The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 5, October 2025
https://doi.org/10.33022/ijcs.v14i5.5022

Multi-Level Stress Classification Using the Electroencephalogram Based on Mental Load Tasks

Walaa Alajali[1]
walaakhshlan@utq.edu.iq[1]
[1]College of Education for Pure Science, University of Thi-Qar, Nasiriyah, Iraq,64001.

Article Information

Received : 13 Oct 2025
Revised : 22 Oct 2025
Accepted : 23 Oct 2025

Keywords

Stress, Classification, Electroencephalogram

Abstract

Stress is considered one of the major global health issues contributing to cardiovascular disease and depression among other disorders. This study examines the amounts of stress and performance on tasks using electroencephalogram (EEG) data and machine learning. Raw EEG data is preprocessed to remove noise and segment epochs. Empirical Mode Decomposition (EMD) is followed by Butterfly Optimization Algorithm (BOA) for feature extraction and dimensionality reduction. Five machine learning classifiers (SVM, Naive Bayes, Random Forest, KNN, Decision Tree) classify four levels of stress (neutral, low, medium, and high) based on cognitive load during two tasks: the Stroop color-word task and an arithmetic task. Results indicate the Naive Bayes classifiers for the Stroop and arithmetic tasks had accuracies of 98.82% and 98.87% respectively, while the SVM classifier achieved 99.02% accuracy for both tasks combined. Such results attest to the growing interest and application of machine learning on EEG data for mental health monitoring and the possible enhancement of task performance.

A. Introduction

Stress is psychological and physiological response to challenging or threatening situations. It is a typical human reaction that enables individuals to confront challenges and dangers in daily life[1]. However, a prolonged or excessive stress can adversely affect health, leading to elevated blood pressure sleep disturbances, an increased risk of heart attacks, and other cardiovascular disorders [2]. Furthermore, research has shown that mental stress affects particular cognitive functions like working memory, attention, and decision-making. Cognitive deficiencies are greatly exacerbated by scholastic pressure and workplace stress, underscoring the necessity of specialized therapies in these types of environments. The effects of this stress manifest in a range of behavioral symptoms, such as a diminished capacity for adaptation and a slowed rate of information processing[3].

In terms of biology, stress involves intricate interactions between the brain, nervous system, and several hormones. Stress is the body's physiological and psychological reaction to mental, emotional, or physical stress. Such a shift in the surroundings triggers a series of physiological reactions (stress response) in the body and brain[4], [5]. This reaction, sometimes known as the "fight-or-flight" response, is a component of the body's innate adaptation mechanism, which enables individuals to handle potentially hazardous situations[6].

Stress comes in two primary forms: acute stress and chronic stress. Acute stress is defined as temporary tension that passes rapidly. Long-term, persistent stress is referred to as chronic stress. Chronic stress can become so commonplace that people fail to recognise its negative effects, which might result in further health issues[7].

There are several ways to determine someone's mental stress level. Self-report measures are among the most widely used techniques, in which participants use standardized questionnaires to indicate their level of stress. However, self-report measures may not accurately reflect the true amount of stress because they are subjective[8]. Physiological measures are an additional method of measuring mental stress. Through a variety of physiological markers, including Galvanic Skin Responses (GSR), Heart Rate Variability (HRV), Electroencephalography (EEG), Photoplethysmography (PPG), Electrocardiograms (ECG), and others, this entails tracking the body's reaction to stress[9].

Artificial intelligence and machine learning methods have also been used in recent years to identify and quantify mental stress. In order to detect indicators of stress, these methods use algorithms to examine behavioral patterns and physiological markers. In general, assessing mental stress can be a multidimensional and complex process, and the most accurate way to determine a person's stress level may include combining many techniques[10]. Our study will use electroencephalograms (EEGs), one of the most widely used methods for assessing stress. With a high temporal resolution, they provide an instantaneous millisecond reflection of the brain's electrical activity. Moreover, a significant amount of emotional processing is controlled by the prefrontal cortex. Anxiety and sadness are common bad moods brought on by stress, which lowers right prefrontal activity[11].

In this study, the effectiveness of machine learning models in classifying EEG data under various stress-inducing circumstances is assessed. In particular, the effectiveness of these models is examined when they are trained and evaluated using data gathered during a variety of cognitive tasks that cause stress, such as the Stroop color-word test and arithmetic problem-solving tasks, in situations that involve both classifying stress levels within a task and classifying stress levels across multiple tasks.

The novelty and contributions of this study can be summarized as follows:

1. This study investigates the capacity of machine learning models to classify stress levels across a range of stressful multi-task.
2. Combining EEG data from different cognitive tasks to provide more accurate data that represents a variety of stress levels.
3. The classification abilities of five traditional machine learning models support vector machines (SVM), K-nearest neighbors (KNN), and others, are methodically evaluated for various stress conditions, offering a comparative performance study.
4. Analyzing and accurately extracting information from EEG data using the empirical mode decomposition transform, which takes into account the nonlinear nature of neural signals.
5. Utilizing the Butterfly Optimizer technique to increase model efficiency and decrease dimensionality.

The following is the order of the article's following sections: The literature on stress detection using EEG signals is reviewed in Section 2, while the suggested model and its framework are covered in Section 3. The experimental results and discussion are presented in Section 4, and Section 5 concludes the study and outlines its future directions.

B. Related Work

Several previous studies have used machine and deep learning to classify stress levels, such as: Kumar et al.[12] propose an end-to-end multilevel classification of mental stress utilising 64-channel EEG sensors through the integration of extreme gradient boosting (XGBoost), a squeeze excitation (SE) block-enhanced convolutional neural network (CNN), and a long short-term memory (LSTM) model augmented with self-attention, known as StreXNet. The findings exhibit a classification accuracy of 91.81 ± 4.26% for four-class classification, only one session for each level was conducted in this study. There are a number of drawbacks to this study, even though it achieved classification results. the study involved university-educated adults with identical MMSE scores, it only used male volunteers to eliminate the possibility that hormone cycles could affect the stress induction process, and it only had one session for each level.

BADR et al.[13] developed a method to distinguish various mental states across diverse population samples with differing stress baselines. Employing EEG signals, graph convolutional neural networks (GCNs), and binaural beat stimulation (BBs). The research examines stress detection and relief in two population sample groups with differing baselines (group 1: low daily baseline, and group 2: high daily stress baseline). The experiment consists of four phases: resting state, control alertness, stress induction, and stress mitigation. The PDC-

GCN method attained average classification accuracies of 76.43% and 76.32% for Groups 1 and 2, respectively, and 76.37 ± 8.40% overall. These results may be prone to mistakes or personal bias due to the utilisation of a self-report scale, specifically the PSS-10 questionnaire, to assess perceived stress levels. The PSS-10 results did not consistently align with the behavioural assessments. This may be attributed to individuals' inability to appropriately assess their mental state by subjective methods, or due to certain participants expressing discomfort with the BBs post-experiment, rendering the results susceptible to inaccuracies.

KHAN et al. [14] presented the classification of perceived mental stress by EEG data. involves transforming EEG signals into spectrograms by the Short-Time Fourier Transform and optimising three pre-trained deep neural networks, namely ResNet50, EfficientNetB0, and DenseNet121. The classification was conducted for both two-level (non-stressed and stressed) and three-level (non-stressed, mildly stressed, and stressed) stress categories. Research findings demonstrate that the ResNet50 model attained accuracy 95.80% for binary classification and 86.02% for ternary classification, utilising a 90:10 train-test split ratio. Vazquez et al.[15] employed a recurrent neural network (RNN) utilising a structure based on gated recurrent units (GRU) to evaluate the feasibility of estimating the user's perceived stress level associated with a specific task through an EEG system integrated with a Serious Game featuring a multi-level stress driving tool. The RNN model achieves prediction and classification accuracy for stress levels of up to 94%. The integration of an EEG system with a serious game can quantify stress responses; however, the game design may not encompass all types of real-life or environmental stress, potentially constraining the model's practical applicability in diverse everyday scenarios.

Hemakom et al.[16] employed machine learning models that were based using ECG and EEG signals to classify stress at multiple levels and to differentiate between mental exertion and stress in mixed sexes, males, and females during various phases of the menstrual cycle. Combinations of binary classifiers (MLP, AdaBoost, kNN, LR, NB, Max-RF, Log2-Rf, Sqrt-RF, SVM, and RBF-SVM), as well as combinations of classification and voting schemes, were employed in this study. The classification accuracies attained through the utilisation of combinations of binary classifiers are 76.55% for Mixed, 73.33% for Female in the follicular phase, 81.43% for Female in the luteal phase, and 87.14% for Male.

Yang et al.[17] introduced a paradigm for recognising driving stress aimed at quantifying stress through EEG and behavioural data. Utilising the extraction of essential EEG and behavioural features alongside feature selection, low, medium, and high levels of driving stress were discerned through the SHapley Additive exPlanation (SHAP) approach, and fuzzy rules were constructed utilising the decision tree method. The results demonstrated that following feature selection, an accuracy rate of 84.93% was attained. The use of a driving simulator instead of actual driving to collect EEG signals is one limitation of the study. The psychological response of drivers may not be accurately represented in virtual environments. Also, there are a lot of variables that might affect how people rate their own stress levels, so using a self-report measure isn't the most trustworthy option. The stress level classification using this data becomes less accurate as a result. Hafeez et al. [18]sought to investigate the correlation between students'

stress levels and the decline in their following examination performance. Additionally, investigated whether certain EEG frequency bands can serve as indicators for stress levels. Utilized two distinct deep learning frameworks to categorize the data, achieving an accuracy of 70.67% with Long Short-Term Memory (LSTM) and 90.46% with Convolutional Neural Networks (CNN).
The study had some drawbacks, including a limited sample size of just 14 people and difficulties handling noise and artefacts when gathering data. As a result, certain data were excluded.

C. Research Method

In this section the research methodlogy is presented. Figure 1 shows the block diagram of the propsed method.

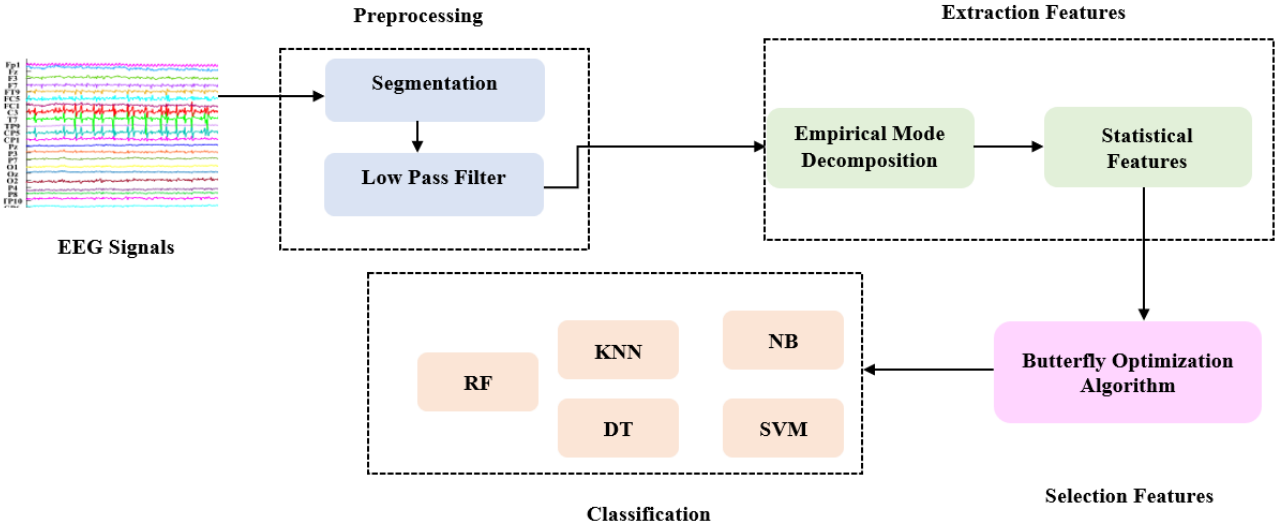

Figure 1. Proposal Method for Classification of Multi-Level Stress.

Dataset Description

An electroencephalography (EEG) dataset was used to analyze cognitive load patterns. This publicly available dataset contains EEG signals from a group of 15 of graduated students including eight females and seven males with an average age of 21.5 years [19]. Rerecording was collected while participants performed various task including the Stroop Color Work task (SEWT) and arithmetic problem-solving exercises. Participants were selected based on their health ensuring that no neurological or psychological abnormalities were detectable. This facilitated the reproduction of EEG recordings associated with cognitive stress. Every training session was conducted for all participants before data collection to ensure stability of task performance [20]. Figure 2 shows three levels of mental stress state is collected using the SEWT and math problems. The study resulting indicated better generalizability due to the diverse ethnic backgrounds of participants. To understand EEG data related to stress and cognitive load, population profile is crucial. This provides a diverse and controlled set of cognitive responses during the scan recordings for each recording were divided into four levels of induced mental stress normal law and high. Three trials were conducted to collect comprehensive data, with each task lasting between 10 and 20 seconds. The Opanci device, when used with an 8-channel Cyton board, records complex frontal brain responses to cognitive tests, such as the Stroop and Arithmetic Tests, at a rate of 250 Hz. The channels that were used include Fp1, Fp2, F7, F3, Fz, F4, F8, and C4. Figure 2 shows how the eight channels are set up, Figure 3 illustrate Placement of 8-Electrode Channels.

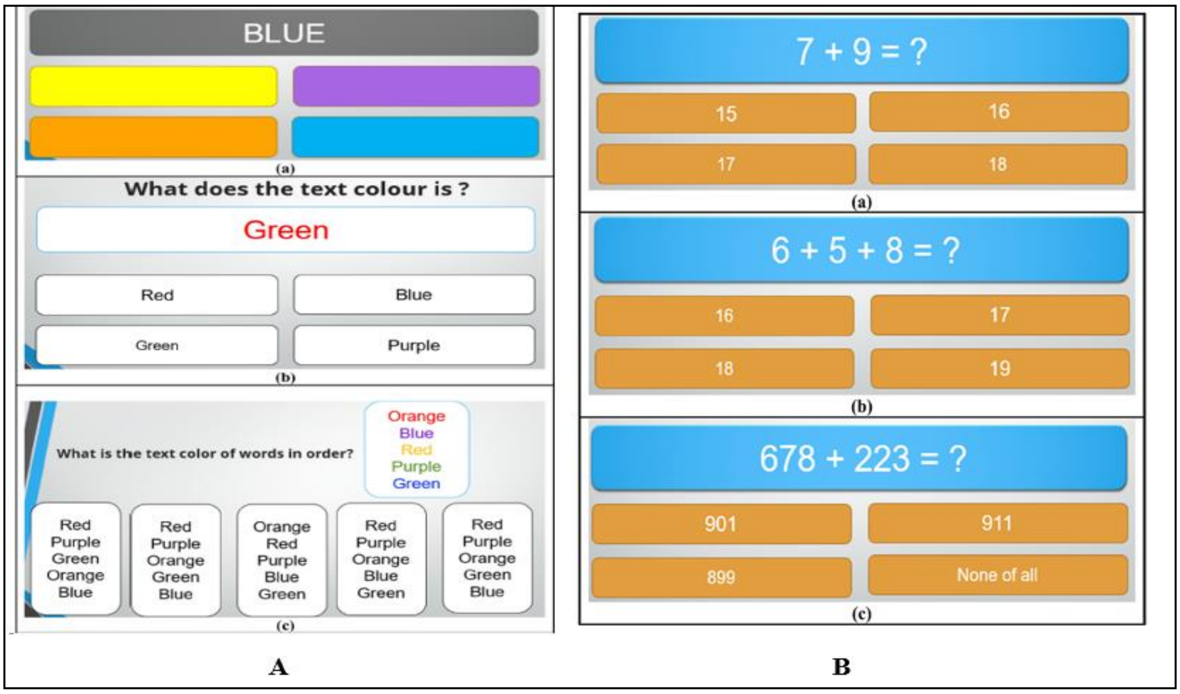

Figure 2. A- SCWT B- The Arithmetic Stimuli Used in the Proposed Study at Stress Levels at Low, Mid, and High Levels.

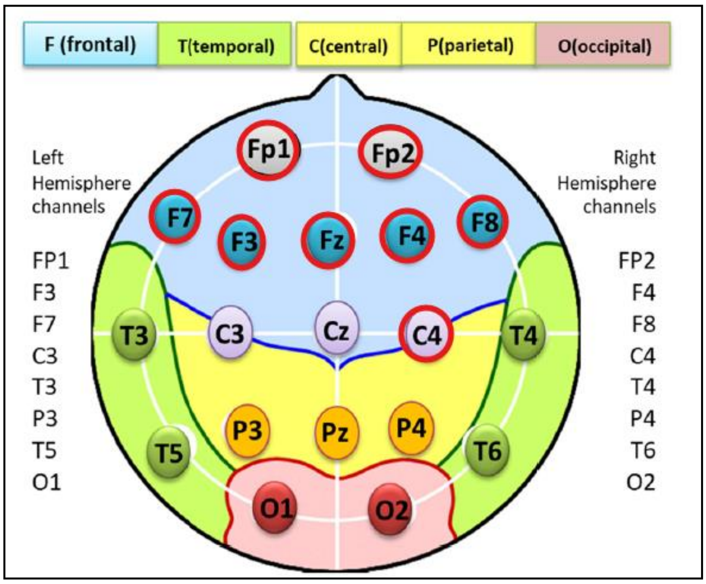

Figure 3. The Placement of 8-Electrode Channels[21].

Preprocessing

Artifacts present during EEG signal capture and transmission can lead to inaccurate interpretations among participants. To reduce artifacts, EEG signals should be filtered appropriately[22]. This study filters signals using a Butterworth fifth-order low-pass filter with a cutoff frequency of 45 Hz as shown in Figure 4.

To deal with the non-linear and non-stationary character of extended EEG signals, they must be split into short parts. Short-term segment statistics are expected to be identical[23]. After filtering, signals are divided into 2.5 seconds non-overlapping time windows (625 samples each).

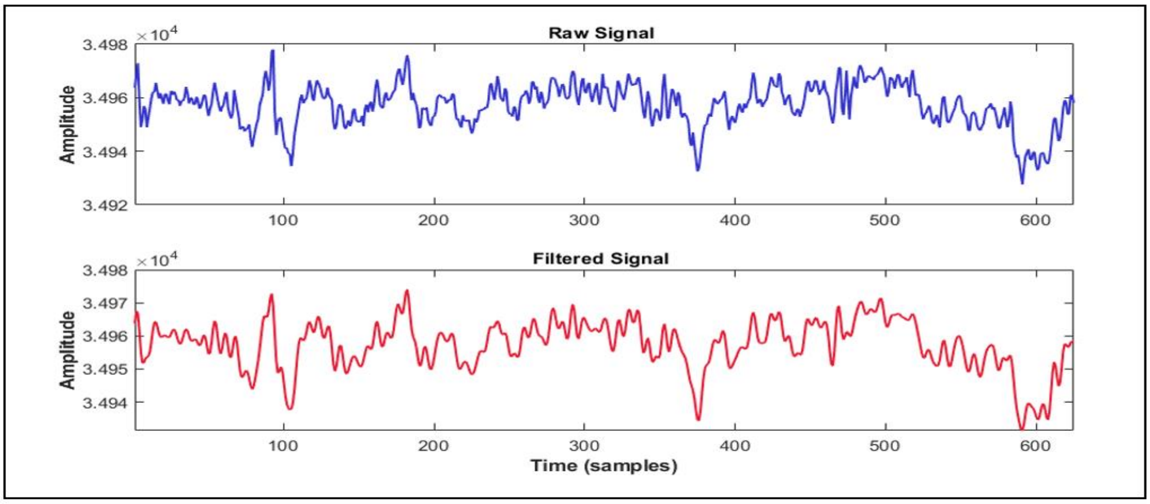

Figure 4. Signal Processing Using a Butterworth Filter(Low-Pass-Filter)

Empirical Mode Decomposition

The signal decomposition step is essential to conventional machine learning techniques because it facilitates the extraction of features required for the classification procedure. Empirical Mode Decomposition is a popular method for stress level categorization feature extraction from nonstationary signals, such EEG[24]. EMD is an adaptive technique that can be used to analyze time series signals that are nonlinear and non-stationary. The complex signal can be broken down using this method into a number of physically meaningful Intrinsic Mode Functions (IMF) and a residual component that represents the signal trend[25].

The time-frequency domain data of the signal is contained in the IMFs, which also have the following characteristics[26]:

1. There are either the same number of zero crossings of the IMFs or a difference of one between the number of extremes (maximum and minimum values).
2. The top and bottom envelope means of each IMF are zero, and the IMFs are locally symmetric.
3. IMFs have at least one maximum and one minimum value.

The following is the mathematical definition of EMD decomposition for raw signals x(t):

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + r(t)$$

Where $MiF_i(t)$ is the component of IMF and r(t) is the residual term of decomposition. The inherent mode function definition states that the signal can be repeatedly screened to separate it into a residual and a finite number of IMF components. The following is the precise screening procedure for the breakdown of the EMD algorithm[27]:

Step 1: Determine the signal's maximum and minimum values x(t);

Step 2: Using curves, the upper and lower envelope lines are connected to the maximum and minimum values of the signal x(t). All of the x(t) sampling points are contained in the two envelope lines. The average value m(t) is made up of the upper envelope $E_{max}(t)$ and the lower envelope $E_{min}(t)$. The formula for calculating m(t) is as follows:

$$m(t) = \frac{e_{max}(t) + e_{min}(t)}{2}$$

Step 3: To obtain the middle signal, F(t), subtract the signal x(t) from the mean m(t). The following is the calculating formula:

$$F(t) = x(t) + m(t)$$

Step 4: Assess if the signal F(t) satisfies the IMF's requirements. If not, step (2) is carried out again until the F(t) satisfies the component criterion, obtaining the signal x(t)'s first IMF component $F_i(t)$ and residual component r(t). The aforementioned procedures can be repeated with i = 1, 2, n to produce the remaining IMF component $F_i(t)$.

Step 5: Keep the residual component rig(t) when the signal has less than two extreme points. At that time, the decomposition process is complete. After the signal x(t) is divided into i IMF components and i residual components, it can be rebuilt using Eq. (4):

$$x'(t) = \sum_{j=}^{p} F(t) + r(t)$$

where p is the number of IMF. The EEG eigenmatrix is then rebuilt after each IMF component is positioned in each channel in accordance with the frequency scale to create multiple channels of signal.

After segmenting the EEG signal into 2.5-second segments, each segment from each channel was passed through the EMD algorithm for analysis. Four IMF components were taken from each segment, in addition to a residue as shown in Figure 5. which were then utilized for calculating statistical features. Figure 5 shows the analysis of each stress level using EMD analysis.

Feature extraction

The feature extraction stage is one of the most crucial steps in classification because it considerably improves the model's performance and accuracy. In order

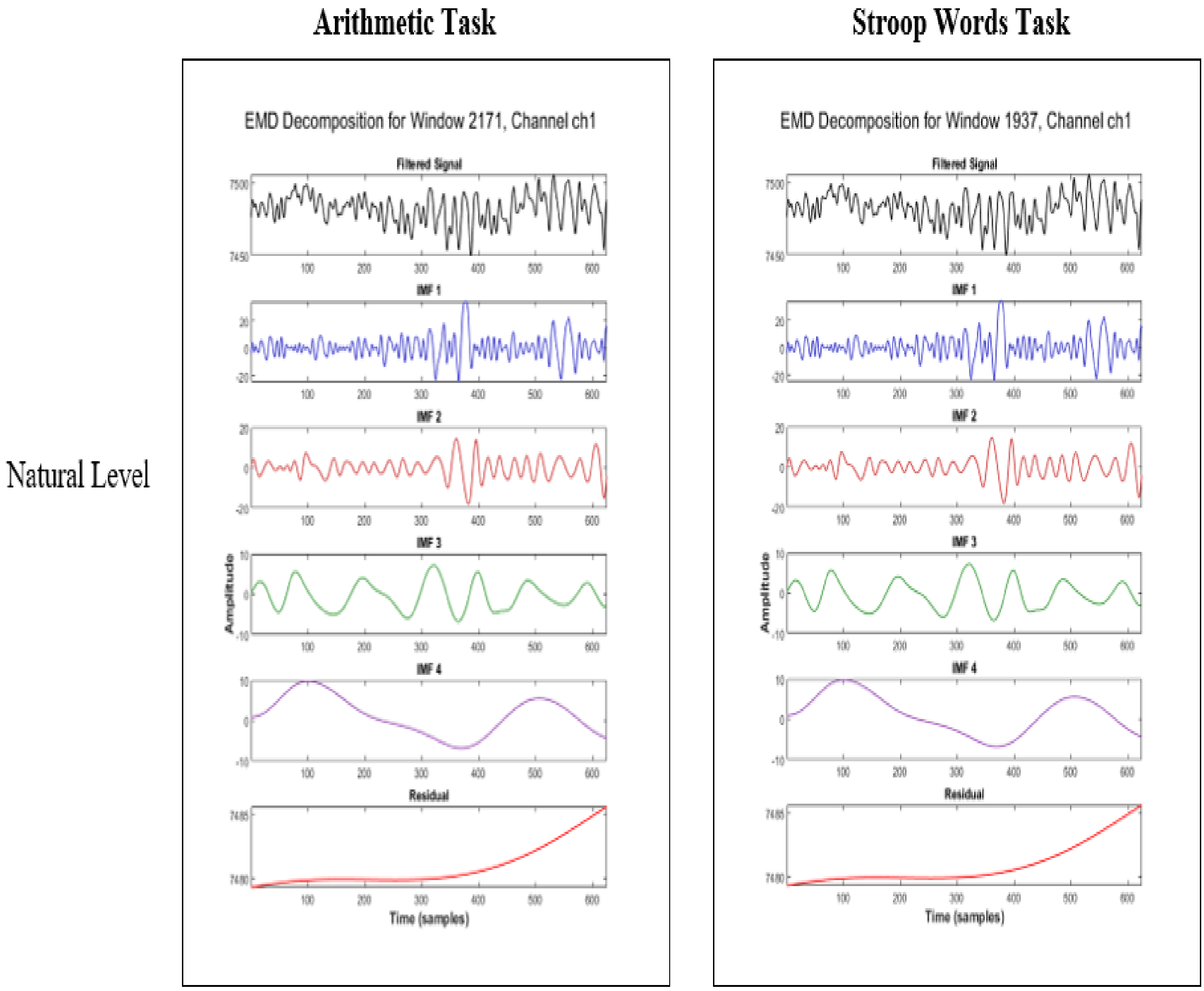

Figure .5 Analysis Using EMD For: A- Natural Level And B- for Each Stress Level(Low, Mid And High) in One-Channel.

to properly classify various categories, it is crucial to choose features that accurately reflect the nature of the data. However, reducing the number of extracted features is required to avoid issues like overfitting, decrease training and classification time, and simplify computations. This study aimed to extract a set of fundamental statistical features that assist create a clear and meaningful representation of the data, which facilitates faster and more accurate learning and decision-making by the classification model. These features consist of five primary features: mean, median, standard deviation, maximum, and minimum, which are extracted from each IMF and residue component.

- **Minimum value**

  The smallest value of a given signal is represented by the minimum value.
- **Maximum value**

A signal's peak value is referred to as its maximum value.

- **Mean**

For a given signal, the mean is the average. The mean is given by the following equation:

$$y = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where, X refers for the signal, N for the total number of samples, and Y is the mean.

- **Median**

The median is the value that lies in the middle of all the values.

- **Standard Deviation**

The standard deviation is the variance squared. The standard deviation can be found using the following formula[28]:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} |x_i - Y|}$$

where standard deviation is referred by σ.

Butterfly Optimization Algorithm

The BOA is a strong swarm-based metaheuristic that imitates the way butterflies forage for food[29]. BOA is an innovative feature selection technique that looks for and uses the most pertinent characteristics from a dataset in order to optimize the performance of machine learning models. The way that butterflies forage and mate is an inspiration for BOA, which uses chemoreceptors to detect scent and direct movement. As a function of stimulus intensity, smell is essential to BOA, helping butterflies with both local and global search operations[30]. It is possible to formulate the inspiration and movement behavior of butterflies as an optimization method in which the created fragrances indicate the fitness values and the butterflies represent the search agents. The butterflies (search agents) in BOA have the ability to produce fragrance (fitness value) that sets them apart from other scents. Other search agents may be able to adjust their location in the search space with the aid of such behavior[29].

The following is a mathematical model of the fragrance intensity:

$$fi = cI^a \qquad \text{---------} \quad (1)$$

where ***fi*** refers for the butterfly's fragrance strength, ***I*** for stimulus intensity, **c** for sensory modality, and a for power exponent based on modality, which shows a different degree of absorption.

According to the BOA algorithm, the phase known as "global search" occurs when a butterfly detects the scent of another butterfly and approaches it. In another situation, a butterfly will travel at random when it is unable to detect scents in its surroundings; this stage is known as local search[31]. Several iterations are carried out using the algorithm. All of the butterflies in the solution space transfer to new locations throughout each iteration, after which their fitness values are assessed. The fitness values of each butterfly at various locations throughout the solution space are initially determined by the algorithm. These butterflies will then use Equation (1) to produce fragrance at their locations.

The algorithm consists of two essential steps: the local search phase and the global search phase. The butterfly moves closer to the fittest butterfly (solution)g∗ during the global search phase, which is represented by Eq. (2).

$$x_i^{t+1} = x_i^t + (r^2 \times g^* - x_i^t) \times f_i \qquad \text{-----------(2)}$$

where x_i^t is the ith butterfly's solution vector $x_i$for iteration number t.The current best solution among all the solutions in the current iteration is denoted by g∗. *fi* represents the fragrance of the ith butterfly, whereas r is a random number in the range [0, 1].

The local search phase could be represented as

$$x_i^{t+1} = x_i^t + (r^2 \times x_j^t - x_k^t) \times f_i \qquad \text{------------(3)}$$

where x_j^t and x_k^t represent the *j*th and *k*th butterflies in the solution space, respectively. If r is a random number in [0, 1] and x_j^t and x_k^t are members of the same swarm, then Eq. (3) becomes a local random walk.

The fitness function uses a K-nearest neighbor classifier to assess specific feature sets, with classification accuracy serving as the main performance indicator. The fitness function indicates how well the chosen feature set enhances the model's capacity to discriminate between various stress levels in the dataset. In order to explore the feature space and choose the appropriate ones, the butterfly optimization technique was employed. A K-nearest neighbor classifier is used to assess the many feature sets that the algorithm iteratively creates in order to maximize classification accuracy. In every iteration, the algorithm keeps the best collection of features that were acquired at that stage.

This guarantees that the most effective options are tracked. To get the greatest performance, the classifier always uses a predetermined number of neighbors (K), which are chosen based on the properties of the data. The model's capacity to generalize to previously unseen data is confirmed by dividing the dataset into a training and test set using the optimal set of features produced by the algorithm at the conclusion of the optimization procedure. By balancing exploration and exploitation tactics in the feature space, this integrated strategy helps to enhance model performance, eliminate superfluous dimensions, and boost the effectiveness of the data-driven categorization system.

Butterfly Optimization Algorithm[31]

1. Objective function *f(**x**)*, where x= (x_1, x_2,..,X_{dim}), and dim = number of dimensions
2. Create ***n*** butterflies as the initial population using x_i=(i=1,2,...n)
3. *f(x_i)* determines the stimulus strength $\boldsymbol{I_i}$ at $\boldsymbol{x_i}$.
4. Define power exponent ***a***, switch probability ***p***, and sensor modality ***c***.
5. **while** the stopping requirements are not fulfilled, **do**
6. **For each** butterfly ***bf*** in the population, **do**
7. Determine the ***bf*** fragrance using Equation (1).

8. **end for**
9. Finding the best ***bf***
10. For each butterfly ***bf*** in the population, do
11. Create a random number r from the range [0, 1].
12. **if r**<**p** then
13. Use Eq. (2) to move toward the best butterfly or solution.
14. Else
15. Move at random with Eq. (3)
16. end if
17. end for
18. Update the value of ***a***
19. end while
20. Provide the best solution that was found.

Classification

The butterfly optemaizer algorithm selects features, which are then input into five ML classifiers (DT, NB, RF, KNN, and SVM) for comparison. The operation of the algorithms is covered in detail in the following.

- ***Support vector machine (SVM)***

The most often used supervised machine learning method for classification tasks is the Support Vector Machine (SVM). The SVM is producing the hyperplanes, which are used to differentiate between classes in the training data set[28]. By comparing members of one class to those of another, the best hyperplane for classification is found. In this study, the rbf kernel function for SVM.

- ***Random Forest (RF)***

RF is an ensemble learning method that generates precise predictions by employing a large number of decision trees. Each tree is constructed using an arbitrary combination of features and a chosen subset of data. Each tree iteratively divides the data into subgroups according to the chosen characteristics during the training phase. The forecasts from each tree are combined to create the final forecast. It effectively handles high-dimensional data and overfitting issues. Its accuracy and capacity to understand intricate correlations between data are its main advantages[21].

- **Naïve Bayes**

The Bayes theorem serves as the foundation for the probabilistic method known as the naïve Bayes classifier. This straightforward yet effective approach may predict classes based on the likelihood that a particular feature set belongs to each class. The Naïve Bayes classifier is named thus because it makes a strong assumption about the independence of the features[32].

- **K-Nearest Neighbor Algorithm**

The K-nearest neighbor algorithm is the least complex machine learning algorithm. It is a non-parametric algorithm built on the supervised learning method. It classifies unknown data based on adjacent data points. The KNN classifier's performance depends on two factors: the number of neighbors considered ('k') and the distance used to determine the nearest data points[33].

- **Decision tree**

Decision trees are supervised machine learning techniques. This method organizes the trained data into a tree of rules. Data for testing is randomly taken from the trained dataset.
After training and accuracy checks, untrained data is classified. The information gain rate is used to establish the most appropriate grouping class. This algorithm is widely utilized in multiple applications[34].

Performance evaluation metrics

This study assesses the performance of suggested models on test data using five parameters: accuracy, precision, sensitivity (recall), specificity, and F1-score. Using these evaluation criteria, we can anticipate classifier behavior on extracted feature data.

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$\boldsymbol{Sensitivity}(\boldsymbol{Recall}) = \frac{TP}{TP + FN} \times 100\%$$

$$\boldsymbol{Specificity} = \frac{TN}{TN + FP} \times 100\%$$

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \times 100\%$$

$$\mathbf{F1}\boldsymbol{Score} = 2 \times \frac{precision \times recall}{precision + recall} \times 100\%$$

where, TP is True Positive samples, TN is True Negative samples, FP is False Positive samples, and FN is False Negative samples[35].

D. Experiment Results

In this section the classification of various stress levels using EEG signals for two tasks from a publicly accessible dataset: an arithmetic test and a Stroop word test is presented.
The EEG signals were preprocessed using a low-pass filter at a cutoff frequency of 45 Hz. To facilitate subsequent signal analysis, each channel signal was divided into 2.5-second segments, each containing 625 points (frequency × 2.5 seconds).
Each channel signal was then analyzed using the EMD method to produce four IMFs and a residue. Five statistical features were used to calculate each component, and each segment included 200 features (number of channels, number of components, and number of statistical features).

The Butterfly optimization approach was utilized to reduce computational complexity, enhance model performance, and choose the most crucial features. After applying BOA, the dimension was reduced from 4385 × 200 to 4385 × 60. Each segment now has only 60 features. Five classifiers (SVM, NB, KNN, RF, and DT) were utilized to assess the suggested model's accuracy in identifying stress levels (normal, low, medium, and high) using BOA's selected features. The suggested model was evaluated in three scenarios: each task independently and then combined.

In this study, the proposed method was initially applied to EEG signals during a Stroop word task at three levels (low, mid, and high), which rely on attention and information incongruity. A word was displayed, and instead of reading it, participants were asked to determine its color. Due to the incongruity between the displayed word and the color, this causes a cognitive load. Signals from the EEG show the neuronal interplay caused by this cognitive stress. High performance was attained by Naive Bayes using 98.89% accuracy was attained. Table 1 displays classifiers' performance in utilizing EEG signals to classify stress levels during the Stroop Word Test.

After that, the suggested model is utilized to analyze the EEG signals produced by arithmetic tests, which are cognitive exercises meant to cause cognitive stress. Participants are required to execute arithmetic tasks of increasing difficulty (low, medium, and high) on each trial, which stimulates certain brain regions. With the Naive Bayes classifier, the greatest accuracy of 98.82% was attained. Table 2 shows the classifiers' performance in classifying stress levels using EEG signals during the arithmetic test.

Finally, the proposed model was used after merging data from the two tasks (Stroop Word Test and Arithmetic Test), which comprised both visual and arithmetic tasks. The purpose of the combination is to evaluate the generalizability of a single stress level classification across various cognitive situations. The results indicated a minor improvement, with the SVM classifier achieving 99.021% accuracy. Table 3 exhibit performance of classifiers when the two tasks are combined. The model's enhanced performance after task combining indicates that adding variety to cognitive tasks enriches the signal and improves the model's ability to classify various stress levels. Figure 5: Displays the classifiers' performance across every scenario.

Table 1. Classifiers' performance in utilizing during the Stroop color Word Test.

Model	Accuracy	Precision	Sensitivity	F1_Score	Specificity
SVM	98.531	98.611	98.534	98.543	99.511

Naive Bayes	98.898	98.934	98.90	98.903	99.633
Random Forest	98.164	98.186	98.137	98.166	99.3
KNN	97.43	97.468	97.431	97.438	99.143
Decision Tree	92.534	92.613	92.546	92.552	97.611

Table 2. The classifiers' performance during the arithmetic test.

Model	Accuracy	Precision	Sensitivity	F1_Score	Specificity
Naive Bayes	98.821	98.874	98.479	98.375	99.607
SVM	98.703	98.767	98.826	98.712	99.568
Random Forest	98.585	98.599	98.738	98.587	99.528
KNN	98.467	98.484	98.692	98.471	99.489
Decision Tree	96.934	96.961	96.971	96.942	98.978

Table 3. Performance of classifiers when the two tasks are combined.

Model	Accuracy	Precision	Sensitivity	F1_Score	Specificity
Naive Bayes	98.592	98.599	98.513	98.591	99.531
SVM	99.021	99.053	99.119	99.025	99.673
Random Forest	98.225	98.23	98.403	98.223	99.408
KNN	98.042	98.05	98.221	98.713	99.347
Decision Tree	94.553	94.576	94.555	94.563	98.184

E. Discussion

The current study's findings indicated significant efficiency in discriminating between multiple levels of mental stress using EEG data. A novel methodology was proposed that combines the EMD method to analyze signals into a collection of IMF and residual components and calculates a set of statistical features for each component, as well as employing the Butterfly optimization process to choose the most relevant and impactful features. This resulted in a significant improvement in the performance of categorization models employed in a variety of cognitive tasks, including the Stroop Word Test and the Arithmetic Test. The NB classifier achieved a classification accuracy of more than 98%, which increased to 99% when these tasks were combined using the SVM classifier. This indicates the suggested approach's usefulness in detecting stress-related brain alterations across diverse cognitive settings. Comparing these findings to previous studies indicates the effectiveness of the proposed strategy through:

In the Kumar et al.[12] study, only male volunteers were recruited to exclude the chance that hormonal cycles would influence the stress induction process, as opposed to the current study, which is more generalizable. BADR et al.[13] utilized a subjective measure, the PSS-10 questionnaire, to quantify perceived stress levels, making the results prone to errors or bias, whereas our suggested study used an objective assessment based on EEG features, making it unbiased.

According to KHAN et al. [14], the triple classification's classification accuracy dropped from 95.80% to 86.02% when compared to the binary classification. Additionally, only one task was used in the study. In comparison, our recent investigation attained an accuracy of over 98% and was conducted across various scenarios. An EEG method combined with a game was used by Vazquez et al.[15] to estimate felt stress levels. As a result, not all forms of stress might be represented in the game design. In contrast, the Stroop word task and the arithmetic test were realistic tasks in our current study.

The classification accuracy of the models suggested by Hemakom et al.[16] and Yang et al.[17] is lower than the findings of our current study. In addition to the challenge of managing noise and artifacts, Hafeez et al. [18] employed tiny samples in their investigation. In our present investigation, we used a larger sample size and improved noise processing. Table 4 shows a summary of the performance of previous studies with the proposed model. Our current study has some limitations such as a small sample size, despite the good performance. Consequently, the model can be used to bigger samples in order to generalize it. The model can be tested in settings with a variety of everyday pressures in subsequent research. The model can be applied to data with and without time limits because the amount of time allotted to each task level can also have an impact.

Table 4. Comparison of performance between previous studies and the proposed model.

Authors	Methods	Number Of Levels	Accuracy

Kumar et al.[12]	CNN+LSTM with self-attention	Four	91.81 ± 4.26%
BADR et al.[13]	GCNs	Four	76.37 ± 8.40%
KHAN et al. [14]	STFT+ ResNet50	Three	86.02%
Vazquez et al.[15]	RNN	Multi	94%
Hemakom et al.[16]	Combinations of classification and voting methods	Multi	87.14%
Yang et al.[17]	SHAP+ fuzzy rules+ decision tree method	Three	84.93%
Hafeez et al. [18]	CNN	Multi	90.46%
Proposed methods	**EMD + Statistical Features +BOA**	**Four**	**98.898 % (Stroop color words tasks) 98.82 % (Arithmetic tasks) 99.021%(combination)**

F. Conclusion

In this study, the EEG signals were analyzed using the empirical mode decomposition technique, and extracted a set of statistical features from the analyzed signal. The butterfly optimizer method was employed to select the optimal features, and five algorithms were utilized to determine stress levels from Stroop color words and arithmetic tasks, aiming to assess the models' accuracy in classifying stress levels.

When a Naive Bayesian classifier was used to classify stress levels in the Stroop color word and arithmetic tasks, respectively, 98.82% and 98.87% accuracy were obtained. The accuracy reached to 99.021% after merging the two tasks with an SVM classifier. The study found that using the proposed model with EEG signals to analyze individuals' mental responses in diverse settings is an effective tool since it can distinguish between different levels of stress. This helps to further the development of technological solutions for stress monitoring and psychological and mental health improvement in the workplace and study environments.

G. References

[1] M. Bolpagni, S. Pardini, M. Dianti, and S. Gabrielli, “Personalized Stress Detection Using Biosignals from Wearables: A Scoping Review,” May 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/s24103221.

[2] V. Dham, K. Rai, and U. Soni, “Mental Stress Detection Using Artificial Intelligence Models,” in Journal of Physics: Conference Series, IOP Publishing Ltd, Aug. 2021. doi: 10.1088/1742-6596/1950/1/012047.

[3] F. Al-Shargie, S. M. Taresh, and A. Al-Ezzi, “Mental Stress and Cognitive Deficits Management,” Apr. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/brainsci14040316.

[4] P. Boucher and P. Plusquellec, “Acute stress assessment from excess cortisol secretion: Fundamentals and perspectives,” Front Endocrinol (Lausanne), vol. 10, p. 749, 2019.

[5] K. Dedovic, ; Catherine D’aguiar, and J. C. Pruessner, “What Stress Does to Your Brain: A Review of Neuroimaging Studies.”

[6] W. B. Cannon, “The wisdom of the body,” 1939.

[7] R. Wesley and H. Pham, “Stress Studies: A Review,” in Springer Series in Reliability Engineering, vol. Part F15, Springer Science and Business Media Deutschland GmbH, 2025, pp. 267–282. doi: 10.1007/978-3-031-72636-1_13.

[8] T. Razavi, “Self-report measures: An overview of concerns and limitations of questionnaire use in occupational stress research,” 2001.

[9] G. S. Kumar and B. Ankayarkanni, “Leveraging Conv-XGBoost algorithm for perceived mental stress detection using Photoplethysmography,” Intell Based Med, vol. 11, Jan. 2025, doi: 10.1016/j.ibmed.2025.100209.

[10] S. Gedam, S. Dutta, and R. Jha, “Analyzing mental stress in Indian students through advanced machine learning and wearable technologies,” Sci Rep, vol. 15, no. 1, p. 20610, Jul. 2025, doi: 10.1038/s41598-025-06918-6.

[11] L. Abdul Kader, F. Al-Shargie, U. Tariq, and H. Al-Nashash, “One-Channel Wearable Mental Stress State Monitoring System,” Sensors, vol. 24, no. 16, Aug. 2024, doi: 10.3390/s24165373.

[12] A. Kumar, J. K. Barath, P. Shanmukh, and D. Joshi, “StreXNet: A Novel End-to-End Deep Learning Based Improved Multi-Level Mental Stress Classification from EEG Sensors,” IEEE Sens J, 2024, doi: 10.1109/JSEN.2024.3506984.

[13] Y. Badr et al., “EEG-Based Multi-Level Mental State Classification Using Partial Directed Coherence and Graph Convolutional Networks: Impact of Binaural Beats on Stress Mitigation,” IEEE Access, vol. 13, pp. 61284–61298, 2025, doi: 10.1109/ACCESS.2025.3553932.

[14] S. Khan, S. H. Noorani, J. Frnda, U. Rauf, A. Arsalan, and S. M. U. Saeed, “A Transfer Learning-Based Framework for Enhanced Classification of Perceived Mental Stress Using EEG Spectrograms,” IEEE Access, vol. 13, pp. 89266–89286, 2025, doi: 10.1109/ACCESS.2025.3571437.

[15] J. J. Gonzalez-Vazquez, L. Bernat, J. L. Ramon, V. Morell, and A. Ubeda, “A Deep Learning Approach to Estimate Multi-Level Mental Stress from EEG Using Serious Games,” IEEE J Biomed Health Inform, vol. 28, no. 7, pp. 3965–3972, 2024, doi: 10.1109/JBHI.2024.3395548.

[16] A. Hemakom, D. Atiwiwat, and P. Israsena, “ECG and EEG based machine learning models for the classification of mental workload and stress levels for women in different menstrual phases, men, and mixed sexes,” Biomed Signal Process Control, vol. 95, Sep. 2024, doi: 10.1016/j.bspc.2024.106379.

[17] L. Yang, R. Zhou, G. Li, Y. Yang, and Q. Zhao, “Recognizing and explaining driving stress using a Shapley additive explanation model by fusing EEG and behavior signals,” Accid Anal Prev, vol. 209, Jan. 2025, doi: 10.1016/j.aap.2024.107835.

[18] M. A. Hafeez and S. Shakil, “EEG-based stress identification and classification using deep learning,” Multimed Tools Appl, vol. 83, no. 14, pp. 42703–42719, Apr. 2024, doi: 10.1007/s11042-023-17111-0.

[19] A. Nirabi, F. Abd Rahman, M. H. Habaebi, K. A. Sidek, S. H. Yusoff, and L. A. M. Jefri, “Cognitive Load Assessment Through EEG: A Dataset from Arithmetic and Stroop Tasks,” Mendeley Data, vol. V, no. 1, pp. 1–10, Mar. 2024.

[20] A. Nirabi, F. A. Rahman, M. H. Habaebi, K. A. Sidek, and S. Yusoff, “Cognitive load assessment through EEG: A dataset from arithmetic and Stroop tasks,” Data Brief, vol. 60, Jun. 2025, doi: 10.1016/j.dib.2025.111477.

[21] M. Deshmukh, M. Khemchandani, and M. Mhatre, “Impact of brain regions on attention deficit hyperactivity disorder (ADHD) electroencephalogram (EEG) signals: Comparison of machine learning algorithms with empirical mode decomposition and time domain analysis,” Appl Neuropsychol Child, 2025, doi: 10.1080/21622965.2025.2511691.

[22] R. Upadhyay, P. K. Padhy, and P. K. Kankar, “EEG artifact removal and noise suppression by Discrete Orthonormal S-Transform denoising,” Computers & Electrical Engineering, vol. 53, pp. 125–142, 2016, doi: https://doi.org/10.1016/j.compeleceng.2016.05.015.

[23] K. Singh and J. Malhotra, “Deep learning based smart health monitoring for automated prediction of epileptic seizures using spectral analysis of scalp EEG,” Phys Eng Sci Med, vol. 44, no. 4, pp. 1161–1173, 2021.

[24] I. H. Elshekhidris, M. B. Magdi, and A. Fragoon, “Automated classification of epileptic seizures using modified one-dimensional convolution neural network based on empirical mode decomposition with high accuracy,” Neuroscience Informatics, vol. 5, no. 2, Jun. 2025, doi: 10.1016/j.neuri.2025.100188.

[25] J. A. A. de Menezes, J. C. Gomes, V. de Carvalho Hazin, J. C. S. Dantas, M. C. A. Rodrigues, and W. P. dos Santos, “Classification based on sparse representations of attributes derived from empirical mode decomposition in a multiclass problem of motor imagery in EEG signals,” Health Technol (Berl), vol. 13, no. 5, pp. 747–767, Sep. 2023, doi: 10.1007/s12553-023-00770-2.

[26] J. Shen et al., “Exploring the Intrinsic Features of EEG Signals via Empirical Mode Decomposition for Depression Recognition,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 356–365, 2023, doi: 10.1109/TNSRE.2022.3221962.

[27] X. Geng et al., “A method of EEG signal feature extraction based on hybrid DWT and EMD,” Alexandria Engineering Journal, vol. 113, pp. 195–204, Feb. 2025, doi: 10.1016/j.aej.2024.11.056.

[28] I. H. Elshekhidris, M. B. Magdi, and A. Fragoon, “Automated classification of epileptic seizures using modified one-dimensional convolution neural network based on empirical mode decomposition with high accuracy,” Neuroscience Informatics, vol. 5, no. 2, Jun. 2025, doi: 10.1016/j.neuri.2025.100188.

[29] S. N. Makhadmeh et al., "Recent Advances in Butterfly Optimization Algorithm, Its Versions and Applications," Mar. 01, 2023, Springer Science and Business Media B.V. doi: 10.1007/s11831-022-09843-3.
[30] A. Asha, G. Ardhra, and K. R. Kavitha, "Butterfly Optimization and Neural Network Based Feature Selection and Classification of Gene Expression Data," in 2024 5th IEEE Global Conference for Advancement in Technology, GCAT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/GCAT62922.2024.10923963.
[31] S. Arora and S. Singh, "Butterfly optimization algorithm: a novel approach for global optimization," Soft comput, vol. 23, no. 3, pp. 715–734, Feb. 2019, doi: 10.1007/s00500-018-3102-4.
[32] M. G. Murariu, F. R. Dorobanțu, and D. Tărniceriu, "A Novel Automated Empirical Mode Decomposition (EMD) Based Method and Spectral Feature Extraction for Epilepsy EEG Signals Classification," Electronics (Switzerland), vol. 12, no. 9, May 2023, doi: 10.3390/electronics12091958.
[33] P. Patel, S. Balasubramanian, and R. N. Annavarapu, "Cross subject emotion identification from multichannel EEG sub-bands using Tsallis entropy feature and KNN classifier," Brain Inform, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40708-024-00220-3.
[34] A. Abdulrahman and M. Baykara, "Feature extraction approach based on statistical methods and wavelet packet decomposition for emotion recognition using EEG signals," in 2021 International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Aug. 2021. doi: 10.1109/INISTA52262.2021.9548406.
[35] A. Zülfikar and A. Mehmet, "Empirical mode decomposition and convolutional neural network-based approach for diagnosing psychotic disorders from eeg signals," Applied Intelligence, vol. 52, no. 11, pp. 12103–12115, Sep. 2022, doi: 10.1007/s10489-022-03252-6.