International Information and Engineering Technology Association
*Advancing the World of Information and Engineering*

# Classification of Human Mental Stress Levels Using a Deep Learning Approach on the K-EmoCon Multimodal Dataset

Praveenkumar Shermadurai*[ID], Karthick Thiyagarajan[ID]

Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur 603203, India

Corresponding Author Email: praveens11@srmist.edu.in

**ABSTRACT**

The idea of "stress stacking" is that the psychological stress many people face in modern society can lead to depression, heart disease, and cancer, among other long-term illnesses. Thus, managing and tracking a person's stress is crucial. This paper proposes that a modified Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model can extract features from Electroencephalogram (EEG), Electrocardiogram (ECG), and Accelerometer (ACC) data. A lengthy feature vector combines relevant features from the various modalities. Combining all of the features could improve classification performance, but it could also increase the number of dimensions and lead to bad performance because of redundant information and inefficiency. This paper uses Kruskal-Wallis analysis to suggest a new way to deal with the high-dimensionality problem by automatically finding the best subset of features. To categorize the stress based on the feature vector, we utilized a K-Nearest Neighborhood (KNN), a Random Forest (RF), a Support Vector Machine (SVM), and a Decision Tree Classifier (DT). SVM outperformed the other three classifiers with a performance accuracy of 94.58%, which is 3.72% Superior to cutting-edge techniques.

## 1. INTRODUCTION

The nervous system's response to a threat or directive is called stress [1]. Stress has received substantial attention recently because it affects so many people. This propensity might result from shifting work practices, cultural expectations, lifestyles, etc. [2]. In some situations, such as high-stress ones at work, during exams, and the like, tension can be advantageous up to a point. Once it reaches a certain point, stress is no longer suitable. Additionally, it harms one's productivity, the value of life, health, and mental condition [3]. If a person experiences recurring trials and feels intense anxiety, their body will be under prolonged strain [4]. Consequently, the importance of stress detection systems has grown in comparison to ten years ago. It is imperative to safeguard people against the escalating impacts of stress, particularly since it cannot be prevented. Therefore, prompt identification and management of stress are essential for enhancing mental health and general well-being [5].

Psychological, physiological, and behavioral primary modalities are used most frequently in automatic stress detection [6]. The Hypothalamic Pituitary Adrenal axis and the Autonomic Nervous System are the two primary processes that work to try to restore physiologic stability in response to stress [7]. Variations in heart rate variability, sweat gland activities, skin temperature, etc., bring this on. Thus, physiological signals can offer details on ANS function as reliable stress markers. Additionally, ECG and EDA, among physiological features, provide an accurate picture of a person's state of stress [8].

Several psychological instruments commonly used in research and clinical settings to assess stress levels include the Perceived Stress Scale [9], the Stress Response Inventory [10], the Holmes-Rahe Stress Inventory [11], and the Hamilton Rating Scale for Depression [12]. The evaluations include self-reports or professional ratings based on arbitrary judgments and estimations to get exact data on cognitive, emotional, or behavioral stress reactions. Nevertheless, due to their subjective nature, these methods are deficient in their ability to discern nuanced patterns of mental state. In machine learning, feature selection, particularly infinite latent feature selection, is crucial (ILFS). The method uses all possible feature subsets to produce ranking [13, 14].

When an employee's capacity to complete a task falls short of the assignment's difficulty, work-related stress develops [15], which harms people's health and society. Persistent mental stress can lead to psychological and physical ailments like headaches, depression, anxiety, and sleeplessness [16-20], raising healthcare expenditures and lowering the quality of life. These conditions can also lead to cancer, diabetes, hypertension, cardiovascular diseases, and more. Chronic stress may also reduce performance, increase absenteeism, and ultimately result in job loss, which raises expenses [21, 22]. Stress-oriented costs were calculated to be $30 billion and $100 billion in nations like the US and Australia. Acute stress left untreated for a prolonged period can lead to chronic stress, which can have detrimental long-term effects. As a result, it's crucial to keep a close eye on employees' mental health.

According to recent studies, physical and mental health are intimately intertwined [23-25]. As a result, more people are turning to technology to identify people's mental states. Stress and pressure are significant and integral parts of daily living for humans. Global stress is a concern. One in three people [26] has severe anxiety or stress that manifests in psychological and physical symptoms. Numerous stressful circumstances, such as long-term higher stress levels on socioeconomic characteristics and interpersonal interaction problems, excite the human autonomous nervous system. A rise in heart rate, perspiration, and muscular contraction is brought on by releasing chemicals like adrenaline or cortisol during this period [27]. These reactions enhance one's ability to react correctly under pressure. Stress may prevent fatalities in an emergency by, for example, prompting someone to use the brakes to avoid a car collision. Human health is jeopardized when the body fails to revert to its normal state, and the mind as well as the body endure persistent and prolonged stress.

The autonomic nervous system did poorly in distinguishing between sensitive and physical risks. For instance, the body responds to a disagreement with a friend like in a real-life-or-death situation. Physical pain, mental sickness, digestive problems, sadness, and reduced productivity may all be brought on by stress. According to the data from the American Psychological Association, work-related stress has a financial impact of $200 billion each year on the US economy [28]. The probability of accidents is also increased by drivers' tension and exhaustion [29]. Stress should be identified and addressed as soon as it manifests to avoid chronic stress and its ensuing issues and costs. Wearable sensors have shown the ability to assess mental stress and anxiety levels [30]. Stress diagnosis has become more accurate when using sensor fusion techniques, which collect data from multiple sensors. Too far, there needs to be a complete, widely-accepted standard for how to combine different signal kinds. In this study, we create a brand-new stress detection system that uses machine learning algorithms and a sensor board.

Because they provide researchers and inventors with the resources they need to gather and interpret data, sensor-based knowledge is crucial for e-Health. According to conventional definitions, a wireless sensor network (WSN) may connect with nearby devices to exchange or collect data, interact with the atmosphere through its sensors, analyze data nearby, or link with other wireless communication technologies [31]. Since the vast majority of contemporary devices used in people's everyday lives come with various sensors/sensor-oriented applications, they have been broadly utilized in healthcare and are only continuing to gain popularity. WSNs, ML, or DL influence how advanced healthcare apps are developed [32]. Modern e-Health applications may considerably benefit from the burgeoning fields of ML and DL since they need sophisticated techniques for exploration and analysis.

By analysing the patterns and variations in physiological signals such as EEG (Electroencephalography), ACC (Accelerometry), and ECG (Electrocardiography), mental stress may be categorised into three degrees. Each signal can contribute to the categorization of stress levels in the following manner:

EEG is a technique employed to identify and examine the electrical impulses in the brain. This enables us to comprehend the distinct brainwave patterns associated with different mental states. Specific cognitive and emotional states are linked to distinct frequency bands in EEG data, including alpha, beta, theta, and delta waves. During periods of mental stress, there may be alterations in EEG signals, characterised by heightened beta activity and reduced alpha activity. Researchers can detect different degrees of mental stress by examining the power spectral density or coherence of EEG data across various frequency bands.

ACC is commonly used wearable devices that monitor the acceleration and movement patterns of the body. Stress can lead to alterations in movement patterns, such as heightened physical activity or posture modifications. Researchers can deduce the degree of physiological arousal and agitation linked to mental stress by examining factors like activity level, gait dynamics, and postural alterations obtained from ACC signals. Elevated levels of physical activity or agitated movements may indicate heightened stress levels.

ECG is a technique employed to evaluate the electrical signals produced by the heart, enabling the measurement of heart rate, heart rate variability (HRV), and other pertinent cardiac parameters. During periods of psychological stress, the body's autonomic nervous system may exhibit alterations in heart rate and heart rate variability. Heightened sympathetic nervous system activity can result in a raised heart rate and decreased heart rate variability, whereas parasympathetic activation can have the opposite impact. Examining characteristics such as heart rate variability, heart rate dynamics, and heart rate recovery from ECG data might aid in classifying degrees of stress. For instance, a reduction in HRV and an extended period of recovery following the removal of a stressor may suggest elevated levels of stress. Researchers may employ multimodal techniques to correctly measure an individual's stress response by integrating data from EEG, ACC, and ECG signals and analysing their temporal dynamics. This enables the categorization of mental stress into three degrees or more. Machine learning methods, such as classification models or pattern recognition approaches, can assist in accurately detecting significant characteristics and patterns within these signals to efficiently categorise stress levels.

