowadays, and earn so much time and money to newly created company that has limited budget to manually analyze their user feedback.

The high scalability offered by the developed systems raises a lot of possible improvements or future work to be done. One of these possible lines is to develop new classification methods based on the average rating or the user experience detection in order to extract even more valuable information from app reviews. In addition, it would be also interesting to obtain latest application downloaded by the user and develop a recommendation system based on application of the same category place in a high position inside the ranking.

Acknowledgements

This research work is supported by the Spanish Ministry of Economy and Competitiveness under the R&D projects SEMOLA (TEC2015-68284-R) and ITEA3 Emo-Spaces (ITEA3-14012/RTC-2016-5053-7) and for the project ITEA3 SoMeDi (ITEA3-15011).

References

[1] V. N. Inukollu, D. D. Keshamoni, T. Kang, and M. Inukollu, "Factors influencing quality of mobile apps: Role of mobile app development life cycle," *CoRR*, vol. abs/1410.4537, 2014.
[2] "Number of available android applications - appbrain." [Online]. Available: http://www.appbrain.com/stats/number-of-android-apps
[3] P.-L. Yin, J. Davis, and Y. Muzyrua, "Entrepreneurial innovation: Killer apps in the iphone ecosystem," Stanford Institute for Economic Policy Research, Discussion Papers 13-028, 2014.
[4] I. J. Mojica, M. Nagappan, B. Adams, T. Berger, S. Dienst, and A. E. Hassan, "An examination of the current rating system used in mobile app stores," *IEEE Software*, 2015, preprint.
[5] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh, "Why people hate your app: Making sense of user feedback in a mobile app store," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '13. New York, NY, USA: ACM, 2013, pp. 1276–1284.
[6] "Best practices for replying to ios & google play reviews." [Online]. Available: https://support.appbot.co/help-docs/best-practices-replying-to-ios-and-google-play-reviews/
[7] G. Kwakyi, "Why you need to reply to your app's google play user reviews," Oct 2016. [Online]. Available: https://medium.com/@incipiagabe/why-you-need-to-reply-to-your-apps-google-play-user-reviews-88226e617f0b
[8] N. Genc-Nayebi and A. Abran, "A systematic literature review: Opinion mining studies from mobile app store user reviews," *Journal of Systems and Software*, vol. 125, pp. 207 – 219, 2017.
[9] H. Demirkan, S. Earley, and R. R. Harmon, "Cognitive computing," *IT Professional*, vol. 19, no. 4, pp. 16–20, 2017.
[10] M. Coronado, C. A. Iglesias, Á. Carrera Barroso, and A. Mardomingo Mardomingo, "A Cognitive Assistant for learning Java featuring Social Dialogue," *International Journal of Human-Computer Studies*, Oct 2018.
[11] A. Soffer, D. Konopnicki, and H. Roitman, "When watson went to work: Leveraging cognitive computing in the real world," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2016, pp. 455–456.
[12] M. a. Hearst, "'Natural' search user interfaces," *Communications of the ACM*, vol. 54, no. 11, p. 60, 2011.
[13] M. Canonico and L. De Russis, "A comparison and critique of natural language understanding tools," in *CLOUD COMPUTING 2018: The Ninth International Conference on Cloud Computing, GRIDs, and Virtualization*. IARIA, 2018, pp. 110–115.
[14] J. F. Sánchez-Rada, C. A. Iglesias, I. Corcuera-Platas, and O. Araque, "Senpy: A Pragmatic Linked Sentiment Analysis Framework," in *Proceedings DSAA 2016 Special Track on Emotion and Sentiment in Intelligent Systems and Big Social Data Analysis (SentISData)*, October 2016.
[15] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *Journal of machine learning research*, vol. 9, no. Aug, pp. 1871–1874, 2008.
[16] K. P. Murphy, "Naive bayes classifiers," https://datajobsboard.com/wp-content/uploads/2017/01/Naive-Bayes-Kevin-Murphy.pdf, Oct 2006.
[17] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
[18] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, no. Aug, pp. 975–1005, 2004.
[19] W. M. C. Stanik, D. Pagano, ""Review 2015 Training Set - SQL format"," https://mobis.informatik.uni-hamburg.de/wp-content/uploads/2015/03/Data-and-Coding-Guide-.zip, 2015.
[20] W. Maalej and H. Nabil, "Bug report, feature request, or simply praise? on automatically classifying app reviews," in *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, Aug 2015, pp. 116–125.
[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.