IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 1, March 2024, pp. 337∼347
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i1.pp337-347

Machine learning-based stress classification system using wearable sensor devices

Varun Chandra, Divyashikha Sethia
Department of Software Engineering, Delhi Technological University, New Delhi, India

Article Info

Article history:

Received Aug 6, 2022
Revised Feb 13, 2023
Accepted Mar 10, 2023

Keywords:

Electrodermal activity
Electroencephalogram
Heart rate
K-nearest neighbors
Montreal imaging stress task
Random forest
Skin temperature
Stress

ABSTRACT

University students often become victims of high-stress levels due to the highly competitive work environment. Unmonitored stress levels in students can inflict severe physiological health problems. This work aims to build a stress classification framework using wearable sensor devices to predict mental stress levels for undergraduate engineering students. It comprises a study to collect a data set of 23 university students using wearable devices for four physiological signals, i.e., electroencephalogram (EEG), electrodermal activity (EDA), skin temperature (SKT), and heart rate (HR), when the students perform the montreal imaging stress task (MIST) for the mental workload. The machine learning models proposed in this work help classify stress into three levels: rest, moderate, and high. The models achieve a classification accuracy of 99.98% using the EEG signals' time-frequency domain features and an accuracy of 99.51% using the EDA, HR, and SKT signals. The proposed models achieve better scores than all the previous studies on stress classification, using EEG signals and EDA, HR, and SKT signals. This study is novel since it also demonstrates the applicability and proficiency of wearable sensor devices in developing accurate stress classification models to help build real-time stress monitoring systems.

Corresponding Author:

Divyashikha Sethia
Department of Software Engineering, Delhi Technological University
Bawana Rd, Rohini, New Delhi, Delhi 110042, India
Email: sethiadivya@gmail.com

1. INTRODUCTION

Stress occurs due to a person's inability to handle his mental and emotional states during a challenging situation. It is described by Hans Selye as "an unspecific response of a human body to the demand of task" [1]. There are two categorizations of stress, namely short-term and long-term stress. Cohen *et al.* [2] concluded that there is a direct association between long-term psychological stress and diseases like depression, human immunodeficiency virus (HIV) / acquired immune deficiency syndrome (AIDS), and cardiovascular diseases. High stress can also cause chronic illnesses such as stroke and diabetes. Tasks involving a high mental workload can induce stress, especially for academic and placement assessments of students. People in academia are constantly engaged in such tasks. Students in universities and schools face many mentally demanding and challenging situations, like examinations, peer pressure, teachers, and job interviews. In such a socially competitive and mentally exhausting environment, it is often the case that students become the victims of stress and its associated health risks. Nandi *et al.* [3] surveyed university medical students and found 53% of the students who participated in the study were stressed, and there were significant effects on the mental and social well-being of these participants. Behere *et al.* [4] proposed a study that involved a questionnaire-based survey of 100 random students. The study found that medical and engineering students had high-stress levels requiring immediate medical attention. It also concluded that students not attending to their high-stress levels could cause severe mental and psychosocial problems.

Although several psychological tests can help assess stress levels, they require assessment from a psychologist and depend on how well the participant has answered the questions. Several studies have recorded an individual's stress response using self-report questionnaires, like the perceived stress scale (PSS) [5]. However, these questionnaires are subjective methods of evaluating stress and might not have any significant correlation with the stress levels of an individual [6].

An alternative is to have stress classification systems with physiological signals that can overcome the limitations of these questionnaires, assess and quantify stress levels, take timely preventive measures, and avoid risks. Physiological signals are more related to the body's vital elements, including cardiac activity electrocardiogram (ECG), blood volume pulse (BVP), brain function, exocrine activity, and muscle excitability estimated using electromyography (EMG). Various wearable sensors in the market can help measure such signals accurately. In this study, our primary focus is on four physiological signals — electroencephalogram (EEG), electrodermal activity (EDA), skin temperature (SKT), and heart rate (HR). Numerous studies use EEG for applications such as assistive smart homes [7], seizure detection [8], driver drowsiness [9], classification of autism [10] and mental stress detection. EEG can capture localized signals of the brain regions that generate the stress response and can give many robust features for assessing stress [11]. Many researchers have observed the associations of variations in alpha, beta, and theta power bands of EEG signals with mental states and have exploited them for feature extraction [12]. In addition to EEG, many stress studies consider EDA, a measure of electric current flowing through the skin. Human skin has millions of sweat glands that can get activated under stress. Activation of sweat glands increases the amount of skin moisture secreted from the body [13], which varies the skin conductance at that region of the skin. Several studies show a high correlation between the stress levels of an individual and skin conductance [14]. High stress and anxiety can also cause variations in SKT. Different parts of the body show different patterns of temperature variations. Contrary to most beliefs, SKT can increase or decrease during stress. In a study by Vinkers *et al.* [15], when participants took the trier social stress test (TSST) [16], the authors observed that the upper region of the participant's arm showed a significant increase in SKT. Another commonly adopted measure for detecting stress is the heart rate (HR), measured as the number of heartbeats per minute. During stress, the human body releases adrenaline, a hormone that can cause the heart rate and breathing rate of a person to rise. In literature, many studies have reported that during stress, a person's heart rate rises significantly [15], [17]–[19].

Related work: several stressors can be used to study the impact of stress, such as academic and mental tasks. Montreal imaging stress task (MIST) task [20] for the mental workload. It comprises several mental arithmetic questions with single-digit answers and varying difficulty levels. MIST is nearest to the scenario when the students appear for an exam or a placement test in real life. Hence, this paper considers using MIST task to monitor stress in university students.

