IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 July 2025, accepted 31 July 2025, date of publication 6 August 2025, date of current version 14 August 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3596517

RESEARCH ARTICLE

Mental Workload Detection and Assessment Through Statistical Features Extraction and Optimization Using GEL-RF Method for EEG Signals Using N-Back Dataset

WALEED MANZOOR[1,2], NOMAN NASEER[3], (Senior Member, IEEE), IMRAN FAREED NIZAMI[4], SYED HAMMAD NAZEER[3], (Senior Member, IEEE), AND HUSAM A. NEAMAH[5,6,7], (Member, IEEE)

[1]Department of Electrical and Computer Engineering, Air University, Islamabad 44000, Pakistan
[2]Department of Computer Engineering, Bahria University, Islamabad 44000, Pakistan
[3]Department of Mechatronics and Biomedical Engineering, Air University, Islamabad 44000, Pakistan
[4]Department of Electrical Engineering, Bahria University, Islamabad 44000, Pakistan
[5]Department of Electrical Engineering and Mechatronics, Faculty of Engineering, University of Debrecen, 4028 Debrecen, Hungary
[6]Technical Engineering College, Al-Ayen University, Thi-Qar 64001, Iraq
[7]College of Engineering, National University of Science and Technology, Dhi Qar 64001, Iraq

Corresponding authors: Noman Naseer (noman.naseer@mail.au.edu.pk) and Husam A. Neamah (husam@eng.unideb.hu)

ABSTRACT Mental workload (MW) assessment measures the cognitive effort required to perform tasks and is crucial in fields such as aviation, clinical medicine, and healthcare. This study addresses the critical need to optimize cognitive load and resource allocation by evaluating MW through EEG-based analysis. Recently, MW assessment using EEG signals has gained importance, as they show a high correlation with specific cognitive processes and MW. Although previous studies have employed various techniques for MW assessment using EEG, detailed statistical feature selection remains underexplored. This work analyzes the N-back dataset, a standard benchmark for cognitive load, using four combinations of binary and multiclass classification to assess three MW levels. We propose a novel model, which integrates genetic, evolutionary and linear (GEL) feature selection techniques with a Random Forest (RF) classifier as GEL-RF model. Experimental results demonstrate that the GEL-RF model achieves classification accuracies of 97% for binary tasks and 96.3% for multiclass tasks MW assessment, outperforming existing methods. These results shows that our proposed model can improve MW assessment accuracy, helping to increase safety and efficiency in demanding mental tasks.

INDEX TERMS Mental workload assessment (MWA), statistical features, electroencephalogram (EEG), genetic, evolutionary, linear (GEL) search algorithms, N-back, random forest (RF).

I. INTRODUCTION

Mental workload (MW) assessment is widely studied and investigated in recent years [1]. MW is the ability to process information, which is used while executing a task [2]. To perform better in any environment, people need to make minimal mistakes. Measuring MW helps identify performance issues and how well task can be done [3]. Studies on MW seek to improve how the brain handles information [4], [5] High mental workload can cause tiredness, reducing performance and efficiency. Heavy workloads often lead to serious accidents due to mistakes or poor decisions [5]. Dealing with new tasks and interruptions during high MW makes performance and decision-making harder [6]. Human performance may be safety-critical for human lives in various circumstances, as in jobs carried out by first responders, firefighters, and pilots [7]. Most aviation accidents (70%) happen because of human errors, which are due to poor

The associate editor coordinating the review of this manuscript and approving it for publication was Ines Domingues.

mental responses [8]. In these situations, assessing someone's ability to do a task well is key to preventing accidents and saving lives [7].

Clinical-grade equipment is commonly used in the literature to track psychophysiological reactions [9], [10]. However, these devices typically take a long time to set up and are uncomfortable to wear, which may prevent the monitored subject from moving around freely to do their activities. Using wearable technology is important for monitoring mental workload in real-world situations [7].

Different methods, like the NASA task load index (NASA-TLX), Subjective Workload Assessment Technique (SWAT), etc., are used for subjective measurement of MW. Body signals like brain activity (EEG), heart rate through electrocardiogram (ECG), and eye movement are used to measure MW [11], [12]. Recent developments in low-cost EEG technologies have pushed new fields of study beyond the diagnosis of neurological illnesses, such as interacting with household appliances, educating students, and controlling robotic limbs. For the specific purpose of assessing MW, EEG is quickly replacing other sensors as the preferred sensor [13], [14].

Measuring mental workload during complex tasks is important to understand stress levels and prevent accidents. In brain-computer interfaces (BCI), common steps that are usually followed are data acquisition, preprocessing, feature extraction, and classification, but optimal feature selection is the unattended domain where feature optimization can be done. Choosing the right features is the key to getting useful information. This helps improve how accurately we classify data by using the most important details [15]. Time domain, frequency domain, and spatial features are extracted for MW classification [15], [16].

One of the key benefits of using brain activity (EEG features) to infer cognitive workload is that it has close to ideal temporal resolution with regards to the resolution of under a minute. Currently, no comprehensive set of features has been able to established to assess mental work-load within individuals and across different individuals. For that reason, it is required to establish a single universal set of features that accounts for all the individuals. If such a model is identified, the subsequent objective is to ascertain whether a common workload classification model exists that can be used across subjects [1]. Comprehensive information about which EEG features are linked to cognitive workload in multitask environments is seen in [17]. Stepwise discriminant analysis (SWDA) technique is utilized to determine optimal features by maximum absolute loading and ranking them with respect to their relative importance of features with their boundaries, but their high correlation makes it difficult to find out the redundant features easily [1]. The signal-to-noise ratio (SNR) metric assists in understanding the relative importance of each element as compared to that of the noise. SNR saliency screening is employed for selection of artificial neural network (ANN) features, but SNR leads to miscalculation of its values due to variation in noise, neglecting correlation among features and hindering the performance of feature selection of ANN [18].

The aim of this study is to assess the MW of the human brain using EEG signal analysis through optimal statistical feature extraction and selection. In order to minimize feature space and feature irrelevancy, it's important to identify and use the most relevant features at the same time. In order to work on this, researched various features and feature selection strategies that can be used in a variety of applications. A novel model using the GEL search algorithm has been proposed to find out the optimal features and perform classification with the RF classifier, which has provided accuracy of 97.6% for binary tasks and 96.3% for multiclass tasks.

The novelty of the proposed GEL-RF method lies in its hybrid feature selection framework and enhanced classification accuracy for EEG-based mental workload (MW) assessment. The proposed method achieves an accuracy of 97.0% for binary class problem and 96.3% for multiclass problem. The proposed framework optimization systematically eliminates redundant features while preserving discriminative statistical properties across time, frequency and entropy domains. The optimized features are then classified using a RF classifier, which is well-suited for handling EEG signals and improves classification accuracy.

Filter-based methods Relief-F [19] neglects feature interdependencies, or hierarchical algorithms [20], which struggle in determining multicollinearity. In contrast, GEL-RF's hybrid feature selection integrates genetic, evolutionary, and linear search strategies which eliminate redundant features and achieve adequate accuracy. Furthermore, in comparison to computationally intensive 3D-CNNs [2] or transfer learning [21], proposed method offers better accuracy as validated by t-SNE visualizations showing enhanced class separability in Figure 3. This clearly presents proposed model's improved feature optimization and generalizability, making it distinguishable from existing methodologies.

The remainder of this paper is organized as follows: Section II which shows the previous work done on MW assessment using EEG. Section III explains the method used to recognize the different levels of workload using the GEL-RF model, feature extraction and selection, and classification. Section IV is devoted to the experimental results and discussion of binary and multiclass MW assessment. Finally, section V outlines the conclusions and future work of MW assessment using EEG signals.

II. RELATED WORK

The most widely used methods of quantifying workload can be divided into two primary groups: objective scores derived from physiological responses and subjective measures based on the subject's perception [6]. Subjective measures remain the most common way to evaluate MW; the NASA-TLX [22] is the most widely used test to learn about the perceived

workload levels using different human-machine interface systems [23], [24]. Based on a weighted average of six subvariables (mental demand, physical demand, temporal demand, performance, effort, and frustration), this questionnaire calculates the MW. It is frequently used in aviation to evaluate pilots mental strain when they are operating aircraft controls [25], [26]. Physiological measures have gained popularity among researchers recently because they offer more accurate data on workload, as they measure physiological changes that are unconscious [27], [28], [29].

The most widely used measurements and sensors for gathering physiological data are respiration rate sensors, electrodermal activity (EDA) to measure skin surface temperature, oxygen density of blood in the brain, eye movement trackers, ECG to register heart electrical activity, electromyography (EMG) to read skeletal muscle electrical activity, EEG to detect electrical activity in the brain, photoplethysmography to register volumetric changes in blood flow, and others [30]. One can evaluate the perceived workload using TLX surveys, although it is very subjective [22]. Physiological data come from natural processes and, when combined with TLXs, offer more accurate information [11], [23], [27]. Previous studies in the field of MW assessment using EEG signals and machine learning techniques have made significant contributions to understanding and quantifying cognitive load in various tasks [19]. Because of its ability to capture temporal information in real time, the EEG approach is chosen [19]. Due to its non-invasive nature, EEG is also regarded as a safe technique for frequent and prolonged use. Interestingly, no medical treatments and the presence of a skilled operator are required [31].

High mental stress reduces pilot performance. Situation Awareness and Workload in Aviation examines, through the use of surveys, subjective questionnaires, and task performance data. The NASA-TLX index and the Situation Awareness Rating Technique (SART) are used in the study to quantify workload [25]. Objective real-time workload insights may be obtained by fusing subjective data with physiological measurements like EEG [25]. EEG, ECG, and eye tracking are just a few multimodel methods used in measuring neurophysiological signals in aircraft pilots, exhaustion, and drowsiness in car drivers for the assessment of MW [32]. The results show that EEG is a powerful predictor of mental strain and stress, especially in the alpha and beta frequency ranges. Sensitivity of EEG to cognitive processes is one of its advantages, but movement-related noise produces difficulties for algorithms. We should focus on improving noise removal techniques and making good machine learning models [13].

Researchers analyze brain signals in specific frequency ranges (delta, theta, alpha, beta, and gamma) to understand MW. Identifying the best patterns in brain signals is key to making EEG better at measuring mental workload [33]. Within their work, they also briefly discussed the most used features for BCI classified into time domain, frequency domain, time-frequency domain, and spatial domain [34]. Different changes in a subject's behavior can be identified by locating and extracting a few unique features from the EEG. As applications for EEG analysis are numerous and features are chosen based on the kind of applications [35]. As a single modality, EEG in particular has been widely used in combination with various machine learning methods [12]. In the study, various machine learning models, including support vector machine (SVM), RF, and neural networks, as well as signal processing techniques like Fourier and wavelet transforms, are used for the MW assessment using EEG signals [13].

Improving EEG signal classification for motor imagery tasks involves using statistical wavelets and harmony search to enhance accuracy by selecting the best features in brain-computer interfaces (BCI). The approach is highly effective for feature selection but is computationally expensive, as search is dependent on Harmony memory considering rate (HMCR) and pitch adjusting rate (PAR), providing issues for MW assessment applications [15]. Feature selection is a crucial aspect in reducing the feature space and complexity while also identifying relevant features. According to the taxonomy of feature selection, the evaluation strategy can be categorized into five main groups: filter method, wrapper method, embedded method, hybrid method, and ensemble method. It is used in channel selection and feature selection problems for various EEG-based applications [15].

The use of multilevel feature fusion with 3D convolutional neural networks (CNN) is suggested for EEG-based workload prediction in order to increase the precision. The method provides good classification accuracy in the evaluation of MW by capturing both spatial and temporal information from EEG data. Although the technique is quite good at automatically extracting complicated information, its high computational cost makes deployment challenging [2]. In the investigation executed by Liu et al. [36], subjects are directed to engage in an n-back memory assessment to evaluate the workload levels of 21 participants using EEG signals across three distinct experimental conditions. In the study, feature correction analysis and transfer learning methodology on EEG signals produced more accurate results in comparison to their isolated implementation. As transfer learning makes it computationally expensive, and on limited data system performance gets compromised.

Jusas and Samuvel [37] perform analysis through linear discriminant analysis (LDA) for four unique motor imagery tasks on nine subjects. These tasks necessitated that participants participate in cognitive processes that involved the rehearsal or simulation of particular actions. As this method leads to an overfitting issue, and the feature selection method depends on the specific dataset and classification task. In their scholarly investigation involving seven participants, Yin and Zhang [9] developed a predictive model for cognitive workload utilizing binary classification techniques. The study tested how accurate the model's predictions were compared

to standard methods like SVM and k-nearest neighbor (kNN). The results showed this approach works better than other classifiers. The method struggles to stay consistent across different session variations.

Pei et al. [38] executed an empirical investigation involving a cohort of seven subjects, wherein they attained better accuracy through the implementation of the RF algorithm. They created a model to predict outcomes using various EEG features collected from a 64-channel EEG device during a simulated aviation task. Moreover, the classification efficacy of the three-level model, which exclusively employed band power metrics, achieved a better performance. The study did not go for compressed features, due to which it is unable to detect the concerned brain area.

In a study conducted by Wang et al. [7], the hierarchical Bayes model's cross-subject classifier is trained and tested using EEG data from a cohort of eight participants. The performance of the cross-subject classifier is similar to that of individual-specific classifiers. Multi-model classification reduces the accuracy, as multi-model data synchronization is a complex task. In the research undertaken by Kaczorowska et al. [39], a cohort of 29 participants engaged in a numerical symbol matching assessment, which is categorized into three tiers of complexity. The methodologies of logistic regression and RF are utilized to ascertain the precision of the 20-variable model, which is formulated utilizing ocular tracking data, but with parameter tuning, performance can be improved. Grimes et al. [31] undertook an investigation that categorized MW through the implementation of n-back tasks, resulting in a notable degree of precision. Through the application of the Naive Bayes (NB) algorithm, the model with two levels of difficulty exhibited higher accuracy, whereas the model featuring four levels of difficulty attained little low accuracy comparatively. So it can improve the accuracy of more workload levels with the improvisation of model adaptability.

Wu et al. [40] executed an empirical investigation centered on predictive analytics employing the eye-tracking methodology. The research comprised a sample of 32 individuals categorized as non-experts and seven individuals recognized as experts. The investigators incorporated metrics such as pupil diameter, blink frequency, focusing velocity, and saccadic movement speed as variables within their analytical model, whereas the NASA-TLX score is utilized as the dependent variable. The ANN model's prediction accuracy gets lower because fewer samples make it less reliable.

Cognitive load experienced by automobile drivers illustrated that a Feedforward Neural Network (FNN) attained good prediction accuracy in assessing levels of alertness, drowsiness, and sleep states [41]. The study shows that the overfitting issue is there, which can be overcome with a larger neural network. In an alternative investigation, a cohort of 28 subjects participated in a computer-mediated aviation task. Throughout this investigation, the cerebral signals of the participants are captured utilizing a 32-channel EEG apparatus, where as its performance degraded due to the inclusion of eye tracking data [42]. Various methods, including RF, SVM, ANN, and linear regression, are used in the study. In the three-category classification task, the model that combined EEG variables produced the better accuracy rate among these algorithms. SVM has been used in a number of studies in the literature to classify MWs using EEG or eye tracking data [21], [43], [44], [45], [46], [47], [48], [49]. [20] utilized effective brain connectivity measures, selecting significant features through hierarchical algorithms, resulting in effective accuracy. The above methodologies utilized for feature selection are designed to reduce repetition within the dataset. There's a chance that some repeated features could still provide useful information, possibly improving the model's performance if included.

Various investigations have utilized feature selection methodologies, including Relief-F [19] to augment model efficacy and reduce dimensionality. Although EEG-based evaluations offer significant perspectives on cognitive workload, obstacles such as individual variability and the impact of external factors remain. Enhancement of models should be prioritized to accommodate these issues for predictive precision [20], [50].

This unified approach helps in achieving high accuracy, resolving class overlap issues and reducing dimensionality. The model reduces the number of features from 38 to 19 while preserving discriminative power across time, frequency, and entropy domains. The model's integration with a RF classifier achieves accuracy of 97.0% for binary class and 96.3% for multiclass MW classification. By addressing challenges such as noise sensitivity and inter-subject variability GEL-RF establishes a new benchmark for MW assessment, with significant implications for applications in aviation, healthcare, and BCI.

In summary, previous studies on MW assessment using EEG signals and machine learning techniques have demonstrated promising results with varying accuracies. However, challenges such as feature selection and optimization, model complexity, and interpretability remain. Addressing these limitations in future research will contribute to the development of more optimized and applicable models for MW assessment.

III. METHODOLOGY

Before using data analysis techniques, the EEG signal goes through a lengthy preparation process to eliminate noise and artifacts. Getting useful information about the signal depends heavily on data preprocessing [13]. EEG MW assessment is done through the GEL-RF model, which is shown in figure1. EEG signals are recorded from the portable headset EMOTIV EPOC+, which gives 14 channels of data according to the 10-20 standard system shown in figure 2 for electrode placement [14], [51]. This sensor offers power spectra and raw data for the primary brain frequencies. (Theta (θ) Alpha (α), Beta low (β_{low}) Beta high (β_{high}) and Gamma (γ)). The

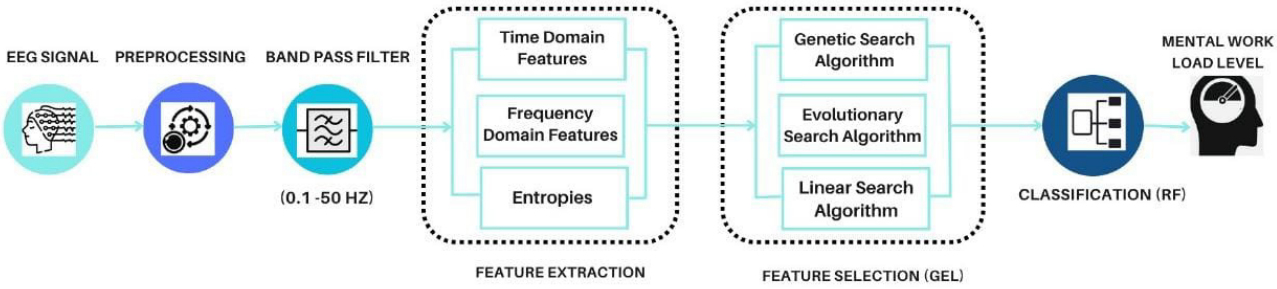

FIGURE 1. Mental workload assessment using statistical feature extraction and selection.

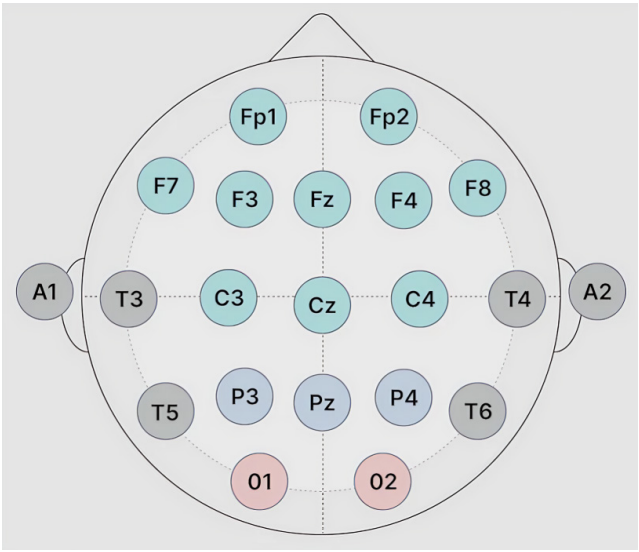

FIGURE 2. Standard 10-20 system for electrode placement.

subsection includes preprocessing, normalization, filtering, feature extraction, feature selection, and classification detail.

A. PREPROCESSING

Normalization: The N-back dataset comes with basic preprocessing and filtering, but still EEG signals need to be preprocessed for better efficiency and detection of mental states. Before feature extraction, Z-score normalization is performed on subject files, which can improve the performance of machine learning algorithms. It makes the feature set more comparable across different subjects, experimental conditions, and other factors. Band Pass Filtering: A 0.1 to 50 Hz third-order band-pass filter is applied to make sure of the removal of artifacts, wherein there come alpha, beta, and gamma waves in the brain activity, which covers MW activities. After filtering, it can be examined for certain characteristics or trends that might be connected to the neuronal activity under investigation, as BP filters isolate a specific frequency range of interest.

B. FEATURE EXTRACTION

Feature extraction is very important for the analysis of EEG signals, and hence it helps to extract relevant information from the raw signal. One common method for feature extraction includes the extraction of statistical features. EEG signals are used for the computation of a set of statistical parameters from time, frequency domain, and entropies. These parameters can highlight many signal properties, such as the distribution of frequencies and amplitude overall. Apart from these, more sophisticated statistical properties that can be extracted include entropy, kurtosis, skewness, and fractal dimension of the EEG signal. These features may provide greater insight into the signal's complexity and non-linearity.

All things considered, extracting statistical features from the EEG data will help to diagnose a number of neurological disorders along with revealing very important details on the underlying cerebral activity in the brain. Statistical features give a quantifiable estimate of the signal characteristics; they are important for any study of EEG signals. Statistical features include the mean, standard deviation, and power, which are easy to compute and clearly interpret the signal properties. They could provide information about the total amplitude of a signal, frequency composition, and energy, all of which may be used to detect anomalies or trends [34].

In contrast, auto-feature calculation and interpretation are much more involved in the computation. Those characteristics assume stationarity, which generally is not satisfied in cases of EEG signals, at least while dealing with non-stationary phenomena like seizures or some stages of sleep, as auto-features are less reliable than statistical features. Due to such non-stationarity, the statistical features are more noise-resistant and therefore are a better choice for dealing with EEG data [34]. Moreover, the extraction of statistical features from EEG signals is fast and computationally efficient, thus allowing for the analysis of large datasets within a relatively short period of time. This is particularly important when dealing with long EEG recordings, such as those used in sleep studies or in the monitoring of patients with epilepsy.

Statistical features are highly valuable tools for EEG signal analysis, giving a solid and effective way to extract information from the raw EEG signals. These features are especially valuable in identifying patterns or abnormalities in the signal and also for comparative study of various signals or segments of them. Thus, we have extracted statistical features

in the time domain, frequency domain, and entropies for MW assessment as follows:

C. TIME DOMAIN FEATURES

1) MEAN

The average amplitude of an EEG signal over a given time period is represented by the signal's mean in the time domain.

$$Mean = \bar{u} = \frac{1}{m}\sum_{k=1}^{m} x_k \tag{1}$$

The mean of the signal is determined by adding together all of the data points in the signal and dividing by the total number of data points. This gives an approximation of the average amplitude of the signal during the relevant time frame.

Where u is the EEG signal, m is the cumulative quantity of data points within the specified time interval, and x_k is the kth data point of the aforementioned signal.

2) MEDIAN

The median of an EEG signal within the temporal domain signifies the central value of the signal when it is systematically ordered in a numerical sequence across a defined temporal interval.

$$\text{median}(U) = u_{\frac{M+1}{2}-\frac{M}{2}*(1-\text{mod}(M,2))} \tag{2}$$

Here, in this context, M represents the samples that occur in the EEG signal and $U = u_1, u_2, ., u_N$ represents all signal values of the concerned signal lying in the temporal domain.

It gives the value of the arithmetic median for an odd N. In the case where N is an even number, it produces the value of the average of two middle values.

3) VARIANCE

The variance represents the quantitative measure of the deviation of the data values from the mean. The lower the variance, the closer the data points are to the mean. If the variance is high, then it shows that the data points have spread out widely. In signal processing, variance is commonly used because it gives a mathematical model describing the spread of a signal.

$$\sigma^2 = \frac{1}{M}\sum_{k=1}^{M}(u_k - \mu)^2 \tag{3}$$

where M is the number of samples in the EEG signal, u_k is the value of the signal at each sample point, and μ is the mean of the signal.

4) STANDARD DEVIATION

The standard deviation serves as a quantitative assessment of the variability or dispersion of a dataset in relation to its mean. A minimal standard deviation signifies that the individual data points tend to cluster closely around the mean, whereas a substantial standard deviation implies that the data points are more widely distributed.

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(u_i - \mu)^2} \tag{4}$$

where M is the number of samples in the EEG signal, u_i is the value of the signal at each sample point, and μ is the mean of the signal.

In addition, to be a useful indicator for evaluating MW, it should first determine a base value of the standard deviation associated with a relaxed or low workload condition and then compare that to the standard deviation during the high workload condition. An elevated standard deviation for the high workload state represents a higher MW.

5) SKEWNESS

The skewness of an EEG signal is a measure of the asymmetry of the signal's probability distribution.

$$Skewness = \frac{1}{M}\sum_{k=1}^{M}\frac{(u_i - \mu)^3}{\sigma^3} \tag{5}$$

where M is the number of samples in the EEG signal, u_i is the value of the signal at each sample point, μ is the mean of the signal, and σ is the standard deviation of the signal.

6) KURTOSIS

The kurtosis of an EEG signal serves as an indicator of the extent of flatness within the probability distribution of the signal.

$$Kurtosis = \frac{1}{M}\sum_{k=1}^{N}\frac{(u_i - \mu)^4}{\sigma^4} - 3 \tag{6}$$

where M denotes the number of samples in the EEG signal, the value of the signal is denoted by u_i at each sample point, the mean of the signal is μ, and the standard deviation of the signal is given by σ.

When kurtosis is positive, the distribution of the signal is spikier than that of a normal distribution; when kurtosis is negative, the distribution of the signal is flatter than that of a normal distribution. If the kurtosis is zero, then the signal probability distribution is similar to that of a normal distribution.

7) HJORTH MOBILITY

The ability of an EEG signal to fluctuate over time is measured by its Hjorth mobility in the time domain.

$$Mobility = \frac{\sqrt{var(dU)}}{\sqrt{var(U)}} \tag{7}$$

In the above equation, U is the original EEG signal, and dU is the first difference of the signal, which is computed as the difference between each sample and its predecessor. The denominator is the square root of the variance of the original signal, and the numerator is the square root of the variance of the first difference.

Hjorth mobility is defined as how fast the signal changes; the higher the value, the more it changes with time. It is commonly used for the comparison of a number of signals and changes in the signal.

8) HJORTH ACTIVITY

The energy or power of an EEG signal is measured by its Hjorth activity in the temporal domain.

$$Activity = \frac{var(dU)}{var(U)} \tag{8}$$

where dU is the first-order difference of the signal, or the result of subtracting samples from one another in pairs. The denominator is thus the variance of the original trace, while the numerator is the variance of the first-order difference of the signal.

The more the value, the more energy or power the signal has. This measure is often used to compare several signals and track changes in the signal.

9) HJORTH COMPLEXITY

By evaluating the amplitude and frequency content of a signal, Hjorth Complexity is one time-domain measure that determines the degree of complexity of such a signal. With the concept of entropy in mind, it can be obtained by using the signal's variances, its derivative, and the derivative of its derivative. A number of neurological and psychiatric conditions have employed research based on the application of the Hjorth Complexity to EEG signals that analyze the complexity of brain activity.

$$C = \frac{\sqrt{\frac{(\mathrm{Var}(\Delta u))^2}{\mathrm{Var}(u)}}}{\sqrt{\mathrm{Var}(\Delta u)}} \tag{9}$$

where Δu is the signal's first derivative, and u is the EEG signal. The Hjorth parameters, which are frequently used to calculate the Hjorth complexity, are the foundation upon which the aforementioned equation is written. However, the precise definition of Hjorth complexity can differ depending on the source.

10) MEAN ENERGY

The average squared amplitude of an EEG signal over a particular time period is called the mean energy of the signal. It is mathematically expressed as

$$E = \frac{1}{T}\int_{t=0}^{t=T} u^2(t)dt \tag{10}$$

When T is the complete time period, u(t) represents the EEG signal at time t, and the integral is calculated across the whole duration.

D. FREQUENCY DOMAIN FEATURES

When an EEG signal is converted from the time domain to the frequency domain using the fast Fourier transform, its properties are referred to as frequency domain features.

1) MEAN

The average of the signal's complex values at each frequency can be used to determine the mean of a continuous EEG signal in the frequency domain.

$$M = \frac{1}{B}\int_{f=0}^{f=B} U(f)df \tag{11}$$

where B represents the entire frequency range, U(f) is the signal's complex value at frequency f, and the integral is calculated across all frequencies.

2) MEDIAN

The median value of the complex values of a continuous EEG signal at each frequency can be used to compute the signal's median in the frequency domain.

$$M = \mathrm{median}U(f) \tag{12}$$

where the median is the middle value of the complex values and U(f) is the complex value of the signal at frequency f.

3) VARIANCE

The mean of the squared deviations between the mean and the complex values of the signal at each frequency can be used to determine the variance of a continuous EEG signal in the frequency domain.

$$V = \frac{1}{B}\int_{f=0}^{f=B} (U(f) - \mu)^2 df \tag{13}$$

where μ is the signal mean, B is the entire frequency range, U(f) is the signal's complex value at frequency f, and the integral is calculated across all frequencies.

4) STANDARD DEVIATION

One may calculate the standard deviation of a continuous EEG signal in the frequency domain by taking the square root of the variance of the signal's complex values at each frequency.

$$S = \sqrt{\frac{1}{B}\int_{f=0}^{f=B} (U(f) - \mu)^2 df} \tag{14}$$

where μ is the signal mean, B is the entire frequency range, U(f) is the signal's complex value at frequency f, and the integral is calculated across all frequencies.

5) SKEWNESS

The normalized third moment of the complex values at each frequency of the signal can be used to find the skewness of a continuous EEG signal in the frequency domain.

$$S = \frac{1}{B}\int_{f=0}^{f=B} \left(\frac{U(f) - \mu}{S}\right)^3 df \tag{15}$$

In this case, U(f) is the signal's complex value at frequency f, B is the entire frequency range, μ is the signal's mean, S is its standard deviation, and the integral is calculated across all frequencies.

6) KURTOSIS

Bandpower (BP) by adding up the power of the signal at each frequency in the theta band (4-8 Hz) and normalizing it by the overall power of the signal, one can determine the theta of a continuous EEG signal in the frequency domain.

$$K = \frac{1}{B}\int_{f=0}^{f=B}(\frac{U(f)-\mu}{S})^4 df - 3 \tag{16}$$

In this case, U(f) is the complex value of the signal at frequency f, B is the entire range of frequencies, μ is the mean of the signal, S is its standard deviation, and the integral is calculated over all frequencies.

7) BANDPOWERDELTA

A continuous EEG signal's bandpower delta in the frequency domain can be computed by adding up the signal's power at each frequency in the delta band (0-4 Hz) and normalizing it by the signal's overall power.

$$BP_{\Delta} = \frac{\int_{f=0}^{f=4} P(f)df}{\int_{f=0}^{f=B} P(f)df} \tag{17}$$

where the integrals are taken over the full frequency range and the delta band, respectively, and P(f) is the power of the signal at frequency f, and B is the overall frequency range.

8) BANDPOWERTHETA

BP by summing up the power of the signal at each frequency in the theta band (4-8 Hz) and normalizing it by the overall power of the signal, the theta of a continuous EEG signal in the frequency domain can be determined.

$$BP_{\Theta} = \frac{\int_{f=4}^{f=8} P(f)df}{\int_{f=0}^{f=B} P(f)df} \tag{18}$$

where the integrals are taken over the whole frequency range and the theta band, respectively, and P(f) is the power of the signal at frequency f, and B is the overall frequency range.

9) BANDPOWERALPHA

The strength of a signal inside a particular frequency range is measured by its BP. It is the EEG signal's power in the alpha frequency band, which normally spans 8-12 Hz, when discussing EEG signals. It is frequently used to measure the EEG signal's alpha activity, which is assumed to be connected to cognitive functions including relaxation and concentration.

$$BP_{\alpha} = \frac{\sum_{f=f_{low}}^{f_{high}} P(f)}{f_{high} - f_{low}} \tag{19}$$

In this equation: The alpha band's BP is BP_{α}. f_{low} is the alpha band's lowest frequency limit (8 Hz). The alpha band's upper frequency boundary is f_{high} 12 Hz.

10) BANDPOWERBETA

It describes the frequency domain power of an EEG signal's beta frequency band. Finding the power inside the 13-30 Hz frequency range is the first step in computing the BP. This is known as the beta band.

$$BP_{\beta} = \frac{\sum_{f=f_{low}}^{f_{high}} P(f)}{f_{high} - f_{low}} \tag{20}$$

In this equation: BP_{β} is the beta band's BP, f_{low} is the lower frequency boundary of the beta band (13 Hz), and $P(f)$ is the PSD of the EEG signal at frequency f. The beta band's upper frequency boundary is f_{high} 30 Hz.

11) RATIOBANDPOWERALPHABETA

An effective metric for assessing the balance of alpha and beta activity in an EEG signal is the BP alpha to BP beta ratio. The alpha frequency band's BP divided by the beta frequency band's BP yields the ratio.

$$R_{\alpha/\beta} = \frac{BP_{\alpha}}{BP_{\beta}} \tag{21}$$

In this equation:

The ratio of BP alpha to BP beta is $R_{\alpha/\beta}$. The alpha band's BP, or BP_{α}, is determined by dividing the number of frequency bins in the alpha band by the sum of the PSD inside the alpha frequency band (8-12 Hz). The BP of the beta band, BP_{β}, is determined by dividing the number of frequency bins in the beta band by the sum of the PSD within the beta frequency band (13-30 Hz).

12) RATIOBANDPOWERTHETAALPHA

An effective metric for assessing the balance of theta and alpha activity in an EEG signal is the BP theta to BP alpha ratio. The theta frequency band's BP divided by the alpha frequency band's BP yields the ratio.

$$R_{\theta/\alpha} = \frac{BP_{\theta}}{BP_{\alpha}} \tag{22}$$

In this equation:

The ratio of BP theta to BP alpha is $R_{\theta/\alpha}$. The theta band's BP, or BP_{θ}, is determined by dividing the number of frequency bins in the theta band by the sum of the PSD inside the theta frequency band (4-8 Hz). BP of the alpha band, BP_{α}, is determined by dividing the number of frequency bins in the alpha band by the sum of the PSD inside the alpha frequency band (8-12 Hz).

13) RATIOBANDPOWERTHETAPLUSALPHABETA

An effective metric for assessing the balance of theta+alpha and beta activity in an EEG signal is the ratio of (BP theta + BP alpha) to BP beta. The BP of the alpha and theta frequency bands added together can be divided by the BP of the beta frequency band to determine the ratio.

$$R_{(\theta+\alpha)/\beta} = \frac{BP_{\theta} + BP_{\alpha}}{BP_{\beta}} \tag{23}$$

In this equation:

The ratio of (BP theta + BP alpha) to BP beta is $R_{(\theta+\alpha)/\beta}$. The theta band's BP, or BP_θ, is determined by dividing the number of frequency bins in the theta band by the sum of the power PSD inside the theta frequency band (4-8 Hz). The BP of the alpha band, BP_α, is determined by dividing the number of frequency bins in the alpha band by the sum of the PSD inside the alpha frequency band (8-12 Hz). The BP of the beta band, BP_β, is determined by dividing the number of frequency bins in the beta band by the sum of the PSD within the beta frequency band(13-30 Hz).

14) RATIOBANDPOWERTHETABETA

A suitable metric for measuring the balance of theta and alpha activity in an EEG signal is the BP theta to BP alpha ratio. The ratio is defined as the BP of the theta frequency band divided by the BP of the alpha frequency band.

$$R_{\theta/\beta} = \frac{BP_\theta}{BP_\beta} \tag{24}$$

In this equation:

The ratio of BP theta to BP beta is $R_{\theta/\beta}$. The BP for the theta band, or BP_θ, is calculated by dividing the number of frequency bins in the theta band by the sum of the PSD inside the theta frequency band (4-8 Hz). Similarly, the BP of the beta band, BP_β, is determined by dividing the number of frequency bins in the beta band by the sum of the PSD within the beta frequency band (13-30 Hz).

15) RATIOBANDPOWERTHETAPLUSALPHATHETAPLUSBETA

A reasonable marker of the balance between (theta + alpha) and (theta + beta) activity in an EEG signal is the ratio of (BP theta + BP alpha) to (BP theta + BP beta). The ratio can be found by taking the ratio of the total BP of the alpha and theta frequency bands to the total BP of the beta and theta frequency bands.

$$R_{(\theta+\alpha)/(\theta+\beta)} = \frac{BP_\theta + BP_\alpha}{BP_\theta + BP_\beta} \tag{25}$$

In this equation:

The ratio of (BP theta + BP alpha) to (BP theta + BP beta) is $R_{(\theta+\alpha)/(\theta+\beta)}$. BP, or BP_θ, of the theta band, is found by dividing the number of frequency bins in the theta band by the sum of the PSD inside the theta frequency band (4-8 Hz). BP_α. Calculate the BP for the alpha band, divided by the sum of PSD within the alpha frequency band: 8-12 Hz. Calculate the BP for the beta band, BP_β. Divide the number of frequency bins in the beta band by the sum of the PSD within the frequency band of beta, between 13 and 30 Hz.

E. ENTROPIES

Entropies are utilized for the assessment of EEG signals, measuring signal complexity, as well as for detecting variations in brain electrical activity.

The Shannon, Renyl, and Tsallis entropies are some of the entropies that can be calculated for EEG signals. These entropies, as computed using the amplitude or phase variation of the EEG signal, might be used to identify EEG signal alterations that could be linked to specific neurological conditions or mental states.

1) SHANNONENTROPY

It calculates the amount of uncertainty or unpredictability of a signal. The logarithm of these probabilities multiplied by the total of the probabilities for amplitude or phase value for all EEG signals gives the value of Shannon entropy.

$$H(U) = -\sum_{k=1}^{m} P(u_i) log_2(P(u_i)) \tag{26}$$

The Shannon entropy of the EEG signal is defined as $H(U)$. The probability of the k-th amplitude or phase value of the EEG signal is $P(u_i)$.

2) RENYLENTROPY

A non-extensive generalization of the Shannon entropy, the Renyl entropy measures the amount of uncertainty or lack of predictability in a transmission. Two parameters, α and L, define it and may be used to tune how sensitive the entropy measure is to various signal patterns. Renyl entropy is applied in the analysis of EEG signal complexity and determination of changes in electrical brain activity. The computation of the Renyl entropy thus turns on using the logarithm of the sum of the probabilities of each EEG signal amplitude or phase value raised to the power of the parameter α. Length multiplies.

$$H_{\alpha,L}(U) = L \log_2 \left(\sum_{k=1}^{m} P(u_i)^\alpha \right) \tag{27}$$

In this equation:

The EEG signal's Renyl entropy is $H_{\alpha,L}(U)$. The probability of the EEG signal's k-th amplitude or phase value is $P(u_k)$.

3) TSALLISENTROPY

It is a generalization of the Shannon entropy and is a measure of the degree of uncertainty or unpredictability in a signal. The sensitivity of the entropy measure to various signal pattern types can be changed by varying the parameter q.

Tsallis entropy is calculated by multiplying the difference between the probability and 1 raised to the power of the parameter q by the sum of the probabilities of each amplitude or phase value of the EEG signal.

$$H_q(U) = \frac{1}{q-1} \left[1 - \sum_{k=1}^{n} P(u_i)^q \right] \tag{28}$$

The EEG signal's Tsallis entropy is $H_q(U)$. The probability of the EEG signal's k-th amplitude or phase value is $P(u_k)$. The EEG signal's amplitude or phase values are denoted by m. The sensitivity of the entropy measure is defined by the parameter q.

4) LOGENERGYENTROPY

A metric that combines the signal's energy and entropy information is known as the Log Energy Entropy (LEE). In order to calculate LEE, take the logarithm of the signal's energy and multiply it by the signal's entropy.

$$LEE = log_2(E) \times H(U) \tag{29}$$

In this equation:

E denotes the energy of the EEG signal, computed by adding up the square of all its amplitude values. Then followed by the entropy of the EEG signal, typically determined by multiplying the logarithm with probabilities of all phase or amplitude values.

5) MEANCURVELENGTH

The foundation for mean curve length (MCL) is a measure of the complexity of the EEG signal. It is computed by dividing the total length of the curve and separated by the number of data points. While processing EEG signals, MCL is applied in quantifying the signal's complexity and also the variability of electrical activity of the brain.

$$MCL = \frac{\sum_{k=1}^{m-1} \sqrt{(u_k - u_{k+1})^2 + (v_i - v_{i+1})^2}}{m} \tag{30}$$

In this equation:

MCL represents the EEG signal's Mean Curve Length (u_k, v_k) and (u_{k+1}, v_{k+1}). are the signal's two consecutive data points. m is the signal's number of data points.

6) FIRSTDIFFRENCE

The first difference of an EEG signal is a measurement that represents how quickly the signal changes with respect to time. It can be determined as every signal data point value minus the value immediately preceding it. When analyzing EEG signals, the first difference can be applied to measure dynamic, or variability, or in detecting changes in the electrical activity of the brain.

$$\Delta u_k = u_k - u_{k-1} \tag{31}$$

In this equation:

Δu_k is the EEG signal's first difference at point i, u_k is its value at point k, and u_{k-1} is its value at the point k-1 before it.

7) NORMALIZEFIRSTDIFFERENCE

A normalized first difference of an EEG signal normalizes the first difference in terms of the amplitude of the signal as it measures the rate of change of the signal over time. It is computed as the difference between each subsequent data point of the signal and the past one; then the result is divided by the amplitude of the signal.

$$\Delta u_k = \frac{u_k - u_{k-1}}{A} \tag{32}$$

In this equation:

The normalized initial difference of the EEG signal at location k is Δu_k. The EEG signal value at point k is u_k, and the EEG signal value at the position k-1 before it is u_{k-1}. The EEG signal's amplitude, or A, is determined by subtracting the signal's minimum value from its greatest value.

8) SECONDDIFFERENCE

The second difference of an EEG signal is a reflection of how rapidly the first difference of the signal varies with time. It is derived by subtracting the first difference of each data point of the signal from the first difference of the data point itself. The second difference in EEG signal analysis can be employed to quantify the dynamic or variability of the signal and identify changes in the electrical activity of the brain.

$$\Delta^2 u_k = (u_k - u_{k-1}) - (u_{k-1} - u_{k-2}) \tag{33}$$

In this equation:

$\Delta^2 u_k$ represents the EEG signal's second difference at point k, u_k represents the signal's value at point k, and u_{k-1} represents the EEG signal's value at the point k-1 before it. The EEG signal value at the moment prior to the previous point k-2 is u_{k-2}.

9) NORMALIZEDSECONDDIFFERENCE

The normalized second difference for an EEG signal normalizes the second difference relative to amplitude, recording the rate of changes of the first difference of a signal over time. It is calculated as a subtraction of the first difference of each data point for the signal from that which came before it, minus again, divided by the signal's amplitude.

$$\Delta^2 u_k = \frac{(u_k - u_{k-1}) - (u_{k-1} - u_{k-2})}{A} \tag{34}$$

In this equation:

The normalized second difference of the EEG signal at point k is $\Delta^2 u_k$, the EEG signal value at point k is u_k and the EEG signal value at the point k-1 is u_{k-1}. The value of the EEG signal at time step one before the previous point k-2 is u_{k-2}. The amplitude, or A, of the EEG signal is obtained by subtracting the minimum value of the signal from its maximum value.

10) MEANTEAGERENERGY

A mean teager energy (MTE) operator that uses the square of the first difference of the signal minus the product of the signal and its second difference computes the MTE.

In EEG, the MTE is very useful to determine the non-linear energy in the signal and thus diagnose differences in the electrical activities of the brain.

$$MTE = \frac{1}{m} \sum_{k=1}^{m} (u_k^2 - u_{k-1} u_{k+1}) \tag{35}$$

In this equation:

The value of the EEG signal at position k is u_k. The value of the EEG signal at the previous position k-1 is u_{k-1}. The value of the EEG signal at the subsequent position k+1 is u_{k+1}. The number of data points in the signal is denoted by m.

11) LOGROOT SEQUENTIAL VARIANCE SUM (LRSV)

It computes the Log Root Sum of Sequential Variation, which measures the complexity and non-linearity of the EEG signal, using the logarithm of the square root of the sum of the squared differences between successive data points in the signal.

$$LRSV = log_2(\sqrt{\sum_{k=1}^{m}(u_k - u_{k-1})^2}) \tag{36}$$

In this equation:

The EEG signal value at position k is u_k. The EEG signal value at the previous position k-1 is u_{k-1}. m is the signal's number of data points.

12) MAXIMUM

The maximum value of an EEG signal can be calculated using the following equation:

$$max(u) = \max_{k=1}^{M}(u_k) \tag{37}$$

where:

The discrete time series u represents the EEG signal, which is represented by a series of numbers $u_1, u_2, \ldots, u_M$. The k-th EEG signal sample is denoted by u_k and an EEG signal's total number of samples is M. The mathematical operator $\max_{k=1}^{M}$ yields the highest value in the range of values $u_1, u_2, \ldots, u_M$. The aforementioned formula determines the highest value across all EEG signal samples. Each sample u_k of the EEG signal is examined by the operator $\max_{K=1}^{M}$, which compares it to the highest value discovered thus far. The new maximum value is then designated as the highest sample out of all the samples. Importantly, this formula yields the highest amplitude of signal, which depends on signal values.

13) MINIMUM

The minimum value of an EEG signal can be calculated using the following equation:

$$min(u) = \min_{k=1}^{M}(u_k) \tag{38}$$

where:

The discrete time series u represents the EEG signal, which is represented by a series of numbers $u_1, u_2, \ldots, u_M$. The k-th EEG signal sample is denoted by u_k and an EEG signal's total number of samples is M.

The mathematical operator $\min_{k=1}^{M}$: yields the lowest value in the range of values $u_1, u_2, \ldots, u_M$. This formula determines the EEG signal's minimum value over all samples. Each sample u_k of the EEG signal is examined by the operator $\min_{k=1}^{M}$, which compares it to the lowest value discovered thus far. The new minimum value is then designated as the lowest sample out of all the samples.

F. FEATURE SELECTION

Feature reduction and optimization of EEG data are important for MW evaluation. With high-dimensional EEG data, there are numerous redundant or irrelevant features that may hinder classification accuracy and increase computational expense. We use feature optimization techniques like GEL to find the most discriminative features that distinguish the underlying patterns reflected by varying levels of MW. This process provides some major advantages, such as improved classification accuracy, enhanced model generalization, and better interpretability. Genetic [34], evolutionary [52], and linear search [53] are optimistic to obtain features. Genetic and evolutionary algorithms search the feature space stochastically for the best combinations of features. Identifying the relevant features with these models helps in developing an optimized, accurate, and better MW detection system. An effective model is proposed as the GEL algorithm, which selects optimum features by rejecting the redundant features and selecting the most relevant features for the MW assessment. The working algorithm for the GEL model is explained in algorithms 1-4; each algorithm finds optimal features. The proposed GEL algorithm is the union of all the selected features to obtain the optimal final feature set.

1) GEL MODEL FOR FEATURE OPTIMIZATION

Algorithm 1 describes the genetic search algorithm for feature selection. It initializes the population on the feature subset to evaluate offspring fitness for each chromosome, it replaces the worst with offspring, and it selects the best chromosomes as the best feature. This process will be repeated on all feature subsets till all best chromosomes are selected as best features.

Algorithm 1 Genetic Search for EEG Feature Selection

Require: EEG data $D = \{X, y\}$, Pop. size P, Gen. G, p_c, p_m, Fitness f
Ensure: Feature subset F_G
0: Initialize population *Pop* (binary strings).
0: **for** each chromosome $c \in Pop$ **do**
0: Evaluate fitness $f(c)$.
0: **end for**
0: **for** generation $g = 1$ to G **do**
0: Select parents P_1, P_2.
0: Perform crossover with prob. p_c.
0: Perform mutation on offspring with prob. p_m.
0: Evaluate offspring fitness.
0: Replace worst in *Pop* with offspring.
0: **end for**
0: Select best chromosome c_{best}.
0: $F_G \leftarrow$ Features selected by c_{best}. {Store selected features}
0: **return** $F_G = 0$

Algorithm 2 describes the evolutionary search algorithm, which evaluates fitness for each subset and again evaluates created mutant fitness. It replaces chromosomes with a created mutant to select the best chromosomes as a feature. This process will be repeated for all feature subsets till all the best features get extracted for EEG signals.

Algorithm 2 Evolutionary Search for EEG Feature Selection

Require: EEG data $D = \{X, y\}$, Pop. size P, Gen. G, Mutation rate m, Fitness f
Ensure: Feature subset F_E
0: Initialize population *Pop* (binary strings).
0: **for** each chromosome $c \in Pop$ **do**
0: Evaluate fitness $f(c)$.
0: **end for**
0: **for** generation $g = 1$ to G **do**
0: **for** each chromosome $c \in Pop$
0: Create mutant m_c.
0: Evaluate fitness $f(m_c)$.
0: **if** $f(m_c) > f(c)$ **then**
0: Replace c with m_c.
0: **end if**
0: **end for**
0: **end for**
0: Select best chromosome c_{best}.
0: $F_E \leftarrow$ Features selected by c_{best}. {Store selected features}
0: **return** $F_E = 0$

Algorithm 3 shows the linear search algorithm, which finds the best fitness and the best subset for each subset of features. For each feature and each pair, fitness is calculated for the best subset, which is selected as the best feature. This process is repeated for all feature subsets till the best feature set is selected for EEG signals MWs.

Algorithm 3 Linear Search for EEG Feature Selection

Require: EEG data $D = \{X, y\}$, Feature set F, Fitness f
Ensure: Feature subset F_L
0: Initialize best fitness $best\_f = -\infty$.
0: Initialize best subset $best\_subset = \emptyset$.
0: **for** each feature $f_i \in F$ **do**
0: Create subset $subset = \{f_i\}$.
0: Evaluate fitness $current\_f = f(subset)$.
0: **if** $current\_f > best\_f$ **then**
0: $best\_f \leftarrow current\_f$.
0: $best\_subset \leftarrow subset$.
0: **end if**
0:**end for**
0: **for** each pair $(f_i, f_j) \in F, i \neq j$ **do**
0: Create $subset = \{f_i, f_j\}$.
0: Evaluate fitness $current\_f = f(subset)$.
0: **if** $current\_f > best\_f$
0: $best\_f \leftarrow current\_f$.
0: $best\_subset \leftarrow subset$.
0: **end if**
0: **end for**
0: $F_L \leftarrow best\_subset$. {Store selected features}
0: **return** $F_L = 0$

Algorithm 4 is the GEL model for the final feature set, which unions all features extracted from each search algorithm. This final feature set is then given as input to the classifier for classification.

Algorithm 4 Combined Feature Set and Classification

Require: Feature subsets F_G, F_E, F_L, EEG data $D = \{X, y\}$, Classifier C
Ensure: Classification results
0: $F_{combined} \leftarrow F_G \cup F_E \cup F_L$. {Union of features}
0: $X_{combined} \leftarrow$ Feature matrix from X using features in $F_{combined}$. {Extract combined features from original data}
0: Train classifier C using $X_{combined}$ and y.
0: Evaluate classifier C on a test set.
0: Return classification results. $= 0$

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

EEG signals are recorded using the EMOTIV EPOC+ headset with 14 electrodes and the standard 10-20 system to provide raw data [54]. To find outlier values connected to the peaks of the muscular movement wave, the interquartile range (IQR) filtering technique is used. Signals are further filtered by the headset's built-in filters to improve the quality of EEG signals. The headset provides the quality measure with a number from 0 to 4, which provides the information regarding the contact between the sensor and headset. A quality measure of '4' depicts a good contact, and '0' depicts nil contact. For good quality data and sample size, data with 25% of low-quality signals are excluded. The complete phase of the experiment is excluded if the signal of any of the two sensors is of low quality [54]. Finally, a subject is excluded if workload phases or baseline is discarded during experimental setup. A total of 16 subjects data are selected out of 20 subjects [6].

B. CLASSIFICATION

For the analysis of EEG signals, the RF classifier is employed for MW evaluation. We have tested various machine learning classifiers, which include SVM, NB, Decision tree (DT) and kNN, however, RF yielded the best performance and is, therefore, selected for classification. RF consists of ensemble of DT, which generalizes the model and therefore gives better results on unseen data and it reduces the chances of over-fitting. The proposed method, GEL-RF, which uses the GEL model for feature optimization and RF for classification, provides better results. With regard to this study, the maximum number of tree splits is set at 503,871, allowing for complex decision boundaries within each tree. Furthermore, a total of 100 DT are used to constitute the RF; the predictions of these trees are then combined together by using majority voting. For improving the learning process, a learning rate of 0.1 is applied so as to better influence how the outcome of each tree decided the classification. Results are obtained on multiple parameter settings of the classifier. Such an ensemble classifier is so precise in terms of employing different EEG features to determine the levels of

MW, and on the other hand, the set parameters with DT allow the model to perform well in case it is exposed to newer data. This method has been shown to be very effective for MW assessment in EEG classification tasks.

C. RESULTS

This section provides classification accuracy and comparison of binary and multiclass models for all MW levels using N-back to assess each subject's performance in low, medium, and high workload environments.

D. DATASET

The N-back dataset [6] is used in this research to detect MW assessment. N-Back tests are mentally challenging games that demand task resolution in accordance with a stimulus presented N trials earlier. To create low, medium, and high mental demands, three variations of the n-back tests are employed:

1. Low workload: Every few seconds, a square will appear on the screen in one of eight different places on a standard grid. When a square on the current screen is in the same location as a square in the previous grid, players must press a keyboard key.

2. Medium Workload: Every few seconds, an integer between 0 and 9 will appear on the screen while an audio message will announce an addition, subtraction, multiplication, or division. Players must use the current number and the previous number to solve this task.

3. High Workload: The two previous tests are combined in this one. Every few seconds, an integer between 0 and 9 appears in one of eight different locations on a regular grid. An operator receives an audio message for each number that appears on the screen at the same time. As before, players must use the current number and the number that first appeared twice to solve this operation. Additionally, if the current number's position matches that of the number displayed two screens earlier, players must press a key.

A subject's neurophysiological reaction to mentally challenging tasks depends on his baseline state, which is susceptible to change over time. Prior to the N-back tests, participants watched a calming video for 10 minutes in order to control for variations in the baseline states of the subjects. We conducted a baseline (BL) phase, a video-watching stage, a low workload (WL) stage, a medium workload stage, and a high workload stage for each experiment. As a result, we refer to BL1, BL2, and BL3 as the experiment's baseline phases, while WL1, WL2, and WL3 are the experiment's workload phases. Participants completed a TLX questionnaire after the game to provide their subjective assessments of the workload and level of difficulty. Twenty subjects in all took part in the study. The subjects are all healthy adults between the ages of 20 and 60; none of them had any conditions that might have led to an imbalance in the data that is recorded. Subjects are given a random order of tasks, and each session's recording took place on a different day at a different time [6]. The following sections explain each model and their results subject-wise. Results are shown in Table 2 and Table 3 for the binarized model and multiclass model for all 4 combinations.

In our research on MW Detection and Evaluation, we process EEG signals of the N-back dataset using statistical feature extraction and assess how well they perform in separating three workload classes. To provide a visual representation of the effect of feature selection, we use t-Distributed Stochastic Neighbor Embedding (t-SNE) in 3 plot and compare the two-class distribution before feature (BF) selection and after feature (AF) selection. Prior to feature selection, the t-SNE plots show poor class segregation, which means high overlap and lower discriminative ability in the original feature space. This poor segregation implies the existence of redundant or non-informative features, which can degrade the classification performance of downstream models.

In order to improve class separability, we use our proposed GEL search algorithm for feature selection. The t-SNE plots after feature selection show a much better segregation between the two workload classes, which indicates the efficacy of our method in improving the feature space. This improved representation indicates that the GEL algorithm is able to effectively choose the most discriminative and relevant features, thus improving the discrimination between MW states. By presenting subject-wise class distribution as illustrations, our findings present strong evidence that our approach successfully enhances the interpretability and classification capability of EEG-based workload detection. The better performance of our proposed GEL model supports its potential in developing MW assessment methods.

Table 1 shows the features selected for the entropy-based, time and frequency domain features characterizing mental workload states across neurophysiological domains. Mean, median, skewness, and kurtosis in the time-domain features help in capturing signal asymmetries during cognitive load variations. Mobility and complexity parameters in the time domain quantify a signal's non-stationarity among different workload levels. Frequency domain features, including median, variance, standard deviation, skewness, and kurtosis, are selected to capture spectral variations among different MW levels. Frequency-domain ratios, i.e., $\theta\beta$, $(\theta + \alpha/\theta + \beta)$ are prioritized for their established sensitivity to cortical activation patterns during working memory tasks, particularly in frontal theta (4-8 Hz) and beta (13-30 Hz) bands.

Renyl and Tsallis entropies are selected over Shannon entropy for their parameterized sensitivity to EEG signal changes which enhances the detection of rapid neural dynamics during high workload. Log Energy Entropy quantify the complexity and irregularity of EEG signals, offering insights into non-linearity of signals. Collectively, this optimized feature set maximizes workload separability and yields improved accuracy, as validated by t-SNE visualization in Figure 3.

Figure 4 shows binary classification accuracy comparison of before and after feature selection, With the x-axis representing configurations (BL-WL2, BLs-WL2-WL3,

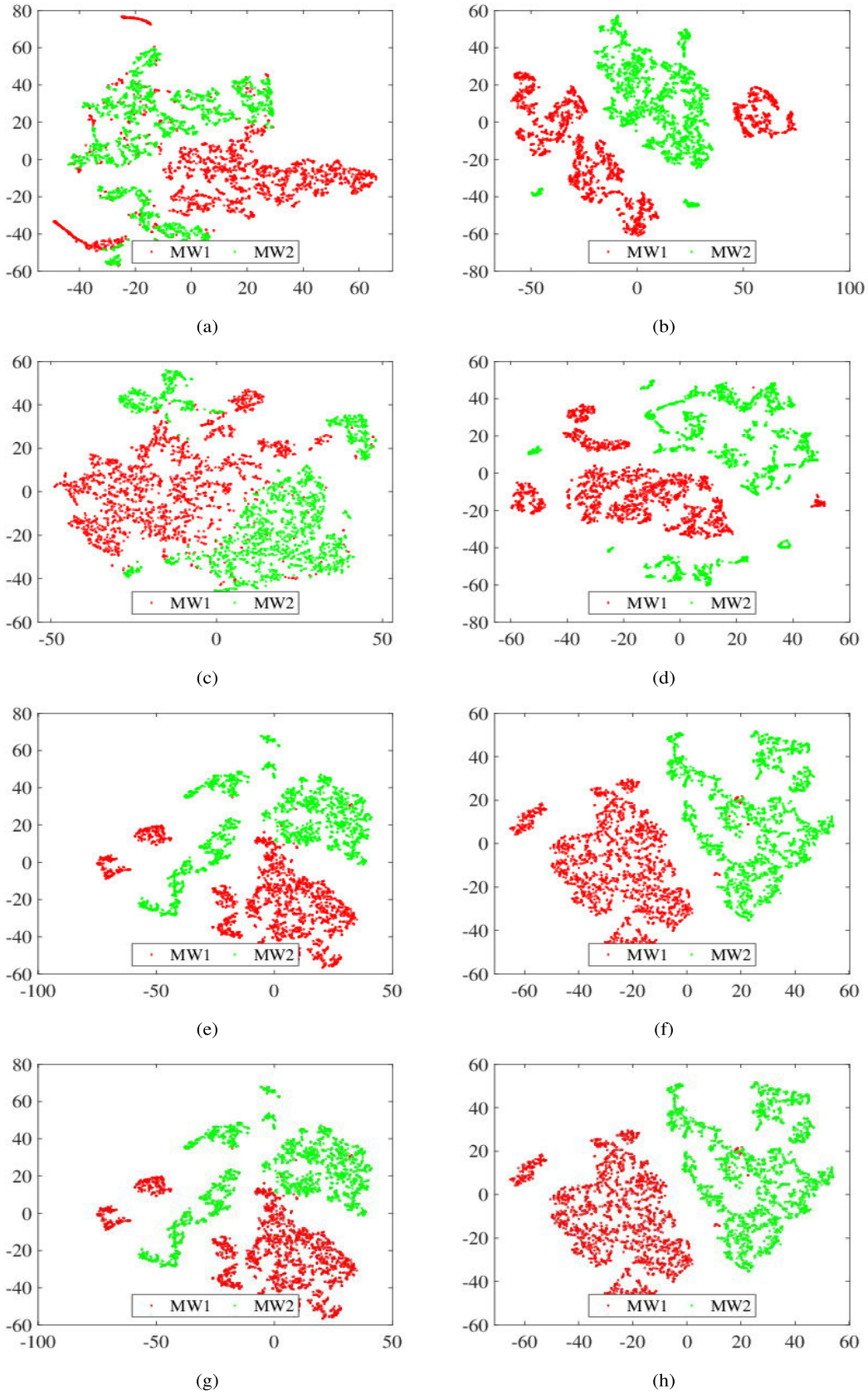

FIGURE 3. TSNE Plots-Comparison of class distribution before and after GEL model. (a), (c), (e), (g) class distribution befor GEL model, (b), (d), (f), (h) class distribution after GEL model.

TABLE 1. Optimized feature set selected by GEL algorithm.

Time Domain Features	Frequency Domain Features	Entropies
Mean	Median	Renyl Entropy
Median	Variance	Tsallis Entropy
Skewness	Standard Deviation	Log Energy Entropy
Kurtosis	Skewness	First Difference
Hjorth Mobility	Kurtosis	Normalized Second Difference
Hjorth Complexity	Ratio Band-Power Theta-Beta	
Mean Energy	Ratio BandPower Theta plus AlphaTheta plus Beta	

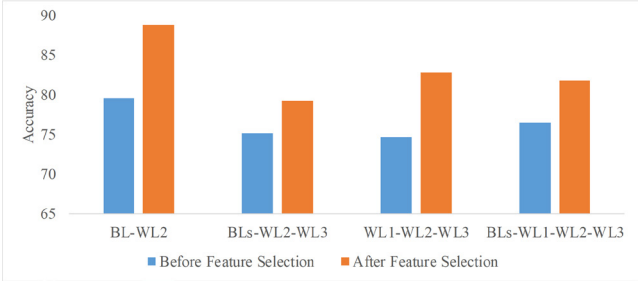

FIGURE 4. Binary class average accuracy comparison before and after feature selection.

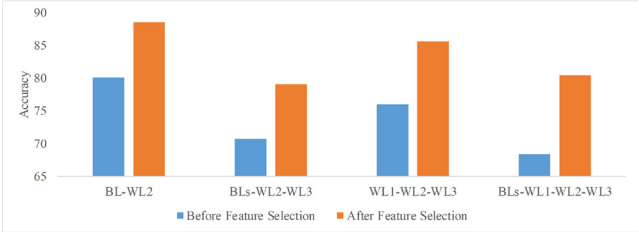

FIGURE 5. Multiclass average accuracy comparison before and after feature selection.

WL1-WL2-WL3 and BLs-WL1-WL2-WL3) and the y-axis showing accuracy (65%-90%), the figure contrasts model accuracies before and after feature selection. All configurations perform better after feature selection, demonstrating how feature reduction can effectively increase model precision. With an increase from 75% to 90%, the BLs-WL1-WL2-WL3 model exhibits the greatest gain, suggesting optimal feature relevance. In a similar way, BL-WL2 shows less noise interference, improving from 65% to 85%. These findings demonstrate how strategic feature selection improves accuracy by giving useful variables priority, with configuration-specific effects highlighting customized methods for the best possible model refining.

Figure 5 shows multiclass classification accuracy comparison of before and after feature selection in which accuracy is represented by the vertical axis and is scaled from 65 to 90 in steps of 5. The performance of four combination configurations—BL-WL2, BLs-WL2-WL3, WL1-WL2-WL3 and BLs-WL1-WL2-WL3—is compared both before and after feature selection.

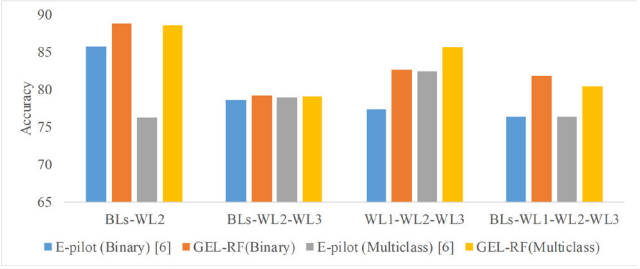

FIGURE 6. Comparison for all combinations using GEL-RF with N-back dataset.

The findings demonstrate how feature selection's efficacy varies throughout models with varying structures. While more intricate designs (BLs-WL1-WL2-WL3) show how feature selection enhancing results by reducing dimensionality, simpler combination (BL-WL2) show noticeable performance adjustments. This comparison highlights how crucial it is to adjust feature selection tactics to model complexity.

Figure 6 shows the relative accuracy of four ensemble-based classification techniques in different ensemble configurations: e-pilot (Binary), GEL-RF (Binary), e-pilot (Multiclass), and GEL-RF (Multiclass). While the x-axis separates four different configurations—BLs-WL2, BLs-WL2-WL3, WL1-WL2-WL3 and BLs-WL1-WL2-WL3—that reflect incremental combinations of base lines (BLs) and work load levels (WL1, WL2, WL3), the y-axis shows accuracy percentages ranging from 65% to 90%.

Important findings include: Impact of Ensemble Complexity: Adding more weak learners to the ensemble (e.g., switching from BLs-WL2 to BLs-WL1-WL2-WL3) usually improves accuracy, especially for multiclass tasks. Performance in Binary vs Multiclass: N-back and GEL-RF, two binary classification techniques, routinely perform better than their multiclass counterparts, with GEL-RF (Binary) having the greatest accuracy across all configurations.

The most reliable findings are obtained with the BLs-WL1-WL2-WL3 configuration, highlighting the synergistic impact of combining base learners with several weak learners. With GEL-RF (Binary) emerging as the most successful method for high-accuracy tasks, our analysis emphasizes the crucial role that ensemble design plays in striking a balance between model complexity and predictive performance.

TABLE 2. Random forest classifier results of binary mental workload assessment.

SUBJECTS	BL-WL2 (BF)	BL-WL2 (AF)	BLs-WL2 -WL3(BF)	BLs-WL2 -WL3(AF)	WL1-WL2 -WL3(BF)	WL1-WL2 -WL3(AF)	BLs-WL1 -WL2-WL3(BF)	BLs-WL1 -WL2-WL3(AF)
Subject 1	77.3%	**83.5%**	69.6%	**70.6%**	75.1%	**79.2%**	75.5%	**79.4%**
Subject 2	77.5%	**89.6%**	64.9%	**66.7%**	63.5%	**87.8%**	69.3%	**80.3%**
Subject 3	79.4%	**85.7%**	73.1%	**75.3%**	72.8%	**76.7%**	74.2%	**81.0%**
Subject 4	83.5%	**87.9%**	75.5%	**77.9%**	76.9%	**87.6%**	79.5%	**84.6%**
Subject 5	70.2%	**82.9%**	63.1%	**65.3%**	62.5%	**80.1%**	65.3%	**76.0%**
Subject 6	83.4%	**87.9%**	71.5%	**80.0%**	70.6%	**76.8%**	72.2%	**81.5%**
Subject 7	80.1%	**90.6%**	82.0%	**84.8%**	82.4%	**70.5%**	77.3%	**84.6%**
Subject 8	86.9%	**94.5%**	79.8%	**84.2%**	76.2%	**77.3%**	79.5%	**87.0%**
Subject 9	81.5%	**89.6%**	79.6%	**82.1%**	82.5	**72.7%**	82.6%	**85.1%**
Subject 10	76.4%	**90.3%**	79.2%	**83.3%**	78.7%	**92.1%**	81.7%	**65.5%**
Subject 11	79.8%	**88.9%**	75.6%	**87.3%**	79.3%	**97.0%**	82.6%	**88.9%**
Subject 12	76.5%	**83.9%**	81.0%	**85.4%**	81%	**85.4%**	78.3%	**80.7%**
Subject 13	78.6%	**89.3%**	73.0%	**74.9%**	76.6%	**79.5%**	78.2%	**79.5%**
Subject 14	81.9%	**92.6%**	77.5%	**79.4%**	74.5%	**82.9%**	77.3%	**81.9%**
Subject 15	72.8%	**87.3%**	68.9%	**79.9%**	66.4%	**82.8%**	69.3%	**81.5%**
Subject 16	91.3%	**96.3%**	88.6%	**90.4%**	76.7%	**95.5%**	81.5%	**91.0%**
AVERAGE	79.57%	**88.80%**	75.18%	**79.22%**	74.73%	**82.74%**	76.52%	**81.78%**
STD	5.17	**3.63**	6.94	**6.46**	6.00	**5.47**	5.04	5.61
Hitcount	0	16	0	16	0	16	0	16

TABLE 3. Random Forest classifier results of multiclass mental workload assessment.

SUBJECTS	BL-WL2 (BF)	BL-WL2 (AF)	BLs-WL2 -WL3(BF)	BLs-WL2 -WL3(AF)	WL1-WL2 -WL3(BF)	WL1-WL2 -WL3(AF)	BLs-WL1 -WL2-WL3(BF)	BLs-WL1 -WL2-WL3(AF)
Subject 1	77.3%	**89.6%**	74.3%	**74.4%**	63.1%	**83.7%**	48.7%	**79.3%**
Subject 2	74.3%	**90.3%**	75.4%	**73.1%**	69.7%	**89.8%**	65%	**80.3%**
Subject 3	79.3%	**88.9%**	75.5%	**78.8%**	81.4%	**76.7%**	69.4%	**75.9%**
Subject 4	83.5%	**82.6%**	76.1%	**79.0%**	78.4%	**86.5%**	68.1%	**76.7%**
Subject 5	70.2%	**89.3%**	51.7%	**70.6%**	60.3%	**79.1%**	65.4%	**79.4%**
Subject 6	83.3%	**92.6%**	71.1%	**79.3%**	77.9%	**81.6%**	66.7%	**78.7%**
Subject 7	81.6%	**87.3%**	76.5%	**83.4%**	83.7%	**81.6%**	73.2%	**77.6%**
Subject 8	86.9%	**96.3%**	78%	**82.3%**	79.2%	**91.2%**	79.1%	**82.4%**
Subject 9	81.3%	**83.5%**	56.8%	**78.1%**	63.4%	**90.2%**	76%	**79.2%**
Subject 10	76.5%	**88.6%**	58%	**81.8%**	64.8%	**91.4%**	75.7%	**78.9%**
Subject 11	79.8%	**85.9%**	77.1%	**86.8%**	92%	**96.1%**	78%	**86.7%**
Subject 12	78%	**86.5%**	75.5%	**76.3%**	60.8%	**84.7%**	44.7%	**77.7%**
Subject 13	82.7%	**83.2%**	57.6%	**73.5%**	75%	**77.7%**	67.1%	**80.5%**
Subject 14	82.1%	**87.9%**	75.7%	**80.2%**	88.8%	**90.2%**	71.7%	**82.3%**
Subject 15	72.9%	**90.6%**	63.9%	**78.7%**	82.6%	**72.0%**	65.6%	**84.3%**
Subject 16	91.2%	**94.5%**	87.7%	**89.4%**	96%	**97.6%**	79.1%	**87.3%**
AVERAGE	80.06%	**88.6%**	70.7%	**79.10%**	76.06%	**85.63%**	68.34%	**80.45%**
STD	5.09	**3.74**	9.64	**4.83**	11.01	**7.01**	9.52	**3.24**
Hitcount	0	16	0	16	0	16	0	16

TABLE 4. Results of student t-test for statistical significance.

Combinations	P-Value(Binary)	P-V(Multiclass)
BLs-WL2	4.822E-08	3.590E-05
BLs-WL2-WL3	2.292E-04	78.558E-04
WL1-WL2-WL3	6.039E-03	5.635E-03
BLs-WL1-WL2-WL3	6.051E-03	7.359E-05

1) BINARY MODEL CLASSIFICATION

In this model our analysis is based on a binary problem in which all baseline data is considered as a single class and all other workload is considered as a second class. Computation is done as a binary problem in a multiclass setting as

a) BL-WL2: In the first combination, baseline data, which is recorded in a relaxation state, is considered as a single class, and WL2 data is taken as the other class.

b) BLs-WL2-WL3: In this combination all baseline session data is taken as a single class, and other workload levels WL2-WL3 are combined to make another class.

TABLE 5. Comparison of multiclass and binary classification with the benchmark dataset.

Combinations	N-back Dataset (Binary)	GEL-RF (Binary)	N-back Dataset (Multiclass)	GEL-RF (Multiclass)
BLs-WL2	85.72±7.52	88.8±3.63	76.25±19.27	88.6±3.74
BLs-WL2-WL3	78.62±16.59	79.2±6.94	79.0±9.22	79.1±4.83
WL1-WL2-WL3	77.34±16.72	82.7±7.47	82.47±45.78	85.6±7.01
BLs-WL1-WL2-WL3	76.44±16.81	81.78±5.61	76.34±15.78	80.45±3.24

c) WL1-WL2-WL3: In the third case, WL1 data is taken as the baseline class, and the combination of WL2 and WL3 is combined to make the other class.

d) BLs-WL1-WL2-WL3: In combination four, all baseline data is considered as a single class, and the combination of all other workloads (WL1-WL2-WL3) data is combined to make a second class.

The binary classification model using the RF classifier attains an average accuracy of MW detection of 88.80% with a low standard deviation of 3.63 and having 97.0% as the highest accuracy, which shows that the proposed methodology provides better results. In Table 2 it shows binary model classification results in which high accuracy subjects of different combinations are highlighted against all MW levels, with the maximum. As the MW detection varies across subjects due to the variation of responses of subjects, our model performance is adequate for all subjects in different MW levels.

Table 2 shows the results over different scenarios before and after applying the GEL feature selection algorithm on the dataset. It can be observed that the performance improves after applying GEL. The average accuracy of BL-WL2 is 88.80% in comparison to without applying GEL, which shows an increase of 9.23%. The BLs-WL2-WL3 shows an average accuracy of 79.22% in comparison to 75.18% before applying GEL. This shows an increase of 4.04%. WL1-WL2-WL3 shows an average accuracy of 82.74%, which is an 8.01% increase of from without using GEL. BLs-WL1-WL2-WL3 shows an average accuracy of 81.78%, which is a 5.26% increase from without using GEL. Table 2 also shows that there is a decrease in standard deviation values for all the scenarios when the GEL model is applied.

2) MULTICLASS MODEL CLASSIFICATION

In this model our analysis is based on a multiclass problem in which all baseline data is considered as a single class and all other workloads are considered as other classes. Computations are done as a multiclass problem on multiple classes:

a) BL-WL2: In the first combination, baseline data, which is recorded in a relaxation state, is considered as a single class, and WL2 data is taken as the other class, as this model is a binary problem in multiclass settings.

b) BLs-WL2-WL3: In this combination all baseline session data is taken as single class and other workload levels WL2-WL3 are 2 other classes of medium and high workload, which makes it a three-class problem.

c) WL1-WL2-WL3: In the third case WL1 data is taken as the baseline class, and the combination of WL2 and WL3 are other classes, as medium and high workloads, making it a three-class problem.

d) BLs-WL1-WL2-WL3: In combination four all baseline data is considered as a single class, and all other workloads, WL1-WL2-WL3 are 3 other classes, making it a four-class problem.

In the multiclass model, results are also improved using the RF classifier, which gives MW detection of an average of 88.6% and a standard deviation of 3.24, which is quite low as compared to the benchmark dataset results. Accuracy gets better, and achievement of low standard deviation is because of the GEL model, which produces an optimized feature set of the N-back dataset, which makes the GEL-RF model a virtuoso model, unlike traditional classification models.

Table 3 shows the comparison of multiclass classification before and after feature selection. BL-WL2 shows an accuracy score of 88.6%, which is 8.54% higher than when GEL is not utilized. The standard deviation is also reduced by a significant amount when GEL is utilized. BLs-WL2-WL3 shows an accuracy of 79.10% when GEL is utilized. With GEL, this accuracy increases to 8.4%. WL1-WL2-WL3 shows an accuracy of 85.63% with GEL methodology. When GEL is not utilized, this accuracy falls to 76.06%. The standard deviation also increases from 7.01 to 11.01 when GEL is not utilized. BLs-WL1-WL2-WL3 shows an accuracy of 68.34%, when GEL is utilized. The accuracy increases by 80.45% to and the standard deviation improves from 9.52 to 3.24 after utilizing the GEL model.

E. STATISTICAL SIGNIFICANCE

Table 4 presents the statistical significance test, which shows that results after feature optimization are improved a lot compared to the performance before feature optimization. It means the feature selected through the GEL model represents effective EEG parameters, which translate different levels of workload in the N-back low-high pressure task effectively. All combination statistical significance tests for binary and multiclass give a p-value less than 0.05. Which shows our hypothesis for feature optimization is correct and the null hypothesis is rejected.

TABLE 6. Performance comparison of proposed GEL-RF with existing techniques.

Method (Year)	Binary Classfication Accuracy (%)	Multiclass Classification Accuracy (%)	Feature Selection
SVM-RBF (Aksu et al., 2024) [19]	83.4	78.2	Relief-F
Brain Connectivity (Safari et al., 2024) [20]	88.1	82.5	Hierarchical algorithm
3D-CNN (Kwak et al., 2022) [2]	91.5	87.3	Multilevel fusion (no selection)
Transfer Learning (Zhou et al., 2022) [21]	89.2	84.6	Domain adaptation
EEG-FNIRS Fusion (Liu et al., 2022) [36]	85.7	80.1	Manual selection
Bandpower (Hernandez-Sabate et al., 2022) [6]	85.7	76.3	Bandpower features
GEL-RF (Proposed)	**97.0**	**96.3**	**Hybrid GEL**

F. DISCUSSION

MW assessment is critical for optimizing human performance and preventing accidents in high-stakes environments such as aviation, healthcare, and industrial operations. EEG signals have emerged as a reliable tool for MW evaluation due to their high temporal resolution and direct correlation with cognitive states. However, challenges such as noise, non-stationarity, and individual variability in EEG signals necessitate robust feature extraction and selection methods to ensure accurate classification. Traditional approaches often rely on subjective assessments (e.g., NASA-TLX) or physiological metrics with limited generalizability, underscoring the need for automated, objective systems leveraging machine learning.

Prior research has explored diverse methodologies for MW assessment, including subjective questionnaires, multimodal physiological measurements (e.g., ECG, eye tracking), and machine learning models such as SVM, ANN, and RF. While these studies achieved moderate success, limitations persist. For instance, SVM-based models struggle with high-dimensional EEG data, and ANN architectures often act as ''closed boxes,'' compromising interpretability. Feature engineering remains a bottleneck, with many studies relying on manual or filter-based selection, leading to redundant features and suboptimal performance. Recent advancements in deep learning, such as 3D CNNs, improved accuracy but incurred high computational costs. Additionally, benchmark datasets like the N-back task reported accuracies of 76-85% with significant inter-subject variability, highlighting the need for robust feature optimization frameworks.

Table 5 shows the comparison result of the binary and multiclass models with the benchmark dataset results, which clearly shows that the GEL-RF model performs better than the traditional algorithms and achieves a maximum accuracy of 97.0% and the lowest standard deviation of 3.24. All combinations in both binary and multiclass configurations achieve maximum accuracy and eliminate the deviation among subject performance as well. All combination results show that previous deviation among subjects is high on average, which questions the model generalizability, but our model reduced the performance deviation among the subjects to the highest extent.

Table 6 shows the performance comparison of the proposed GEL-RF model with other techniques on the N-back dataset. The proposed method achieves an accuracy of **97.0%** for binary class problem and **96.3%** for multiclass problem. It outperforms recent band-power [6] and brain connectivity models [20] in binary and multiclass tasks. The hybrid GEL feature selection eliminates redundant features while enhancing discriminative power, unlike filter-based methods e.g., Relief-F in [19]. GEL-RF's hybrid feature selection reduces dimensionality, enhances class separability as evident from t-SNE plots in Figure 3, and offers improved accuracy across 16 subjects for both binary and multiclass tasks.

Testing is done on all combinations in binary and multiclass classification, where the GEL model ensures to have the finest feature set for the classifier and results presented. A further point of focus is feature reduction, which is done during feature optimization, in which a significant number of features are reduced in each combination for all subjects but still achieving the highest accuracy of 97.0%. Comparison table 5 shows that the RF classifier with the combination of the GEL model achieves higher accuracy than the other models.

Related literature shows that RF models are broadly used in MW assessment using EEG signals. These studies conclude that RF classifier' yields improved accuracy for distinguishing between MW levels [55], [56].

The proposed GEL-RF model achieved superior performance, attaining a maximum classification accuracy of 97.0% for binary tasks and 96.3% for multiclass tasks, with notably low standard deviations (3 and 3.24, respectively). As shown in Table 4, the GEL-RF outperformed benchmark datasets across all combinations. The t-SNE visualization in Figure 3 further validated these results, showing distinct class separation after GEL-based feature selection, unlike the overlapping clusters in the original feature space.

The performance gains stem from two key innovations,

- GEL Feature Selection: The hybrid GEL algorithm efficiently eliminates redundant features while retaining discriminative ones. By stochastically exploring the feature space, GEL reduces the curse of dimensionality. This is evident in the t-SNE plots, where post-selection features yielded class boundaries, reducing intra-class variability.
- RF Classifier: RF's ensemble structure, combining 100 DT with majority voting, inherently reduces overfitting. The model's ability to handle non-linear

relationships in EEG data, coupled with parameter tuning (e.g., 503,871 tree splits, 0.1 learning rate), enabled robust generalization across subjects. Furthermore, the union of features from GEL's three search strategies enriched the feature set, capturing both global and local patterns in EEG signals.

Despite its advancements, this study has several limitations. (a) Hardware Dependency: EEG data were collected using the EMOTIV EPOC+ headset, which has lower spatial resolution (14 channels) compared to clinical-grade systems. Performance may vary with other devices. (b) Computational Overhead: While RF is computationally efficient, the GEL algorithm's hybrid search strategy could increase training time for larger datasets. (c) Real-World Applicability: The N-back task, though widely used, simulates cognitive load in controlled settings. Future work should test the model in dynamic, real-world environments (e.g., piloting simulations).

V. CONCLUSION

In this paper, our aim is to show the significance of statistical features in MW assessment using EEG signals with optimal feature optimization. Another problem addressed is to have a generalized model for binary and multiclass classification of cognitive workload using proposed model. We have proposed GEL-RF model for the optimal feature selection and optimization of statistical features and RF for classification which is performed on N-back dataset [6].

To support the classification results, the statistical test is also carried out, which depicts a high correlation between MW level and EEG features. Results show the proposed model provides an optimal feature set for N-back EEG signals and acquires the highest accuracy of 97.0% and a low standard deviation of 3.24. These results show the significance of statistical feature extraction, optimal feature selection, and utilization for efficient classification. This model provides better performance in the detection of MW in both binary and multiclass task problems, which proves its generability for classification. In EEG signal preprocessing and feature extraction-selection for MW assessment, complexity increases as the intensity of N-back inflammation increases, which needs a few improvements. One factor of consideration is important, as a lot of user state factors need to be considered when developing algorithms for identifying cognitive state. In this case, an important improvement is the use of ensemble models, which process each sensor individually but use different learned weights. More contemporary architectures include convolutional networks combined with LSTM and also Lambda Networks that can incorporate attention. Further analysis can be done in this direction to look at MW assessment in more tough scenarios.

REFERENCES

[1] T. I. Laine, K. W. Bauer, J. W. Lanning, C. A. Russell, and G. F. Wilson, "Selection of input features across subjects for classifying crewmember workload using artificial neural networks," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 32, no. 6, pp. 691–704, Nov. 2002.

[2] Y. Kwak, K. Kong, W.-J. Song, B.-K. Min, and S.-E. Kim, "Multilevel feature fusion with 3D convolutional neural network for EEG-based workload estimation," *IEEE Access*, vol. 8, pp. 16009–16021, 2020.

[3] J. Ting, A. Del Vecchio, D. Friedenberg, M. Liu, C. Schoenewald, D. Sarma, J. Collinger, S. Colachis, G. Sharma, D. Farina, and D. J. Weber, "A wearable neural interface for detecting and decoding attempted hand movements in a person with tetraplegia," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 1930–1933.

[4] F. G. W. C. Paas and J. J. G. Van Merriënboer, "Instructional control of cognitive load in the training of complex cognitive tasks," *Educ. Psychol. Rev.*, vol. 6, no. 4, pp. 351–371, Dec. 1994.

[5] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educ. Psychologist*, vol. 38, no. 1, pp. 63–71, Jan. 2003.

[6] A. Hernández-Sabaté, J. Yauri, P. Folch, M. À. Piera, and D. Gil, "Recognition of the mental workloads of pilots in the cockpit using EEG signals," *Appl. Sci.*, vol. 12, no. 5, p. 2298, Feb. 2022.

[7] I. Albuquerque, A. Tiwari, M. Parent, R. Cassani, J.-F. Gagnon, D. Lafond, S. Tremblay, and T. H. Falk, "WAUC: A multi-modal database for mental workload assessment under physical activity," *Frontiers Neurosci.*, vol. 14, Dec. 2020, Art. no. 549524.

[8] D. A. Wiegmann and S. A. Shappell, *A Human Error Approach To Aviation Accident Analysis: The Human Factors Analysis and Classification System*. Evanston, IL, USA: Routledge, 2017.

[9] Z. Yin and J. Zhang, "Cross-session classification of mental workload levels using EEG and an adaptive deep learning model," *Biomed. Signal Process. Control*, vol. 33, pp. 30–47, Mar. 2017.

[10] R. Hefron, B. Borghetti, C. Schubert Kabban, J. Christensen, and J. Estepp, "Cross-participant EEG-based assessment of cognitive workload using multi-path convolutional recurrent neural networks," *Sensors*, vol. 18, no. 5, p. 1339, Apr. 2018.

[11] T. Heine, G. Lenis, P. Reichensperger, T. Beran, O. Doessel, and B. Deml, "Electrocardiographic features for the measurement of drivers' mental workload," *Appl. Ergonom.*, vol. 61, pp. 31–43, May 2017.

[12] S.-Y. Han, N.-S. Kwak, T. Oh, and S.-W. Lee, "Classification of pilots' mental states using a multimodal deep learning network," *Biocybernetics Biomed. Eng.*, vol. 40, no. 1, pp. 324–336, Jan. 2020.

[13] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, "A review on mental stress assessment methods using EEG signals," *Sensors*, vol. 21, no. 15, p. 5043, Jul. 2021.

[14] S. Wang, J. Gwizdka, and W. A. Chaovalitwongse, "Using wireless EEG signals to assess memory workload in the n-back task," *IEEE Trans. Human-Machine Syst.*, vol. 46, no. 3, pp. 424–435, 2015.

[15] S. Mohdiwale, M. Sahu, G. R. Sinha, and V. Bhateja, "Statistical wavelets with harmony search- based optimal feature selection of EEG signals for motor imagery classification," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14263–14271, Jul. 2021.

[16] J. Wang, Z. Feng, N. Lu, L. Sun, and J. Luo, "An information fusion scheme based common spatial pattern method for classification of motor imagery tasks," *Biomed. Signal Process. Control*, vol. 46, pp. 10–17, Sep. 2018.

[17] D. Damos, *Multiple Task Performance*. Boca Raton, FL, USA: CRC Press, 1991.

[18] K. W. Bauer, S. G. Alsing, and K. A. Greene, "Feature screening using signal-to-noise ratios," *Neurocomputing*, vol. 31, nos. 1–4, pp. 29–44, Mar. 2000.

[19] Ş. H. Aksu, E. Çakıt, and M. Dağdeviren, "Mental workload assessment using machine learning techniques based on EEG and eye tracking data," *Appl. Sci.*, vol. 14, no. 6, p. 2282, Mar. 2024.

[20] M. Safari, R. Shalbaf, S. Bagherzadeh, and A. Shalbaf, "Classification of mental workload using brain connectivity and machine learning on electroencephalogram data," *Sci. Rep.*, vol. 14, no. 1, p. 9153, Apr. 2024.

[21] Y. Zhou, Z. Xu, Y. Niu, P. Wang, X. Wen, X. Wu, and D. Zhang, "Cross-task cognitive workload recognition based on EEG and domain adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 50–60, 2022.

[22] S. G. Hart, "Nasa-task load index (NASA-TLX); 20 years later," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50. Newbury Park, CA, USA: Sage, 2006, pp. 904–908.

[23] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehavioral Rev.*, vol. 44, pp. 58–75, Jul. 2014.

[24] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task load Index): Results of empirical and theoretical research," in *Advances in Psychology*, vol. 52. Amsterdam, The Netherlands: Elsevier, 1988, pp. 139–183.

[25] C. D. Wickens, "Situation awareness and workload in aviation," *Current Directions Psychol. Sci.*, vol. 11, no. 4, pp. 128–133, Aug. 2002.

[26] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs," *J. Cognit. Eng. Decis. Making*, vol. 2, no. 2, pp. 140–160, Jun. 2008.

[27] Z. Wang, Y. Lei, and J.-S. Ding, "Application of heart rate variability in evaluation of mental workload," *Chin. J. Ind. Hygiene Occupational Diseases*, vol. 23, no. 3, pp. 182–184, 2005.

[28] N. A. Stanton, P. M. Salmon, L. A. