Article:

Muñoz, S. orcid.org/0000-0002-2070-8976, Iglesias, CÁ orcid.org/0000-0002-1755-2712, Mayora, O. orcid.org/0000-0002-5773-3876 et al. (1 more author) (2022) Prediction of stress levels in the workplace using surrounding stress. Information Processing & Management, 59 (6). 103064. ISSN 0306-4573

https://doi.org/10.1016/j.ipm.2022.103064

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
0306-4573/

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

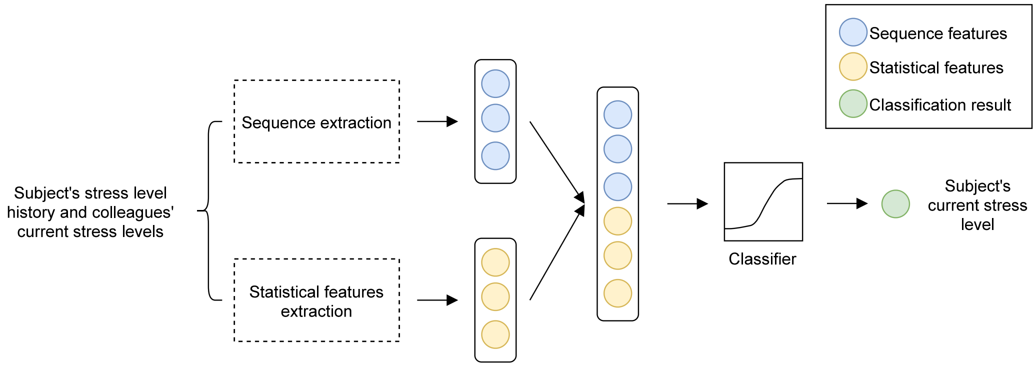

Fig. 1. General architecture representation of the proposed method.

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	$E_{std} = \sqrt{\frac{1}{w-1} \sum_{i=t}^{t-w} (e_i - \bar{e})^2}$	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 (Gooday & 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.