EEG is the most widely used biochemical signal to study brain functions due to the availability of non-invasive, easy-to-use, portable, and low-cost EEG devices. In a recent study by Sharma and Khyati [21], the authors made an EEG-based database for early stress detection. They used Hilbert-Huang transform (HHT) to decompose the EEG signals into intrinsic mode functions (IMFs) and extract related features in the Time-Frequency domain. They classified stress into three levels, low stress, medium stress, and high stress. A hierarchical support vector machine (SVM) model was trained, which achieved an accuracy of 92.86%. Blanco *et al.* [22] used a computer-based version of the Stroop test [23] and the Emotiv EPOC Headset to collect a dataset of 18 subjects. The raw EEG signals obtained were cleaned by subtracting the least-squares line of best fit and bandpass filter network of Chebyshev Type II filters. They used logistic regression, quadratic discriminant analysis (QDA), and k-nearest neighbors (KNN), achieving a maximum accuracy of 78.70% in the stress classification task.

On the other hand, many researchers choose physiological signals other than EEG to study stress. Airij *et al.* [24] proposed a system that monitored only three physiological signals-heart rate, skin conductance, and skin temperature of patients. The system stored the patients' physiological signals and stress level data for record maintenance and sent it to the concerned doctors for remote tracking of the patients. The authors used a rule-based fuzzy logic algorithm for stress classification, which obtained an accuracy of 96.19%.

There have been previous studies to evaluate stress using MIST [20]. Minguillon *et al.* [25] used MIST with a small number of 10 subjects and achieved an accuracy of 50% with LDA with EEG features and

86% with EEG, EMG, ECG, and GSR features. Xi *et al.* [26] achieved a lower accuracy with SVM using only EEG features. However, they did not explore Time Domain features. Al-Shargie *et al.* [27], used MIST [20] to elicit stress in 12 participants and recorded their EEG signals using the Brain Master 24E system. The collected data was filtered using a 0.5 Hz to 30 Hz bandpass filter implemented by using a Butterworth 3rd-order filter. The independent component analysis (ICA) was applied to remove the artifacts. They used an SVM classifier for a three-level classification and obtained mean accuracies of 94%, 85%, and 80% for three mental stress levels one, two, and three, respectively. Jun and Smitha [28] designed an automatic EEG-based stress recognition system that used Stroop test [23], and MIST [20] to induce low levels and high levels of stress, respectively. They trained an SVM model on the power band features from the EEG signals, which achieved an accuracy of 75% in three-level stress classification. Table 1 compares the related work for stress classification using physiological signals with mental/academic tasks.

Table 1. Comparison of work for stress classification using physiological signals with mental/academic tasks

Study	Dataset/Task	Biosignals	Classifiers	Accuracy
Blanco *et al.* [22]	Stroop Test	EEG	QDA, KNN	78.70%
Airij *et al.* [24]	Arithmetic Task	EDA,SKT,HR	Fuzzy Logic KNN	96.19%
Minguillon *et al.* [25]	MIST	EEG	LDA	50%
Xia *et al.* [26]	MIST	EEG	SVM	86%
Al-Shargie *et al.* [27]	MIST	EEG	SVM	94%
Jun and Smitha [28]	Stroop Test & MIST	EEG	SVM	75%
Our Study	MIST	EDA,SK,HR	Random Forest	99.51%
Our Study	MIST	EEG	Random Forest, KNN	99.98%

Contribution: these studies motivate us to build intelligent models to evaluate and predict stress levels in undergraduate engineering students using multimodal physiological sensors using the MIST task [20]. The contributions of this paper are as follows:

- Collection of a dataset of 23 students, recording physiological signals from Emotive EPOC+ for EEG and E4 Empatica wristband for EDA, SKT, and HR when performing MIST task [20]. The participants' stress levels can be categorized into three classes-rest (no stress), moderate stress, and high stress.
- Machine learning models that extract and select the dataset's optimal features. It trains them for mental stress classification using Random Forest and k-nearest neighbors on this dataset.
- Proposal of a stress classification model using the EDA, HR, and SKT features. The model achieves state-of-the-art accuracy of 99.51% in classifying stress into three levels.
- Proposal of a stress classification model achieving 99.98% accuracy using Time-Frequency Domain Features of the EEG data, higher than all previous works in stress classification using EEG.

Roadmap: the rest of the paper comprises four sections. Section 2 discusses the data collection process used in this paper and the proposed methodology to analyze and process the collected data. Section 3 presents the details of the experimentation techniques. Section 4 describes the results of this work and a discussion. This is followed by section 5, which presents the conclusion and future work.

2. METHODOLOGY

2.1. Data acquisition

2.1.1. Participants

This study comprises 23 subjects (20 males and 3 females) between 17 and 25 years. The chosen participants are university students with a minimum formal education of 14 years. All participants are physically and psychologically fit, without any emotional or cognitive disability or any verbal or non-verbal learning difficulties.

2.1.2. Equipment

Emotiv EPOC+ Headset: the study records the brain's EEG signals using the EMOTIV EPOC+ device, a portable, high-resolution, 14-channel EEG system. The electrode's placement is according to the 10-20 International system covering the Anterior Frontal, Frontal Central, Temporal, Parietal, and Occipital regions of the brain. The EEG device uses 14 channels to record raw EEG signals, namely; AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. The positions marked in red represent the placement of the 14 electrodes of the Emotive EPOC+ Device. It records the raw EEG signals from each channel at a sampling frequency of 128 Hz.

E4 Empatica Wristband: the E4 Empatica wristband is a wireless wristband designed to collect real-time data. The E4 features a photoplethysmography (PPG) sensor, an EDA sensor, and an infrared thermopile. The PPG sensor provides a blood volume pulse at a sampling frequency of 1 Hz, which helps obtain the heart rate, heart rate variability, and other cardiovascular signals. The electrodermal activity (EDA) sensor measures skin conductance levels at a rate of 4 Hz. The infrared thermopile measures the skin temperature at a sampling frequency of 4 Hz.

2.1.3. Procedure

The study uses MIST [20] to elicit a short-term stress response in the participant. MIST comprises several mental arithmetic questions with single-digit answers and varying difficulty levels. The standard MIST has three phases: rest, control, and experimental. Participants look at a blank screen with no tasks shown in the rest phase. The control phase presents a series of brief mental arithmetic tasks, including addition, multiplication, subtraction, and division, with various degrees of complexity. The participants need to answer all questions within a given time limit. Each test phase is highly adaptive, i.e., each question's difficulty level increases as the participants correctly answer more questions. The MIST protocol times the control phase but does not provide any feedback to the participant. Further, it instructs the participants to answer the maximum number of questions correctly and monitors their performance, thereby introducing an element of stress in the control phase. The experimental phase presents the same arithmetic problems with the same difficulty levels and a social threat component. Each participant's score is compared with the rest of the participants' average scores and displayed at the top of the screen every two minutes. It instructs all participants to get at least a minimum percentage equal to the global average to qualify. It also times each question and adaptively reduces the time for each question as the participants correctly answer more questions. As a result, this pushes participants to perform to their best ability.

Compared to the standard MIST, this study has two control phases, one experimental phase, and two relax phases. The study incorporates a training phase at the beginning to familiarize the participants with the test's keys and controls. Besides, there are small rest phases between consecutive control and experimental phases. The test takes a total of 45 minutes to complete.

The study labels the signals recorded during the two relax phases at the beginning and the end of the test, along with the two small rest phases in between, as the rest state or a state of no stress of the participant. It labels the two control phases as the state of moderate stress. The experimental phase, which induces more stress in the participants than the control phase, is labeled as a state of high stress. Figure 1 shows all the test phases and their duration.

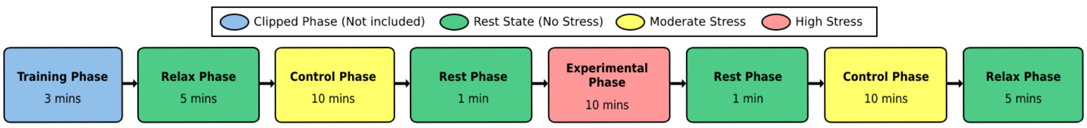

Figure 1. Overview of the test phases

One day prior to taking the test, the participants are informed not to consume any substances containing caffeine since it can alter the EEG signals of the delta and theta frequency bands. Participants are also advised not to use hair products that might interfere with the functioning of the headset. Before the test, participants are asked to consent to strictly voluntary participation. After setting up the devices, the participants are advised to minimize body movement to ensure EEG signals are free of interference-inducing artifacts. It also ensures that the participants are not disturbed until the completion of the test. Once the test is over, the participants are advised to use techniques like meditation, music, or yoga to help them relax and revert to a normal state.

2.2. Data processing

The study collects two sets of physiological signals from a subject, EEG signals and Non-EEG signals, i.e., EDA, HR, and SKT, and we apply separate processing steps to these sets of signals. Figure 2 illustrates the methodologies for building stress classification using EEG and non-EEG signals. The upcoming sections explain the data processing steps for these signals in detail.

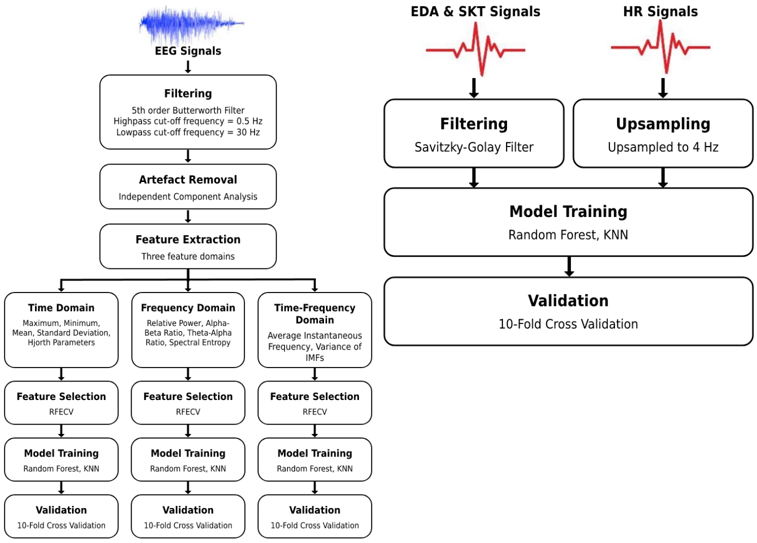

Figure 2. Methodology used for building stress classification models

2.2.1. Filtering

EEG signals are electrical signals often vulnerable to unwanted external noise called artifacts, such as power line noise, muscular movements, eye movements, and faulty electrodes. Artifacts can replicate all kinds of EEG patterns, and their inclusion in automatic analysis can render the analysis of signals unacceptable [29]. The model applies a 5th-order Butterworth Filter with 0.5 Hz and 30 Hz cutoff frequencies to remove the unwanted frequencies from the EEG signals. A Butterworth filter rejects the unwanted frequencies and keeps the required signals' frequency response as flat as possible. The model further applies independent component analysis (ICA) to the EEG signals to eliminate artifacts due to eye blinks and muscle movements.

HR signals are upsampled from 1 Hz to 4 Hz to maintain uniformity with EDA and SKT signals since EDA and SKT signals are recorded at 4 Hz. The study uses the Savitzky-Golay Filter [30] to smoothen the EDA and SKT signals, which also increases the signals' precision without altering their tendency. Since E4 Empatica records heart rate with high precision, it does not perform any filtering for the HR signals. The classification models use the original signals obtained from the device for training.

2.2.2. Feature extraction

The model extracts the EEG features in three domains for processing the EEG signals: Time Domain, Frequency Domain, and Time-Frequency Domain. It does not extract any features from the Non-EEG signals and uses them directly to train the machine learning models. The features extracted from the three domains of the EEG data are explained as follows.

a) Time domain: previous works have used simple time domain features that include characteristics of a signal in the time domain [31]. The proposed model calculates statistical features like minimum, maximum, mean $(\bar{y}_t)$ and standard deviation (σ) given by (1)–(4).

$$Minimum = min(y(t)) \tag{1}$$

$$Maximum = max(y(t)) \tag{2}$$

$$\bar{y}_t = \frac{1}{t}\sum_{i=0}^{t} y(i) \tag{3}$$

$$\sigma = \sqrt{\frac{1}{t}\sum_{i=0}^{t}(y(i) - \bar{y}_t)^2} \tag{4}$$

The model also considers Hjorth parameters, which include: activity, mobility, and complexity [32]. The activity parameter reflects the EEG signal's power and variance in the time domain. The mobility parameter represents the power spectrum's mean frequency or the proportion of standard deviation. The complexity parameter indicates a change in frequency. Hjorth parameters can be described by the (5)–(7).

$$Activity = \sigma^2(y(t)) \tag{5}$$

$$Mobility = \frac{\sigma(\frac{d(y(t))}{dt})}{\sigma(y(t))} \tag{6}$$

$$Complexity = \frac{Mobility(\frac{d(y(t))}{dt})}{Mobility(y(t))} \tag{7}$$

b) Frequency domain: the model analyses signal by converting them to functions of frequency using the Fourier transform. Frequency domain features for EEG are highly correlated with mental workload activities [33], [34]. This study considers five frequency domain features-spectral entropy, alpha by beta low power ratio, alpha by beta high power ratio, theta by alpha power ratio, and relative band power. Spectral entropy is a measure of signal irregularity. Tian *et al.* [35] demonstrated the efficiency of spectral entropy for studying mental workload. The spectral entropy of a signal is given by (8).

$$SE = -\sum_{f=0.5}^{30} P(f)\log[P(f)] \tag{8}$$

Here $P(f)$ is the normalised power spectral density (PSD) of the signal, which helps discover the power distribution of the time domain EEG signals over varied frequency ranges. It provides information about the cortical activation of different parts of the brain. The normalized power distribution of the signal in the frequency domain can be treated as a probability distribution. The Shannon entropy calculated from this distribution is called the spectral entropy of the signal. Relative band power measures the power in a frequency band expressed as a ratio to the signal's total power. We rely on relative band power instead of absolute band power since a stress response corresponds to the change in actual band power relative to the total power. The relative band is defined as (9).

$$\textit{Relative Band Power} = \frac{\textit{Absolute Band Power}}{\textit{Total Power}} \tag{9}$$

Niemiec *et al.* [36] found that during mental arithmetic activities, there is a reduction in alpha activity with a corresponding rise in beta activity. Beta waves can be classified into two categories: low beta waves, often associated with active, busy, and nervous thinking, and high beta waves, which are more pronounced when an unexpected stimulus is received. Hence, this paper uses the alpha by beta high and alpha by beta low ratios as features for the model.

$$\textit{Alpha/Beta} = \frac{\textit{Alpha Band Absolute Power}}{\textit{Beta Band Absolute Power}} \tag{10}$$

Sammer *et al.* [12] showed that theta activity also increases with an increase in the cognitive workload. Hence, Theta by Alpha Ratio is expected to increase during stress.

$$\textit{Theta/Alpha} = \frac{\textit{Theta Band Absolute Power}}{\textit{Alpha Band Absolute Power}} \tag{11}$$

c) Time-frequency domain: the time-frequency domain analyses the signals simultaneously in both time domain and frequency domain. The model computes the time-frequency domain features using Hilbert Huang transform (HHT), which is a two-step process involving empirical mode decomposition (EMD) of a raw EEG signal, followed by Hilbert spectral analysis (HSA) for feature extraction.
- Empirical mode decomposition (EMD): it is a method of breaking down an EEG signal into distinct components known as IMFs. The decomposition of a signal $y(t)$ into its IMFs is as:

$$y(t) = \sum_{i=1}^{n} x_i + r_n \tag{12}$$

where x_i is the IMF, n is the number of IMFs and r_n is the residue after decomposition. We consider only the first four IMFs of an EEG signal.

- Hilbert spectral analysis (HSA): HSA determines the instantaneous frequency of IMFs. EMD and HSA, when applied together, represent the signal's amplitude in the time-frequency domain. Obtained instantaneous frequency $w(t)$ represents the rate of change of phase and is described as:

$$w(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} \tag{13}$$

$$\theta(t) = \arctan(\frac{Hx_i(t)}{x(t)}) \tag{14}$$

where $H[.]$ is the Hilbert transform, $\theta_i(t)$ is the phase, and $x_i(t)$ is the IMF of the raw EEG signal. We extract two time-frequency features: average instantaneous frequency and the variance of the IMFs.

2.2.3. Feature selection

Data often have unnecessary and redundant attributes, which do not contribute to a predictive model's accuracy. At times, these attributes can potentially even reduce the model's accuracy. Feature selection identifies and removes these attributes. By reducing the dimensionality of the dataset, feature selection makes computation faster and increases the overall accuracy of the learning models. With 14 EEG channels and each channel capable of producing many Time Domain, Frequency Domain, and Time-Frequency Domain features, it becomes essential that we extract only the relevant features. For this purpose, we use the recursive feature elimination with cross-validation (RFECV), a technique introduced by Guyon *et al.* [37]. RFECV is a method that selects an optimal subset of features for the model by performing multiple cross-validations and removing 0 to N features in every cross-validation. The set of features that obtains the highest cross-validation score is selected. Although this approach provides high-performance features, it can be very costly in computation.

3. EXPERIMENTATION

For each participant, the study records 720 seconds of data during the rest phase, 1,200 seconds of data during the control phase, and 600 seconds during the experimental phase. It considers only 600 seconds of data for each test phase to reduce any class imbalance. It equalizes the number of labels for the three classes. The model labels all the data points as rest (0), moderate stress (1), and high stress (2) based on the phase of the test. This work implements two machine learning classifiers, Random Forest and k-nearest neighbors, using the scikit-learn module of Python 3 on local computers. It uses the 10-fold cross-validation approach for training and evaluating the classifiers to ensure an unbiased comparison of models and GridSearchCV for hyperparameter optimization for the classifiers. Table 2 shows the different configurations of hyperparameters for the classifiers.

Table 2. Range of hyper-parameters for different classifiers

Classifier	Hyper-parameters
Random Forest	criterion:['entropy', 'gini']
max-depth:[None, 5, 10, 15, 20]	
KNN	n-neighbours: [1, 3, 5, 7, 9, 11]
distance: ['euclidean', 'manhattan', 'minkowski']	

4. RESULTS AND DISCUSSION

The analysis of EDA, HR and SKT signals finds the trends of the signals during the stress state of a participant to be consistent with previous studies [14], [15], [17]–[19]. Heart rate increases sharply during the experimental phase of the test when the participant is under stress. It dips sharply when the participant is brought back to the relaxation phase. Figure 3 illustrates the impact of stress on heart rate. It increases during the experimental phase and dips during the rest and relax phases. The bottom graph shows the average heart rate reading during that phase.

The electrodermal activity (EDA) uniformly increases when the participant enters the stress phases from the relax phases of the test, as shown in Figure 4. The electrodermal activity (EDA) shows a uniform increase throughout the test. The graph in Figure 4 shows the average electrodermal activity (EDA) during that period.

The participant's skin temperature increases during the test's control and experimental phases (refer Figure 5). The skin temperature uniformly increases as the participant goes through various control and experimental phases. The graph in Figure 5 represents the average skin temperature during that phase. Some participants even show a uniform decrease in skin temperature at the end of the test during the relaxation phase.

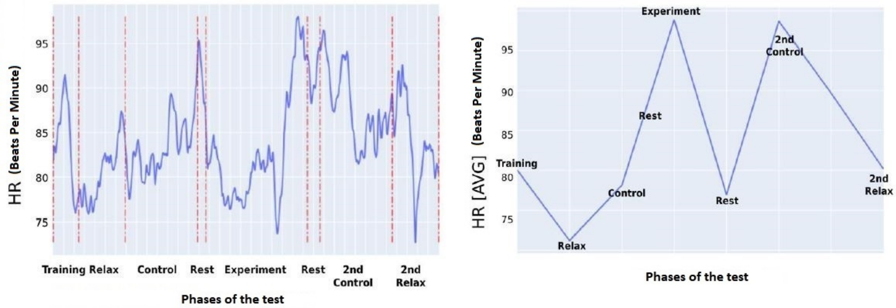

Figure 3. Impact of stress on heart rate

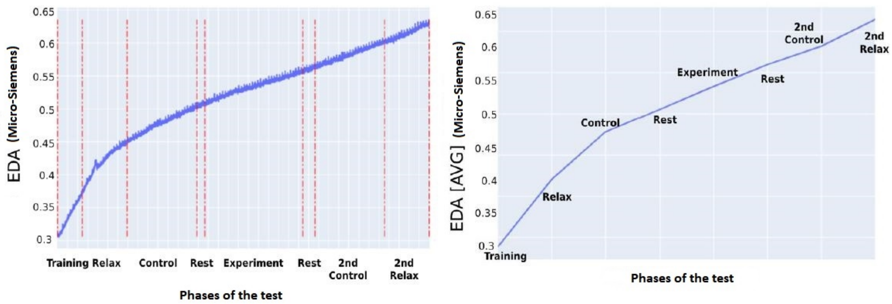

Figure 4. Impact of stress on electrodermal activity (EDA)

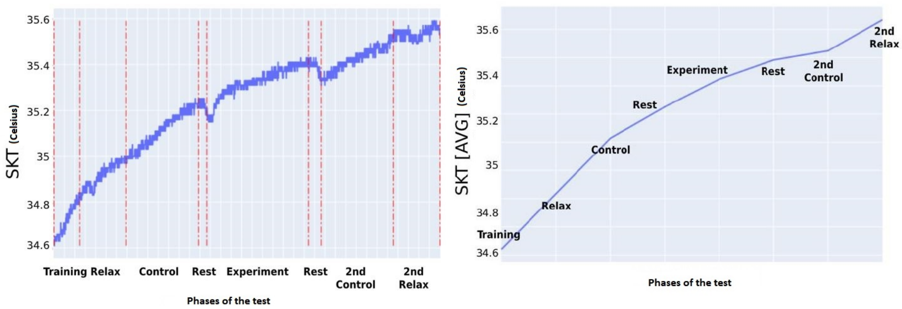

Figure 5. Impact of stress on skin temperature

Meanwhile, from the analysis of EEG signals, we observe the highest theta by alpha ratio value during the experimental phase, followed by the control phase, and the least Theta by Alpha ratio values during the rest and relax phase (refer Figure 6). The model implements Random Forest and k-nearest neighbors for stress classification and compares their performance using two performance metrics; accuracy and area under the receiver operating characteristic curve (AUC-ROC or AUROC). Accuracy is defined as (15).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{15}$$

Here, the paper denotes true positive, false positive, false negative, and true negative as TP, FP, FN, and TN, respectively, and computes them from the confusion matrix of the classifications.

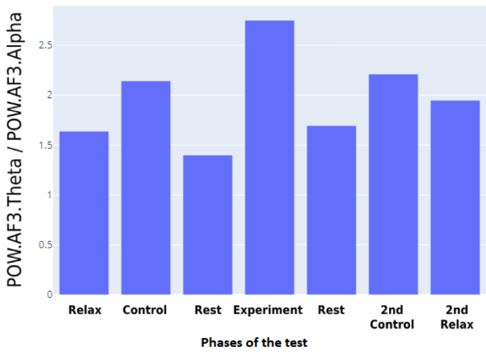

Figure 6. Average theta by alpha ratio of participants using AF3 (dominant) channel

The model also calculates the AUC-ROC score for evaluating the classifiers. The area under the ROC curve represents the classifier's ability to distinguish one class from the rest. It ranges from 0 to 1. An AUC-ROC value of 1 means the classifier can distinguish between the classes. An AUC-ROC value of 0.5 indicates that the classification is random. An AUC-ROC value of 0 signifies that the classifier's performance is the worst, labeling one class as another. Since it is a multi-class classification task, we implement a *one vs. rest* AUC-ROC score. The model computes the three AUC-ROC scores for each class for each classifier. It takes the mean of the three scores as the final AUC-ROC score for the classifier. The results indicate that EDA, HR, and SKT signals are as reliable as EEG for stress classification. Table 3 presents the accuracy and the AUC-ROC scores achieved by both the classifiers on the four feature domains.

Table 3. Results of classifiers for different feature sets

	TD		FD		TFD		EHS	
Class.	Acc.	AR	Acc.	AR	Acc.	AR	Acc.	AR
RF	61.5068%	0.7996	88.2496%	0.9747	99.9803%	0.9999	99.5180%	0.9998
KNN	66.3204%	0.7451	97.3505%	0.9799	99.9753%	0.9998	99.0599%	0.9929

TD1: time domain, FD: frequency domain, TFD: time-frequency domain, EHS: EDA, HR, SKT, AR: AUC-ROC, Acc.: accuracy, RF: Random Forest

A careful evaluation reveals that KNN performs best with N=1 and Minkowski distance as its hyperparameters. Random Forest achieves optimal results for the criterion set to 'gini', and max-depth is set to 'None'. The KNN classifier achieves an accuracy score of 66.32% and 97.35% in time and frequency domain, respectively, outperforming Random Forest in both domains in terms of accuracy. Random Forest only performs better than KNN in terms of AUC-ROC score in the Time Domain. In the time-frequency domain, both the classifiers achieve their best scores, with the Random Forest classifier giving slightly better scores than the KNN classifier. Random Forest achieves the best classification accuracy of 99.98% and an AUC-ROC score of 0.99 in all the domains. KNN achieves the best accuracy of 99.97% and an AUC-ROC score of 0.99 in all the domains. The results for this work are higher than the previous stress classification studies using EEG. This study demonstrates that time-frequency domain features are highly successful in stress classification tasks. The classifiers are also trained on the EDA, HR, and SKT feature pools. This pool attains better results than both time and frequency domain features and achieves very close results to the time-frequency domain features. With this pool, the Random Forest classifier achieves an accuracy of 99.51% and an AUC-ROC score of 0.99, which is higher than [24] and better than all the current works on EEG-based stress classification models.

Conclusively, in terms of domains, the highest accuracy, and AUC-ROC score are obtained in the time-frequency domain using EEG data. The classifiers trained on EDA, HR, and SKT feature pools also outperform classifiers trained on time and frequency domain features. The overall performance of the classifiers increases in the order, time domain $<$ frequency domain $<$ (EDA, HR, SKT) $<$ time-frequency domain.

5. CONCLUSION AND FUTURE WORK

This paper proposes machine learning models for predicting stress levels in undergraduate engineering students while performing the MIST using different wearable physiological biomarkers. The dataset collected in this study comprises EEG, EDA, HR, and SKT signals from 23 participants. Random Forest achieves a classification accuracy of 99.98% and an AUC-ROC score of 0.99, whereas the KNN classifier achieves a classification accuracy of 99.97% and an AUC-ROC score of 0.99. The results using the time-frequency domain features of the EEG data are better as compared to the previous studies. Time-frequency domain features highly correlate with the stress response, and models trained on these features can attain near-perfect accuracy. The classifiers also perform much better on the EDA, HR, and SKT feature pool (non-EEG features) than the features extracted in the time domain and the frequency domain of the EEG data. Using the non-EEG features, Random Forest achieves the best accuracy of 99.51%, and KNN achieves an accuracy of 99.05%. The models trained on EDA, HR, and SKT feature pools achieve state-of-the-art accuracy. Results of stress classification using EDA, HR, and SKT signals are competent to the classification results using only EEG signals. Measuring and recording these signals using cheap, portable, wearable sensor devices with high accuracy is also more manageable. In the future, we aim to build and collect an even more extensive database and use deep learning and federated learning models for real-time prediction with data privacy.

ACKNOWLEDGEMENTS

We acknowledge Dr. Vibha Sharma's contribution, Associate Professor of Clinical Psychology, Institute of Human Behaviour and Allied Sciences (IHBAS), Delhi, for her expert knowledge and help in understanding the clinical aspects of EEG and cognition. We thank Ankit Priyarup for his ideas for the project and Samsung India Electronics Private Limited for providing the Samsung Innovation Campus at Delhi Technological University for pursuing this work.

REFERENCES

[1] H. Selye, "The stress syndrome," *The American Journal of Nursing*, pp. 97–99, 1965, doi: 10.2307/3453119.

[2] S. Cohen, D. Janicki-Deverts, and G. E. Miller, "Psychological stress and disease," *Jama*, vol. 298, no. 14, pp. 1685–1687, 2007, doi: 10.1001/jama.298.14.1685.

[3] M. Nandi, A. Hazra, S. Sarkar, R. Mondal, and M. K. Ghosal, "Stress and its risk factors in medical students: An observational study from a medical college in India," *Indian Journal of Medical Sciences*, vol. 66, no. 1-2, pp. 1–12, 2012, doi: 10.4103/0019-5359.110850.

[4] S. Behere, R. Yadav, and P. B. Behere, "A Comparative Study of Stress Among Students of Medicine, Engineering, and Nursing," *Indian journal of psychological medicine*, vol. 33, no. 2, pp. 145–148, 2011, doi: 10.4103/0253-7176.92064.

[5] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of health and social behavior*, pp. 385–396, 1983, doi: 10.2307/2136404.

[6] F. Al-Shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, "Towards multilevel mental stress assessment using SVM with ECOC: an EEG approach," *Medical & Biological Engineering & Computing*, vol. 56, no. 1, pp. 125–136, 2018, doi: 10.1007/s11517-017-1733-8.

[7] T. Ogawa *et al.*, "Brain-machine interfaces for assistive smart homes: A feasibility study with wearable near-infrared spectroscopy," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 1107–1110, doi: 10.1109/EMBC.2015.7318559.

[8] A. Gardner, A. M. Krieger, G. Vachtsevanos, and B. Litt, "One-class Novelty Detection for Seizure Analysis from Intracranial EEG," *Journal of Machine Learning Research*, vol. 7, no. 6, 2006.

[9] E. M. Elidrissi, E. Essoukaki, L. B. Taleb, A. Mouhsen, and M. Harmouchi, "A new hybrid and optimized algorithm for drivers' drowsiness detection," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 3, pp. 1101–1107, 2022, doi: 10.11591/ijai.v11.i3.pp1101-1107.

[10] N. A. Ali, A. R. Syafeeza, A. S. Jaafar, and M. K. M. F. Alif, "Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 1, pp. 91–99, 2020, doi: 10.11591/ijai.v9.i1.pp91-99.

[11] J. T. Cacioppo, G. G. Berntson, J. T. Larsen, K. M. Poehlmann, and T. A. Ito, "The psychophysiology of emotion," *Handbook of emotions*, vol. 2, pp. 173–191, 2000, doi: 10.1016/B978-0-08-100508-8.00004-7.

[12] G. Sammer *et al.*, "Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload," *Human brain mapping*, vol. 28, no. 8, pp. 793–803, 2007, doi: 10.1002/hbm.20309.

[13] P. Ren, A. Baretto, Y. Gao, and M. Adjouadi, "Affective assessment by digital processing of the pupil diameter," *IEEE Transactions on Affective computing*, vol. 4, no. 1, pp. 2–14, 2012, doi: 10.1109/T-AFFC.2012.25.

[14] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005, doi: 10.1109/TITS.2005.848368.

[15] C. H. Vinkers *et al.*, "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, pp. 520–530, 2013, doi: 10.3109/10253890.2013.807243.

[16] M. A. Birkett, "The Trier Social Stress Test protocol for inducing psychological stress," *Journal of visualized experiments: JoVE*, no. 56, 2011, doi: 10.3791/3238.

[17] G. Giannakakis *et al.*, "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control*, vol. 31, pp. 89–101, 2017, doi: 10.1016/j.bspc.2016.06.020.

[18] V. Engert, A. Merla, J. A. Grant, D. Cardone, A. Tusche, and T. Singer, "Exploring the use of thermal infrared imaging in human stress research," *PloS one*, vol. 9, no. 3, p. e90782, 2014, doi: 10.1371/journal.pone.0090782.

[19] U. Lundberg *et al.*, "Psychophysiological stress and EMG activity of the trapezius muscle," *International journal of behavioral medicine*, vol. 1, no. 4, pp. 354–370, 1994, doi: 10.1207/s15327558ijbm0104_5.
[20] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, "The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain," *Journal of Psychiatry and Neuroscience*, vol. 30, no. 5, pp. 319–325, 2005.
[21] R. Sharma and K. Chopra, "EEG signal analysis and detection of stress using classification techniques," *Journal of Information and Optimization Sciences*, vol. 41, no. 1, pp. 229–238, 2020, doi: 10.1080/02522667.2020.1714187.
[22] J. A. Blanco A. C. Vanleer, T. K. Calibo, and S. L. Firebaugh, "Single-trial cognitive stress classification using portable wireless electroencephalography," *Sensors*, vol. 19, no. 3, p. 499, 2019, doi: 10.3390/s19030499.
[23] F. Scarpina and S. Tagini, "The stroop color and word test," *Frontiers in psychology*, vol. 8, p. 557, 2017, doi: 10.3389/fpsyg.2017.00557.
[24] A. G. Airij, R. Sudirman, and U. U. Sheikh, "GSM and GPS based real-time remote physiological signals monitoring and stress levels classification," in *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, 2018, pp. 130–135, doi: 10.1109/ICBAPS.2018.8527406.
[25] J. Minguillon, E. Perez, M. A. Lopez-Gordo, F. Pelayo, and M. J. Sanchez-Carrion, "Portable system for real-time detection of stress level," *Sensors*, vol. 18, no. 8, p. 2504, 2018, doi: 10.3390/s18082504.
[26] L. Xia, A. S. Malik, and A. R. Subhani, "A physiological signal-based method for early mental-stress detection," *Biomedical Signal Processing and Control*, vol. 46, pp. 18–32, 2018, doi: 10.1016/j.bspc.2018.06.004.
[27] F. M. Al-Shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, "Mental stress quantification using EEG signals," in *International Conference for Innovation in Biomedical Engineering and Life Sciences*, 2015, pp. 15–19, doi: 10.1007/978-981-10-0266-3_4.
[28] G. Jun and K. G. Smitha, "EEG based stress level identification," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 003270–003274, doi: 10.1109/SMC.2016.7844738.
[29] A. Tandle, N. Jog, P. D'cunha, and M. Chheta, "Classification of Artefacts in EEG Signal Recordings and EOG Artefact Removal using EOG Subtraction," *Communications on Applied Electronics (CAE)*, vol. 4, no. 1, pp. 12–19, 2016.
[30] A. Savitzky and M. J. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964, doi: 10.1021/ac60214a047.
[31] S. M. R. Islam, A. Sajol, X. Huang, and K. L. Ou, "Feature extraction and classification of EEG signal for different brain control machine," in *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2016, pp. 1–6, doi: 10.1109/CEEICT.2016.7873150.
[32] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970, doi: 10.1016/0013-4694(70)90143-4.
[33] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine Learning Framework for the Detection of Mental Stress at Multiple Levels," *IEEE Access*, vol. 5, pp. 13545–13556, 2017, doi: 10.1109/ACCESS.2017.2723622.
[34] P. Lahane, A. Vaidya, C. Umale, S. Shirude, and A. Raut, "Real Time System to Detect Human Stress using EEG Signals," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 4, no. 4, 2016.
[35] Y. Tian *et al.*, "Spectral Entropy Can Predict Changes of Working Memory Performance Reduced by Short-Time Training in the Delayed-Match-to-Sample Task," *Frontiers in human neuroscience*, vol. 11, p. 437, 2017, doi: 10.3389/fnhum.2017.00437.
[36] A. J. Niemiec and B. J. Lithgow, "Alpha-band characteristics in EEG spectrum indicate reliability of frontal brain asymmetry measures in diagnosis of depression," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 7517–7520, doi: 10.1109/IEMBS.2005.1616251.
[37] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik , "Gene Selection for Cancer Classification using Support Vector Machines," *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002, doi: 10.1023/A:1012487302797.

BIOGRAPHIES OF AUTHORS

Varun Chandra obtained his B.Tech. degree in Software Engineering from Delhi Technological University in 2021. He has previously worked on EEG signal analysis and stress detection using physiological biomarkers. His research interests include EEG signal analysis, machine learning, and reinforcement learning. He is currently working as a Member of Technical Staff at Adobe, India. He can be contacted at email: varun.chandra214@gmail.com.

Divyashikha Sethia received B.Tech. degree in Computer Science and Engineering from the Maharaja Sayajirao University of Baroda in 1997 and M.Tech. degree in Computer Science and Engineering from the Indian Institute of Technology (IIT) Delhi in 2006. She is an Assistant Professor in the Department of Computer Science and Engineering at Delhi Technological University (DTU), Delhi. Her research interests are distributed systems, mobile computing, telemedicine, cognitive science, artificial intelligence, and cryptography. She is an Assistant Professor in the Department of Software Engineering at Delhi Technological University, India. She can be contacted at email: divyashikha@dtu.ac.in or sethiadivya@gmail.com.