## 1.1 Research contribution

We made numerous contributions to constructing a stress categorization system in this study. One must use an integrative strategy to improve considerably because all these processes are interconnected. Figure 1 displays an overview of the contributions made to this work. Their contributions are presented in this study in the following ways:

**Feature extraction**
Feature extraction automatically identifies pertinent and significant characteristics using unprocessed EEG, ECG, and ACC input data. Reduced dimensions of the data input while retaining the essential information is the aim of feature extraction.

**Feature fusion**
In feature fusion, several features retrieved from various layers or modalities of neural networks are fused to create a single feature representation. This may be done by concatenating the feature representations along the feature dimension during a process known as concatenation.

**Feature selection**
Feature selection is the process of choosing useful features from input features.

**Classification**

In several applications, including machine learning, data analysis, and computer vision, feature categorization may assist in making sense of and organizing incoming data.
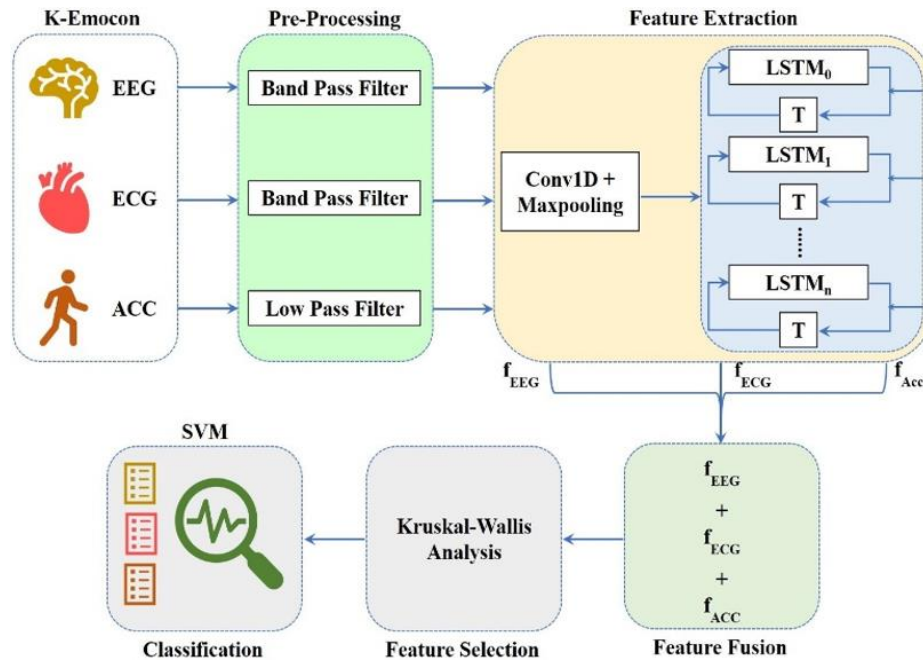


**Figure 1.** Proposed system architecture

## 1.2 Overview

An outline of this study is given in this section. Section 2 presents the Related Work. This is an overview of previous studies that used physiological signs to categorize stress. The novel framework for automatically identifying the key EEG, ACC, and ECG attributes is proposed in Section 3.4. A framework for fusing the information obtained from EEG, ACC, and ECG physiological data is also presented in Section 3.4. The Kruskal-Wallis method's new foundation for choosing the ideal feature set is presented in Section 3.5. Information on the various classifiers used for the classification stress is detailed in Section 3.6.

The specifics of the suggested approach are as follows:

Once pre-processing is complete, pertinent characteristics are derived from the K-EmoCon dataset. Conduct the Kruskal-Wallis test on each feature to ascertain whether there are substantial variations in feature values across different stress levels. The Kruskal-Wallis test is a non-parametric statistical technique employed to compare three or more groups and ascertain the presence of statistically significant disparities among them.

Features are chosen based on their distributions across stress levels, as determined by the Kruskal-Wallis test, if there are significant differences. The SVM classifier is trained using the chosen characteristics. SVM is a supervised learning technique that is specifically developed for completing classification tasks. It operates by identifying the hyperplane that most effectively divides the classes in the feature space. The purpose of cross-validation is to assess the efficacy of the SVM classifier. This aids in evaluating the model's ability to generalise to unfamiliar data and prevents overfitting. The final classifier model based on SVM is evaluated using appropriate metrics such as accuracy, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).

The novelty of this paper is given as follows: By measuring the stress and dividing it into three degrees, we want to create a framework for categorizing mental stress. To extract features, the modified CNN-LSTM model receives the pre-processed EEG, ACC, and ECG signals from the K-EmoCon database. Via feature-level fusion, the retrieved features are combined. The fused features are improved using feature selection and the Kruskal-Wallis statistical analysis. The optimized feature set is then subjected to an SVM-based classifier model employing the poly kernel.

## 2. RELATED WORK

Researchers propose a CNN structure for recognizing activities in a dataset collected from gyroscope and accelerometer sensors. The technique of windowing is employed, utilizing a duration of 2.56 seconds with a 50% overlap. Instead than emphasizing the intricate nature of the network using 1D kernels for convolution and max-pooling in the CNN network, the research investigation centers its attention on multiple feature mappings. The study [33] examines three distinct datasets utilizing Deep Networks (DNNs), CNN, and a modified version of LSTMs. The windowing technique is used with a 50% overlap and 1-second length. Similar to the research [34] uses one-second windowing to test multiple CNN architectures on three different datasets. The report, however, makes no mention of a particular overlap ratio.

The Discrete Fourier Transform (DFT) combines sensory information into a signal and creates an activity picture. Two convolutional layered models have accuracy scores ranging from 97.6 to 99.9%. Lupien et al. [35] displays the sensory input from a second CNN-based model as frequency pictures with a 1s window length and 60% overlap. The holdout validation procedure results in an average F1 score of 0.68. Wearable technology and cell phones are considered for collecting health-related data in research [36]. According to their complexity, activities are further categorized in studies

like [12]. While tasks like cooking, cleaning the home, or dining, which include several basic motions, are considered difficult, standing, sitting, laying down, and other comparable acts are considered easy. Cooper and Cartwright [37] provide a range of approaches for these sorts of tasks.

A CNN model is fed by breaking complex behaviors down into simpler ones. Immediately after the CNN, the outputs are forwarded to a SoftMax unit where an LSTM Network is utilized to differentiate between complex and straightforward operations. Three independent datasets—MHEALTH, PUC-Rio [38], and AREM [39] — evaluate an LSTM model. The suggested model may get an F1 score of 99%, which is much higher. Studies in paper [40] imply an architecture that blends CNN and LSTM. Using the input axis, Srivastava and Salakhutdinov [41] creates a three-dimensional input vector. Combining each vector from a unique sensor produces a stacked sensor image. The human visual system encourages efficient use of the CNN model and Peek networks. The F1-score of the system is calculated with the Leave One Subject Out validation approach, as described by Miah et al. [42]. This approach utilizes a precision metric and incorporates five-fold cross-validation. Zhen et al. [43] suggested a CNN-based inter-learning system for matching text and images. The modalities employed were text and images. Two sub-networks, namely image CNN and text CNN, were developed to establish the cross-modal association between the modalities. These sub-networks share weights at the fully connected layer, which helps in learning the relationship. Liu et al. [44] presented a modality-invariant paradigm for text-image matching. There is no text provided. The proposed approach enhances the uniformity of the distribution of two sets of embeddings by refining a pre-trained CNN image system and a textual RNN network using an auxiliary adversarial loss. Following the implementation of adversarial learning, the distributions of images and text were more similar, resulting in an improvement in retrieval accuracy. Suris et al. [45] presented an inter framework for audio-video retrieval.

Hazarika et al. [46] created representations that are not influenced by different modes of expression, for the purpose of analyzing sentiments across many modes. The Transformer model was utilized to classify text, images, and videos into multiple distinct groupings. The encoder was trained by providing it with text, picture, and video data, resulting in the incorporation of joint modality features. Healey and Picard [47] developed a method for evaluating strain in driving scenarios by utilizing the ECG, EMG, and RESP. The detection technique exhibited exceptional accuracy, above 90%, across the three stress levels. Furthermore, the investigation revealed a robust correlation between the stress level of Heart Rate Variability (HRV) and RESP. Wijsman et al. [48] created generalized estimating equations as a method for evaluating EEG, RESP, and ECG to measure work-related stress.

The approach demonstrated that each signal exhibited a stress correlation and achieved an average accuracy of 84.7% for two stress levels. By utilizing ECG and GSR, Sriramprakash et al. [49] suggested ML models for assessing work-related stress. The scientists concluded that the optimum feature combination was GSR and ECG, and the model's accuracy for two stress levels was 92.75%. Betti et al. [50] developed a wearable device that utilizes ECG, EDA, and EEG to track occupational stress levels. The technology demonstrated an 86% accuracy rate for two stress levels.

Additionally, it was shown that physiological traits were associated with the amount of salivary cortisol, a stress indicator. Elzeiny and Qaraqe [51] spoke on the need to identify mental stress triggers and use early recognition tactics in the workplace. They provided several stress reduction strategies for both the business and its workers. Various methods used in the cited studies compared in Table 1.

**Table 1.** Comparision of various studies

| Reference | ML/DL Algorithms | Result | Demerits |
|-----------|------------------|--------|----------|
| [51] | VGGNet | 69% | Cross-modal retrieval is a very complex process. |
| [52] | ResNet | 88% | Training data with GRL is expensive. |
| [53] | CNN | 72% | The representation of the input data is not accurate. |
| [54] | DL Approach | 82% | Cross modal Transformer is applied in fusing of multimodal data which affect the performance. |
| [55] | CNN | 91% | The size of the dataset was limited. |
| [56] | SVM | 86% | Limited features were extracted for stress classification. |

## 3. METHODS AND METHODOLOGY

### 3.1 K-EmoCon

The multimodal database, K-EmoCon [57], has detailed annotations of continuous emotions during realistic discussions. When examining emotions in social interactions, a new dataset must be used. The dataset contains multimodal measures from 16 sessions of paired arguments on social issues that lasted around 10 minutes each and were captured using technology that is readily accessible on the market. These measures include EEG signals, audiovisual recordings, and peripheral physiological signals. Since it contains emotional annotations from all three potential points of view, the debater, the debate partner, and the audience, which varies from earlier datasets. While viewing the discussion footage every five seconds, raters noted emotional manifestations regarding arousal/valence and 18 other emotion classes. The K-EmoCon, created due to the multi-perspective evaluation of emotions experienced during social interactions, is the first emotion dataset made accessible to the general public. We considered EEG, ECG, and ACC data for our research work.

### 3.2 Pre-processing

The combination of physiological signals with other sounds, such as motion artefacts, baseline wanders, or white noises, can result in distorted recorded waveforms and reduced feature quality. The EEG data has undergone preprocessing steps, which include applying a bandpass filter with a range of 4 Hz to 45 Hz, removing artifacts, and downsampling the original data to a frequency of 128 Hz. Filters with bandstop and bandpass characteristics were utilized to minimize noise from the ACC and ECG signals. The bandwidths of the filters were set between 58 and 62 Hz and 1.5 and 150 Hz respectively [58]. The task was accomplished by employing a low pass filter with a cut-off frequency of 0.5 Hz. The signals were

partitioned into non-overlapping windows of 10 seconds after eliminating the noise. Acute stress can be detected in real-time because of the short-term window. Lastly, a scale between 0 and 1 was applied to the signal values.

### 3.3 Setup for training

The performance of the proposed DNN model was optimized using several hyper-parameters. The network treated the amount of 1D convolutional layers, dropout rate, and modules in the first and second LSTM as hyperparameters. The units in the LSTMs were also considered hyperparameters for processing EEG, ECG, and ACC data. The knowledge proportion and batch size were additional hyperparameters when training with Adam as the optimizer. The epoch was set to 50 after the models were successfully trained. We used Bayesian improvement of the Keras tuner to optimize the hyperparameters for the first 40 trials of each fusion scenario.

A progressive three-step strategy was employed, incorporating an initial phase to expedite the tuning process.

The model was subsequently trained to utilize the updated hyperparameters without premature halting. The evaluation process was then completed using the model with the lowest validation loss. There is no text provided. As the number of epochs increases, the model exhibits a tendency to overfit the training set. The model with the lowest validation loss was selected to create an optimized model with robust generalization abilities, while avoiding overfitting. The calculation of validity failure and modification of hyper-parameters were performed on the validation set. Then, the accuracy, Area Under ROC Curve (AUC), and F1-measure were used for evaluation. AUC and the F1 measure were macro-averaged using a one-to-one technique in the three-level stress categorization.

T-Stochastic Neighbour Embedding (t-SNE) was utilised for feature-level fusion to decrease the dimensionality of high-dimensional features and display them in a two-dimensional environment. The purpose of its usage was to exhibit the output of the final dense layer subsequent to the ReLU activation in the trained model. Clear separation of separate classes in the space signifies successful training of the model. We developed a machine learning baseline model utilising a SVM. Features were derived from EEG, ECG, and ACC data in 10-second intervals. The mean and standard deviation of the supplied data were calculated. Analyse the attributes of each primary frame in the window separately, without merging them. The min-max scaling method was used to standardise each characteristic. The training, hyper-parameter tuning, and validation approaches were implemented using the same approach as the deep learning (DL) method. The coding and statistical analysis were mostly performed using Python (Version 3.6.10). Deep-learning libraries used include Tensorflow (Version 2.1), Keras tuner (Version 1.0.1), and Keras (Version 2.3.1).

### 3.4 Feature extraction and fusion scenarios

We utilised the enhanced framework from a previous study [59] to develop a deep neural network that evaluates physiological inputs. The system utilised convolutional layers for extracting local information and LSTMs for obtaining sequential information. We used the same filter and pooling size for ECG as indicated in paper [60], but EEG used a filter

size of 15 and a pooling size of 20, even though their network architecture was similar. The characteristics were sequentially inputted into bidirectional LSTMs. In the last LSTMs, an attention unit with the same components as the preceding LSTM was incorporated, as seen in Figure 2. Feature-level data fusion was performed. Feature-level fusion aggregates the features obtained from each input and feeds them into a classifier to get a final prediction. Decision-level fusion is efficient in detecting stress even in cases when certain signals are unavailable, while it does not take into account the relationship between the features obtained from each input.
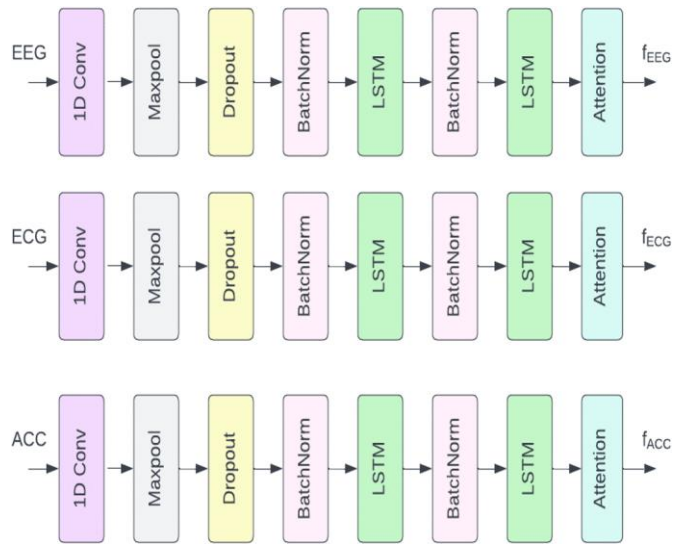


**Figure 2.** Feature extraction from each input

The fusion situations considered are $f_{ECG}+f_{EEG}$, $f_{EEG}+f_{ACC}$, $f_{ECG}+f_{ACC}$, and $f_{ECG}+f_{EEG}+f_{ACC}$. The features collected from each fusion scenario were then concatenated, and the final dense unit was in charge of feature fusion, as shown in Figure 3. The ultimate probability was determined by weighing and summing the scores of each model.
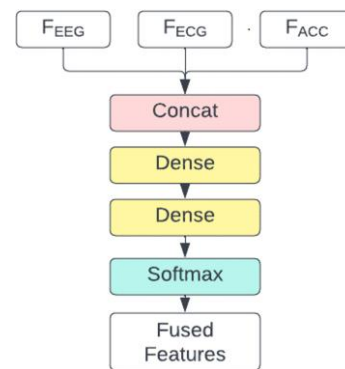


**Figure 3.** Feature-level fusion

The weights with the highest average AUC score were selected in the justification set. A grid search was utilized to find the weights.

### 3.5 Kruskal-Wallis analysis

his section assesses the efficacy of the various methods while presenting a supervised detection system. Because characteristics are not normally distributed, the Kruskal-Wallis test [61] compares data from several stages to

demonstrate a substantial disagreement. This test allows you to statistically assess if samples from different phases come from the same group. For this reason, a confidence interval(P-value), which is a value between 0 and 1, is computed for each attribute. The three types of stress have a significant statistical difference if the P-value is less than 0.01, else the feature is eliminated. Additionally, improving accuracy by developing more models were considered as follows.

1) Samples were categorized into two or more clusters using the Fuzzy C-Means (FCM) method [62], based on the similarity of their feature sets.

2) The Kruskal-Wallis test was employed to remove inappropriate features, and a distinct model was developed for each cluster.

3) Each model computes and stores the average values of each feature.

4) The dataset was initially divided into two segments, 15% and 30% respectively, which were then used to evaluate each model through the following two methods:

a) Features of the test data are retrieved based on the characteristics specific to each cluster.

b) On the second evaluation, the features of a sample are compared with the average of the corresponding features in the models. The sample is then evaluated using the model that best fits its characteristics.

The outcomes of this approach were contrasted with the earlier, more generic outcomes. This technique resolves the issue of individuals' varying stress characteristics and symptoms. However, there needs to be a discernible increase in accuracy in these customized models. Consequently, the generic supervised model is ultimately chosen. Customized models may sometimes outperform generic ones. The right sections extract features, and Kruskal-Wallis analysis is used to choose the best feature set.

## 3.6 Classifier

### 3.6.1 KNN

In this study, KNN [63] are employed for the purpose of categorization. The k-NN classifier is extensively used for classification due to its user-friendly interface and efficient processing capabilities. The process involves classifying the relevant classes by comparing the extracted features and selection technique with the nearby k-learning data. In order to reduce the likelihood of specific outcomes for a given set of learning data, we utilize k-fold validation to separate the testing and training information (where $k = 3$).

### 3.6.2 Random forest

A practical machine-learning approach for classifying stress is a random forest. The algorithm builds numerous decision trees and aggregates their predictions for final categorization. One would need to collect a dataset of stress-related characteristics to utilize random forest for stress classification. These characteristics could include physiological measurements like EEG, ACC, and ECG. Following that, the dataset would be split into testing and training sets. Using various subsets of the training phase and randomly choosing a subset of characteristics for each tree, numerous decision trees would be created using the training set to develop the random forest model.

The decision trees for this analysis will be trained using the Classification and Regression Trees (CART) or ID3 (Iterative Dichotomiser 3) algorithm. Once developed, the random forest model will be employed to classify new data instances related to stress, determining stress levels as either low or high based on stress-related features. The use of random forest for stress classification offers significant advantages; it effectively handles noisy data and missing values and is resistant to overfitting. This resistance is crucial as it prevents the model from being overly complex, which can lead to excellent performance on training data but poor generalization to new data. Consequently, it is imperative to explore the potential of the random forest as a robust and versatile algorithm for stress classification.

### 3.6.3 Decision tree

Another well-liked machine learning technique that may be utilized for stress categorization is the decision tree algorithm [64]. It operates by building a tree-like model, where each leaf node stands in for a class label, each branch for the trial result, and each interior node for a test on an attribute. A database of stress-related attributes would need to be acquired to apply a decision tree method for stress classification comparable to a random forest approach. The decision tree framework would then be constructed using the training data set after dividing the dataset into testing and training sets. Depending on a criterion like knowledge gain or the Gini index, the algorithm for decision trees would select the appropriate characteristic to divide the information into three groups. Unless a stopping requirement is satisfied, such as reaching a specific depth or having minimal occurrences in a leaf node, picking the most suitable attribute and splitting the data would be performed iteratively. The decision tree model may be used to categorize fresh instances of stress-related data after it has been constructed. The model would use the attributes associated with stress as input and proceed through the decision tree until it reached a leaf node, which would then get a class label of either low, medium, or high stress.

Decision trees may also cope with missing values and numerical and categorical data. The decision tree method is vulnerable to overfitting, which happens when the tree is very complicated and attempts to match the data noise. Over-fittings may be avoided using pruning strategies like decreased error or cost complexity. Therefore, it is worthwhile to investigate the decision tree method's potential in this field because it is a straightforward and efficient technique that may be utilized for stress categorization.

### 3.6.4 SVM

The Support Vector Machine technique and clustering models were then used to categorize the stress level. Creating an ideal hyperplane that can distinguish between the three classes using the gathered data is the underlying notion behind an SVM classifier. Selecting the kernel function to be used is the first crucial step in SVM. The feature space and mapping characteristics, essential to non-linear classification and regression in SVM, are determined by choice of the kernel function. For example, it was employed in the optimization procedure due to the fact that the SVM implementation utilizing radial basis functions and polynomial kernels can autonomously determine the quantity of centers, their positions, and their weights. Multiple studies have shown that SVM with a polynomial kernel function performed better than the SVM with a radial basis function (RBF) kernel function in all aspects and achieved the best accuracy in categorization [65]. The poly kernel was chosen for this experiment, and its expression is shown in Eq. (1).

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \qquad (1)$$

Optimal performance of the kernel function requires fine-tuning and enhancement of its parameters. In the given formula, the parameter d (4) represents the extent to which the polynomial kernel influences the adaptability of the classifier.

The retrograde linear kernel at the lowest degree, $d = 1$, is not the best for a non-linear feature. A hyperplane with a dimensionality of 2 is sufficient for effectively differentiating between the two classes and creating a decision boundary that allows for flexibility. The poly kernel with $d = 3$ was claimed to have the lowermost categorization error and increased performance but, without a doubt, substantially more time-consuming processing.

All systems are implemented using the Adam optimizer, with a mini-batch size of 64 and the default learning rate. The loss function employed is Binary Cross-Entropy (BCE) as defined in Eq. (2). Here, $P(Y_i)$ represents the predicted label for all N samples, and Yi denotes the actual label.

$$BCE(y_i, p(y_i)) =$$
$$\frac{1}{N}\sum log(p(y_i)) * log(1 - p(y_i)) * (1 - y_i) \qquad (2)$$

An early-stopping technique is employed to control the duration of training if the loss function [66] does not decrease for a continuous period of 30 epochs. The evaluation of several models is conducted based on their accuracy and F1-score.

**Algorithm:** Minimization of Loss

1: INPUT: $U_{ECG\_EEG\_ACC}$ (concatenated EEG and ECG and ACC features)
2: PARAMETERS: Weights of the hidden units $W_{h1}$, $W_{h2,}$ and $W_{h3}$
3: $W_{h1}$, $W_{h2}$ and $W_{h3} \leftarrow 0$ // Hidden unit initialization
4: $Y_{EEG\_ECG\_ACC} \leftarrow$ null // Recreated input $U_{ECG\_EEG\_ACC}$
5: $N \leftarrow$ Epoch No
6: for (i=0; i<=N; i++)
7: Use the encoder function to convert the input $U_{EEG\_ECG\_ACC}$ into a hidden illustration $h_N$.
8: $Y_{h1} = f_{h1} (U_{EEG\_ECG\_ACC}, Wh_1)$
9: $Y_{h2} = f_{h2}(U_{EEG\_ECG\_ACC}, Wh_2)$
10: $Y_{h3} = f_{h1}(U_{EEG\_ECG\_ACC}, Wh_3)$
11: The decoder function yields $_Y$ from a hidden representation of $h_N$.
12: $Y_{EEG\_ECG\_ACC} = f_Y (Y_{h3}, W_Y)$
13: Loss $\theta = L (U_{ECG\_EEG\_ACC}, Yjoint)$
14: end for
15: return $\theta$

In 10-fold cross-validation, the dataset is partitioned into 10 equivalent segments, known as folds. The overall protocol is as follows:
1. The dataset is divided into 10 subgroups of around the same size using a random process.
2. In each cycle, a single subset is selected as the validation set, while the other 9 subsets are utilised as the training set.
3. The model undergoes training using the training set.
4. Subsequently, the proficient model is assessed using the validation set, and performance measures are computed.
5. The process of steps 2-4 is iterated 10 times, with each of the 10 subsets being utilised precisely once as the validation set.

6. Following all iterations, the performance measures, such as accuracy and error, from each fold are averaged to provide a singular assessment of the model's performance.

It is crucial to acknowledge that throughout each iteration, the hyperparameters of the model are often adjusted by utilising a distinct validation set that is part of the training set. Furthermore, once the cross-validation process is complete, the ultimate model may be trained using the complete dataset, which includes the validation set. This trained model can then be used for deployment or for additional assessment using a separate test set, which constitutes 30% of the data.

### 3.7 Hyperparameter optimization

The primary goal is to enhance the hyperparameters of the LSTM classifier using the advanced WOA algorithm to increase performance in stress classification. This study [67] focuses on the parameters inside the framework, namely the batch size and the quantity of hidden neurons. The enhanced WOA method begins by initialising the hyperparameters using randomly generated initial solutions. It then iteratively works to enhance the accuracy of the stress classification model until the stopping requirements are satisfied. The fitness function in LSTM networks evaluates and provides the accuracy of stress categorization.

**Algorithm:** Enhanced Whale Optimization Algorithm

**Input** ← number of whales (N), maximum iterations (T_max),
Max episodes (E_max ),Max_steps(T)
**Output** ← optimized set of (α actor, α critic, γ, batch size, τ) hyperparameters.
Begin
set number of whales (N)
Initialize whales' positions( $\alpha_{actor}, \alpha_{critic}$, batch size, r) randomly
while (t <T_max )
For (i=1 to N)
Randomly initialize main critic network Q(s,a) and main actor-network μ(s) with weights ω and θ
Initialize target critic network ϕ(s,a) and target actor-network μ'(s) with weights ω and θ
Initialize replay buffer R
Set hyperparameters (a_"acton" α_"critic" Y, batch size, r) as in position vector of the current whale
For (i=1 to E_max ) do
Initialize action exploration process N
Receive initial state $s_1$ from environment
For j=1 to T do
Execute action a_t=μ(s_t|θ)+N
Observe reward $r_t$ a and successor state s_(t+1)
Store experience ($s_t$, $a_t$, $r_t$,$s_{t+1}$) in R
Set $y_i = r_i + \gamma\phi(s_{i+1}, \mu(s_{i+1} \mid \dot\theta) \mid \dot\theta)$
Update the critic by minimizing the loss in Eq $L = \frac{1}{N}\Sigma(y_i - Q(s_i, a_i \mid \omega))^2$
Update the target network weight according to Eq. $\dot\omega \leftarrow \tau\omega + (1 - \tau)\dot\omega$
End for
End for
Set the fitness of each whale to the accumulated train rewards value
Find the best whale X* with the highest fitness

Update parameters (a, A,c,and p)
Update whales' positions
End for
Update the current best whale X*
t = t+1
End while
Return X*
End

## 4. RESULTS AND DISCUSSION

By combining several characteristics, the categorization system may utilise a wide range of information sources to make more knowledgeable selections. Feature fusion approaches amalgamate characteristics in a manner that amplifies their ability to differentiate between stress and non-stress conditions. This can result in a more efficient differentiation of stress-induced patterns. Human stress is a multifaceted occurrence that is affected by a range of elements, such as individual variations, situational circumstances, and measuring inaccuracies. Feature fusion strategies enhance the resilience of the classification system by amalgamating data from different sources, hence mitigating the influence of fluctuations in individual feature domains. Feature fusion approaches can effectively decrease the dimensionality of the input data while retaining important information.

By mitigating the curse of dimensionality, the efficiency of the classification process can be improved. Feature fusion strategies enhance model generalisation by identifying underlying patterns that are constant across diverse populations, situations, or measurement modalities. This can result in enhanced stress categorization accuracy across a wide range of situations. Feature fusion techniques may be modified and tailored to meet the unique needs and limitations of the stress categorization task. Researchers can conduct experiments using various fusion tactics and combinations to enhance performance, taking into account the data and domain

knowledge that is accessible. Feature fusion approaches enable researchers to integrate pre-existing information or specialised knowledge into the process of categorization. To enhance the effectiveness of stress categorization, it may be necessary to incorporate domain-specific connections or restrictions into the fusion model through encoding.

**Table 2.** Metrics based on a different combination of fusion methods

| Combination | Accuracy | AUC | F1-Score |
|---|---|---|---|
| $f_{ECG}+f_{EEG}$ | 64.3 | 0.56 | 0.39 |
| $f_{EEG}+f_{ACC}$ | 68.7 | 0.53 | 0.36 |
| $f_{ECG}+f_{ACC}$ | 82.2 | 0.72 | 0.48 |
| $f_{ECG}+f_{EEG}+f_{ACC}$ | 94.58 | 0.82 | 0.70 |

Table 2 includes the metrics for the three-level stress categorization. The feature-level fusion of $f_{ECG}+f_{EEG}+f_{ACC}$ had the most excellent average accuracy (94.58%), AUC (0.82), and F1 score among the several fusion techniques (0.70). There were substantial variations in accuracy and AUC when performance was compared to the baseline model (t-test, 5% significance level). The lowest average accuracy (64.3%), lowest average AUC (0.53), and lowest F1 score were shown by the feature-level fusion $f_{ECG}+f_{EEG}$ and $f_{EEG}+f_{ACC}$, respectively (0.36). When given physiological signals ($f_{ECG}$, fees, and $f_{ACC}$), $f_{ECG}+f_{EEG}+f_{ACC}$ performed better in feature-level fusion situations than $f_{ECG}+f_{EEG}$, $f_{EEG}+f_{ACC}$, or $f_{ECG}+f_{ACC}$.

The sensitivity, specificity, and the system's accuracy are evaluated using the training and testing datasets, which are picked at random. The average of the rounds' results is calculated using 70% of the data for testing and 30% for training. To find the optimal model, many supervised techniques are used and contrasted. The Gini index is the split criterion while creating a decision tree. One hundred decision tree learners are produced by using the random forest bootstrap. The radial basis function creates the K Nearest Neighbor using $K = 3$ as the SVM kernel. In this instance, the algorithms' chosen parameters produced the greatest accuracy.

**Table 3.** Results of various algorithms based on the K-EmoCon dataset

| EEG+ECG+ACC | Accuracy (During Training) | Accuracy (During Testing) | Sensitivity (During Training) | Sensitivity (During Testing) | Specificity (During Training) | Specificity (During Testing) |
|---|---|---|---|---|---|---|
| Decision tree μ±SD | 87.6±1.8 | 87.4±4.6 | 87.6±1.5 | 81.8±5.9 | 86.1±3.0 | 83.5±9.4 |
| Random forest μ±SD | 90.0±0.0 | 92.7±3.2 | 87.7±1.2 | 84.1±3.1 | 84.2±2.0 | 82.3±5.5 |
| KNN μ±SD | 86.1±1.4 | 93.4±2.5 | 87.0±1.1 | 95.2±3.1 | 94.8±2.2 | 86.6±5.5 |
| SVM μ±SD | 94.0±0.0 | 94.58±1.9 | 90±0.0 | 86.8±2.7 | 90.0±0.0 | 93.4±3.3 |

μ- Mean, SD- Standard Deviation

**Table 4.** Comparison of recent studies using the DL approach

| Research | Input Data | No. of Classes | Classifier | Accuracy |
|---|---|---|---|---|
| [20] | ECG | 2 | KNN | 93.12 |
| [30] | ECG,EDA,EEG | 2 | SVM | 94.4 |
| [29] | EDA | 3 | LDA | 85.05 |
| [28] | HRV | 2 | MLP | 86 |
| [9] | PPG,EDA,ACC | 2 | MLP | 85.06 |
| [22] | ECG,EDA,EMG | 2 | LDA | 82.59 |
| [27] | PPG,EDA,TEMP | 2 | RF | 94 |
| Our Work | EEG, ECG, ACC | 3 | SVM | 94.58 |

For example, a loop function that iteratively assigns multiple values to the variable K and identifies the model with the highest level of accuracy is used to identify the optimal value of K. Along with these techniques, a Decision Tree Network, Random Forest, and a group of learners using KNN were first looked at as well but were disregarded due to their poor accuracy.

Table 3 displays the general efficacy of several approaches. Furthermore, the standard deviation is included. If all of the data are utilized to create the final model, the test column's scores demonstrate how well we expect to be able to identify stress.

Table 4 presents a comparison of different research efforts and their respective performances. Based on our knowledge, our work is the first to introduce a fusion strategy using deep learning that incorporates many types of data and different data sources to investigate job-related stress. This approach is discussed in the Related Works section. While we applied a well-trained network to assess the probability of human emotions, they used CNN-LSTM [20] and Bi-LSTM [22] to interpret physiological data and the order of psychological attributes, respectively. Regarding the classification problem, [29] considered three levels of stress, [27] two levels.

However, in our study, we evaluated three levels of stress. Except for one study [28] that employed a driving simulator as the experimental environment, most trials were static, including the one we looked at. Our study did well in terms of accuracy compared to earlier DL investigations. Next, compared to paper [27], our analysis showed similar accuracy in the three-level stress categorization. Lastly, our examination showed that, compared to the DL study [29], the three-level stress categorization had a lower accuracy. The results of our investigation and previous studies could have been more consistent. This discrepancy might be explained by several factors, such as experimental methods, the number of participants (size of the datasets), the window size, or the kind of data. The confusion matrix summarizes the three-level stress SVM classification's performance.

Table 5 demonstrates that 30 real examples from the low-stress category were accurately identified as belonging to that condition. Out of the total 75 occurrences, this accounted for

40%, and all of those instances were accurately categorized into the appropriate class. Regarding the moderate degree of stress, 25 occurrences, up to 36%, were diagnosed correctly, and the accuracy rate for that class was 93%. Two actual events from the high-stress class, or 1% of the 18 highly stressed forecasts, which amounted to 15 instances, or 20%, were incorrectly categorized as moderate stress. The confusion matrix indicates that the three-level stress categorization achieved an accuracy of 94% in predicting stress levels, with only 6% of predictions being wrong.

Table 6 displays the performance of all models, showing the average accuracy together with the standard deviation in various iterations. The Random Forest method had a shorter processing time than the others, but its performance was lower than KNN and SVM algorithms and greater than Decision Tree algorithms. Conversely, the SVM method had the longest processing time across all iterations. Nevertheless, its performance surpassed that of all other algorithms.

**4.1 t-SNE visualization**

We reveal the t-distributed Stochastic Neighbor Embedding (t-SNE) [68] before and throughout the procedure to better understand the joint feature learning that our model performed. The t-SNE method maintains two points nearby if they have the same distribution by projecting multi-dimensional data into two-dimensional or three-dimensional regions. Similar separations between distant places may be seen in the t-SNE projections. We utilize t-SNE to map the combined characteristics onto a 2-Dimensional plane.

Figure 4 illustrates the feature visualization of UECG-EEG-ACC (standard features) and joint aspects acquired by the MSE cost function on the entire signal of all benchmark datasets. The red dots stand for ECG characteristics, the blue for ACC attributes, and the green for EEG traits. Combining features extraction aims to combine the traits of many modalities in one place. After collaborative feature learning, we saw substantial overlap in visualization across the ECG, EEG, and ACC modalities. Figure 4 shows that the modality gap between the distribution of modalities has significantly decreased.

**Table 5.** Confusion matrix for Three level stress classification by the SVM

| | | Predicted Class | | Accuracy (%) |
| --- | --- | --- | --- | --- |
| | | Low Level | Moderate Level | High Level | |
| Actual Class | Low Level | 30 | 0 | 0 | 100 |
| | Moderate Level | 1 | 25 | 1 | 93 |
| | High Level | 1 | 2 | 15 | 84 |

**Table 6.** The average accuracy and time of different models during training

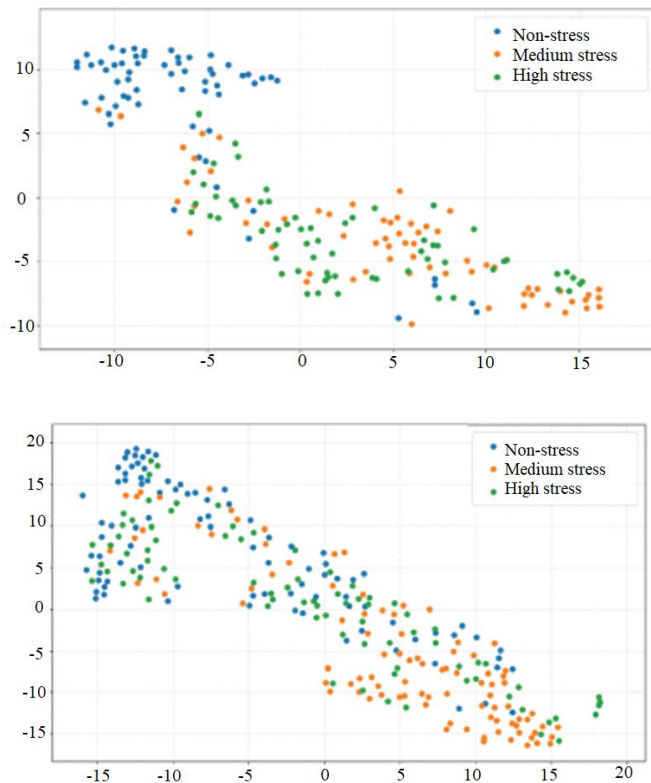| Models | No. Iteration | Time | Accuracy (Mean±std) |
| --- | --- | --- | --- |
| Decision tree | 50 | 13h | 58.16±2.1 |
| | 75 | 17h | 59.95±2.1 |
| | 100 | 21h | 87.6±4.46 |
| Random forest | 50 | 9h | 70.17±1.4 |
| | 75 | 12h | 72.98±1.4 |
| | 100 | 14h | 90.35±1.06 |
| KNN | 50 | 15h | 73.32±0.7 |
| | 75 | 20h | 76.99±9.4 |
| | 100 | 23h | 86.61±1.7 |
| SVM | 50 | 19h | 82.77±1.1 |
| | 75 | 21h | 83.91±1.9 |
| | 100 | 25h | 94.52±1.3 |

**Figure 4.** Outcomes of the feature-based fusion for categorizations of stress using $f_{ECG}$, $f_{EEG}$, and $f_{ACC}$ demonstrate the highest/lowest accuracy

## 5. CONCLUSION

Using multimodal and heterogeneous data, we provided a DL technique in this article for precisely identifying work-related stress. We created the DNN structures that handled the K-EmoCon multimodal dataset's ECG, EEG, and ACC feature data. After that, we fused features at the feature level. The feature-level fusion $f_{ECG}+f_{EEG}+f_{ACC}$ had the most significant average performance in three-level stress categorizations. By employing feature level fusion and utilizing a poly kernel, the SVM achieved an impressive accuracy of 94% in categorizing the data into three distinct stress levels: low, moderate, and high. The model created by the DL technique may assist in enhancing the mental healthcare of workers, reducing stress-related costs, and adequately detecting stress.

This study suggested an alternative framework for developing strong and dependable stress categorization models utilising sophisticated deep learning methods. Although the results of this study are positive, it is important to recognise some limitations. The suggested approach can help identify suitable LSTM hyperparameter values for enhancing stress classification. It is recommended to evaluate the algorithm's performance by comparing it to other classifiers and testing it in various circumstances. In this work, we solely examined the efficacy of the suggested approach on two hyperparameters: batch size and the number of hidden neurons. It would be beneficial to investigate the effectiveness of the suggested framework on different hyperparameters.

The suggested temporal multimodal fusion demonstrated satisfactory performance, however employing a more intricate classifier with additional layers might enhance stress classification performance. It is advisable to apply this method

to several datasets to assess its efficacy in stress categorization. These methods are exclusively available for offline data processing, prioritising accuracy above computing efficiency. Training processes can be placed on the cloud in order to reduce response time in the era of cloud computing. The presented models are appropriate for real-time affective intelligence systems because of the tiny window widths considered, ranging from 1 to 10 seconds.

The suggested algorithms for stress detection may be integrated into many settings like e-health monitoring, mental health treatment, intelligent teaching, or gaming. The algorithms need enhancement to provide quicker and more efficient results for the new applications and should be implemented in cloud computing.

## REFERENCES

[1] Uday, S., Jyotsna, C., Amudha, J. (2018). Detection of stress using wearable sensors in IoT platform. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, pp. 492-498. https://doi.org/10.1109/ICICCT.2018.8473010

[2] Panicker, 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. https://doi.org/10.1016/j.bbe.2019.01.004

[3] 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: 103139. https://doi.org/10.1016/j.jbi.2019.103139

[4] Aalbers, G., McNally, R., Heeren, A., de Wit, S., Fried, E. (2018). Social media and depression symptoms: A network perspective. Journal of Experimental Psychology: General, 148(8): 1454-1462. https://doi.org/10.1037/xge0000528

[5] 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. https://doi.org/10.3390/s19081849

[6] Alberdi, A., Aztiria, A., Basarab, A. (2015). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Journal of Biomedical Informatics, 59: 49-75. https://doi.org/10.1016/j.jbi.2015.11.007

[7] Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A. et al. (2019). Detecting moments of stress from measurements of wearable physiological sensors. Sensors, 19(17): 3805. https://doi.org/10.3390/s19173805

[8] Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., Yang, X. (2018). A review of emotion recognition using physiological signals. Sensors, 18(7): 2074. https://doi.org/10.3390/s18072074

[9] Koh, K., Park, J.K., Kim, C., Cho, S. (2001). Development of the stress response inventory and its application in clinical practice. Psychosomatic Medicine, 63: 668-678. http://doi.org/10.1097/00006842-200107000-00020

[10] Holmes, T.H., Rahe, R.H. (1967). The social readjustment rating scale. Journal of Psychosomatic Research, 11(2): 213-218. https://psycnet.apa.org/doi/10.1016/0022-

3999(67)90010-4

[11] Williams, J. (1988). A structured interview guide for the Hamilton depression rating scale. Archives of General Psychiatry, 45(8): 742-747. https://doi.org/10.1001/archpsyc.1988.01800320058007

[12] Cohen, S., Janicki-Deverts, D., Miller, G.E. (2007). Psychological stress and disease. Jama, 298(14): 1685-1687. https://doi.org/10.1001/jama.298.14.1685

[13] Kuntamalla, S., Lekkala, R.G.R. (2018). Quantification of error between the heartbeat intervals measured from photoplethysmogram and electrocardiogram by synchronization. Journal of Medical Engineering and Technology, 42(5): 389-396. https://doi.org/10.1080/03091902.2018.1513578

[14] Bashar, S.K., Han, D., Fearass, Z., Ding, E., Fitzgibbons, T., Walkey, A., Mcmanus, D., Javidi, B., Chon, K. (2020). Novel density poincare plot based machine learning method to detect atrial fibrillation from premature atrial/ventricular contractions. IEEE Transactions on Biomedical Engineering, 68(2): 448-460. http://doi.org/10.1109/TBME.2020.3004310

[15] Muaremi, A., Arnrich, B., Tröster, G. (2013). Towards measuring stress with smartphones and wearable devices during workday and sleep. BioNanoScience, 3: 172-183. http://doi.org/10.1007/s12668-013-0089-2

[16] Padma, V., Anand, N., Gurukul, S., Javid, S., Prasad, A., Arun, S. (2015). Health problems and stress in Information Technology and Business Process Outsourcing employees. Journal of Pharmacy and Bioallied Sciences, 7(Suppl 1): S9-S13. https://doi.org/10.4103/0975-7406.155764

[17] Smets, E., Rios, V.E., Schiavone, G., Chakroun, I., D'Hondt, E., De Raedt, W., Cornelis, J., Janssens, O., Hoecke, S., Claes, S., Diest, I., Van Hoof, C. (2018). Large-scale wearable data reveal digital phenotypes for daily-life stress detection. Digital Med, 1: 67. https://doi.org/10.1038/s41746-018-0074-9

[18] Park, S., Min, K., Chang, S., Kim, H., Min, J. (2008). Job stress and depressive symptoms among Korean employees: The effects of culture on work. International Archives of Occupational and Environmental Health, 82: 397-405. https://doi.org/10.1007/s00420-008-0347-8

[19] Hajera, S., Ali, M.M. (2018). A comparative analysis of psychological stress detection methods. International Journal of Computational Engineering and Management, 21(2): 1-8.

[20] Ciabattoni, L., Ferracuti, F., Longhi, S., Pepa, L., Romeo, L., Verdini, F. (2017). Real-time mental stress detection based on smartwatch. In 2017 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, pp. 110-111. https://doi.org/10.1109/ICCE.2017.7889247

[21] Alyan, E., Mohamad, S., Mohamad N., Kamel, N. (2020). Effects of Workstation Type on Mental Stress: fNIRS Study. Human Factors: The Journal of the Human Factors and Ergonomics Society, 63: 001872082091317. https://doi.org/10.1177/0018720820913173

[22] Picard, R.W. (2016). Automating the recognition of stress and emotion: From lab to real-world impact. IEEE Multimedia Management, 23(3): 3-7. https://doi.org/10.1109/MMUL.2016.38

[23] Bijalwan, V., Semwal, V., Mandal, T. (2021). Fusion of multi-sensor-based biomechanical gait analysis using vision and wearable sensor. IEEE Sensors Journal, 21(13): 14213-14220. https://doi.org/10.1109/JSEN.2021.3066473

[24] Patil, P., Kumar, K.S., Gaud, N., Semwal, V. (2019). Clinical human gait classification: Extreme learning machine approach. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, pp. 1-6. https://doi.org/10.1109/ICASERT.2019.8934463

[25] Sharma, S., Sharma, M. (2021). A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans. Computers in Biology and Medicine, 134: 104450. https://doi.org/10.1016/j.compbiomed.2021.104450

[26] Adams, P., Rabbi, M., Rahman, T., Matthews, M., Voida, A., Gay, G., Choudhury, T., Voida, S. (2014). Towards personal stress informatics: Comparing minimally invasive techniques for measuring daily stress in the wild. In Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, Oldenburg Germany, pp. 72-79. https://doi.org/10.4108/icst.pervasivehealth.2014.254959

[27] Lee, D., Chong, T., Lee, B.G. (2016). Stress events detection of driver by wearable glove system. IEEE Sensors Journal, 17(1): 194-204. https://doi.org/10.1109/JSEN.2016.2625323

[28] Choi, M., Koo, G., Seo, M., Kim, S.W. (2017). Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness. IEEE Transactions on Instrumentation and Measurement, 67(3): 634-645. https://doi.org/10.1109/TIM.2017.2779329

[29] Zheng, Y., Wong, T., Leung, B., Poon, C. (2016). Unobtrusive and multimodal wearable sensors to quantify anxiety. IEEE Sensors Journal, 16(10): 3689-3696. https://doi.org/10.1109/JSEN.2016.2539383

[30] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E. (2002). Wireless sensor networks: A survey. Computer Networks, 38: 393-422. https://doi.org/10.1016/S1389-1286(01)00302-4

[31] Vemuri, V. (2020). The hundred-page machine learning book. Journal of Information Technology Case and Application Research, 22(2): 136-138. https://doi.org/10.1080/15228053.2020.1766224

[32] Carroll, D., Davey S.G., Shipley, M., Brunner, E., Marmot, M. (2001). Blood pressure reactions to acute psychological stress and future blood pressure status: A 10Year Follow-Up of Men in the Whitehall II Study. Psychosomatic Medicine, 63(5): 737-743. https://doi.org/10.1097/00006842-200109000-00006

[33] Garg, A., Chren, M.M., Sands, L.P., Matsui, M.S., Marenus, K.D., Feingold, K.R., Elias, P.M. (2001). Psychological stress perturbs epidermal permeability barrier homeostasis: Implications for the pathogenesis of stress-associated skin disorders. Archives of Dermatology, 137(1): 53-59. https://doi.org/10.1001/archderm.137.1.53

[34] Adam, T. (2007). Stress, eating and the reward system. Physiology & Behavior, 91(4): 449-458. https://doi.org/10.1016/j.physbeh.2007.04.011

[35] Lupien, S., Maheu, F., Tu, M., Fiocco, A., Schramek, T. (2008). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. Brain and Cognition, 65(3): 209-237. https://doi.org/10.1016/j.bandc.2007.02.007

[36] Atsan, N. (2016). Decision-making under stress and its implications for managerial decision-making: A review of literature. International Journal of Business and Social Research, 6(3): 38. https://doi.org/10.18533/ijbsr.v6i3.936

[37] Cooper, C.L., Cartwright, C. (1997). An intervention strategy for workplace stress. Journal of Psychosomatic Research, 43(1): 7-16. https://doi.org/10.1016/S0022-3999(96)00392-3

[38] Kompier, M., Cooper, C. (1999). Preventing Stress, Improving Productivity. London: Routledge.

[39] Lazarus, R.S. (2020). Psychological stress in the workplace. In Occupational Stress. CRC Press, pp. 3-14.

[40] McVicar, A. (2003). Workplace stress in nursing: A literature review. Journal of Advanced Nursing, 44(6): 633-642. https://doi.org/10.1046/j.0309-2402.2003.02853.x

[41] Srivastava, N., Salakhutdinov, R.R. (2012). Multimodal learning with deep boltzmann machines. Advances in Neural Information Processing Systems, 25.

[42] Miah, A.S.M., Shin, J., Islam, M.M., Molla, M.K.I. (2022). Natural human emotion recognition based on various mixed reality (MR) games and electroencephalography (EEG) signals. In 2022 IEEE 5th Eurasian Conference on Educational Innovation (ECEI), Taipei, Taiwan, pp. 408-411. https://doi.org/10.1109/ECEI53102.2022.9829482

[43] Zhen, L., Hu, P., Wang, X., Peng, D. (2019). Deep supervised cross-modal retrieval. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 10394-10403. https://doi.org/10.1109/CVPR.2019.01064

[44] Liu, R., Zhao, Y., Wei, S., Zheng, L., Yang, Y. (2019). Modality-invariant image-text embedding for image-sentence matching. ACM Transactions on Multimedia Computing, Communications, and Applications, 15(1): 27. https://doi.org/10.1145/3300939

[45] Surís, D., Duarte, A., Salvador, A., Torres, J., Giró-i-Nieto, X. (2018). Cross-modal embeddings for video and audio retrieval. In Computer Vision – ECCV 2018 Workshops, Munich, Germany, pp. 711-716. https://doi.org/10.1007/978-3-030-11018-5_62

[46] Hazarika, D., Zimmermann, R., Poria, S. (2020). Misa: Modality-invariant and-specific representations for multimodal sentiment analysis. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, pp. 1122-1131. https://doi.org/10.1145/3394171.3413678

[47] Healey, J.A., Picard, R.W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems, 6(2): 156-166. https://doi.org/10.1109/TITS.2005.848368

[48] Wijsman, J., Grundlehner, B., Liu, H., Penders, J., Hermens, H. (2013). Wearable physiological sensors reflect mental stress state in office-like situations. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, pp. 600-605. https://doi.org/10.1109/ACII.2013.105

[49] Sriramprakash, S., Prasanna, V.D., Murthy, O.R. (2017). Stress detection in working people. Procedia Computer Science, 115: 359-366. https://doi.org/10.1016/j.procs.2017.09.090

[50] Betti, S., Lova, R.M., Rovini, E., Acerbi, G. et al. (2017). Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers. IEEE Transactions on Biomedical Engineering, 65(8): 1748-1758. https://doi.org/10.1109/TBME.2017.2764507

[51] Elzeiny, S., Qaraqe, M. (2018). Blueprint to workplace stress detection approaches. In 2018 International Conference on Computer and Applications (ICCA), Beirut, Lebanon, pp. 407-412. https://doi.org/10.1109/COMAPP.2018.8460293

[52] Ugulino, W., Cardador, D., Vega, K., Velloso, E., Milidiú, R., Fuks, H. (2012). Wearable computing: Accelerometers' data classification of body postures and movements. In Advances in Artificial Intelligence-SBIA 2012: 21th Brazilian Symposium on Artificial Intelligence, Curitiba, Brazil, pp. 52-61. https://doi.org/10.1007/978-3-642-34459-6_6

[53] Palumbo, F., Gallicchio, C., Pucci, R., Micheli, A. (2016). Human activity recognition using multisensor data fusion based on Reservoir Computing. Journal of Ambient Intelligence and Smart Environments, 8(2): 87-107. https://doi.org/10.3233/AIS-160372

[54] Chen, K., Yao, L., Wang, X., Zhang, D., Gu, T., Yu, Z., Yang, Z. (2018). Interpretable parallel recurrent neural networks with convolutional attentions for multi-modality activity modeling. In 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, pp. 1-8. https://doi.org/10.1109/IJCNN.2018.8489767

[55] Lawal, I., Bano, S. (2019). Deep human activity recognition using wearable sensors. In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Rhodes Greece, pp. 45-48. https://doi.org/10.1145/3316782.3321538

[56] Hwang, B., You, J., Vaessen, T., Myin-Germeys, I., Park, C., Zhang, B.T. (2018). Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals. Telemedicine and e-Health, 24(10): 753-772. https://doi.org/10.1089/tmj.2017.0250

[57] Park, C.Y., Cha, N., Kang, S., Kim, A., Khandoker, A. H., Hadjileontiadis, L., Oh, A., Jeong, Y., Lee, U. (2020). K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. Scientific Data, 7(1): 293. https://doi.org/10.1038/s41597-020-00630-y

[58] Jyoti B.S., Mufti, M., David, B. (2023). Heart rate variability-based mental stress detection: An explainable machine learning approach. SN Computer Science, 4: 176. https://doi.org/10.1007/s42979-022-01605-z

[59] Ronao, C., Cho, S. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 59: 235-244. https://doi.org/10.1016/j.eswa.2016.04.032

[60] Hammerle, N.Y., Halloran, S., Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. arXiv preprint arXiv:1604.08880. https://doi.org/10.48550/arXiv.1604.08880

[61] Jiang, W., Yin, Z. (2015). Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM International

Conference on Multimedia, Brisbane Australia, pp. 1307-1310. https://doi.org/10.1145/2733373.2806333

[62] Peng, L., Chen, L., Ye, Z., Zhang, Y. (2018). AROMA: A deep multi-task learning based simple and complex human activity recognition method using wearable sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(2): 1-16. https://doi.org/10.1145/3214277

[63] Chen, K., Yao, L., Zhang, D., Guo, B., Yu, Z. (2019). Multi-agent attentional activity recognition. In Proceedings of 28th International Joint Conference on Artificial Intelligence, Macao, China, pp. 1344-1350.

[64] Uddin, M.Z., Hassan, M., Alsanad, A., Savaglio, C. (2019). A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. Information Fusion, 55: 105-115. https://doi.org/10.1016/j.inffus.2019.08.004

[65] Iqbal, T., Elahi, A., Wijns, W., Amin, B., Shahzad, A. (2023). Improved stress classification using automatic feature selection from heart rate and respiratory rate time signals. Applied Sciences, 13(5): 2950. https://doi.org/10.3390/app13052950

[66] Kashem, M.A., Ganapathy, V., Jasmon, G.B., Buhari, M.I. (2000). A novel method for loss minimization in distribution networks. In DRPT2000. In International conference on electric utility deregulation and restructuring and power technologies. Proceedings (Cat. No. 00EX382), London, UK, pp. 251-256. https://doi.org/10.1109/DRPT.2000.855672

[67] Ashraf, N.M., Mostafa, R.R., Sakr, R.H., Rashad, M.Z. (2021). Optimizing hyperparameters of deep reinforcement learning for autonomous driving based on whale optimization algorithm. PLoS ONE, 16(6): e0252754. https://doi.org/10.1371/journal.pone.0252754

[68] Arora, S., Hu, W., Kothari, P.K. (2018). An analysis of the t-SNE algorithm for data visualization. In Proceedings of the 31st Conference on Learning Theory, PMLR, 75: 1455-1462.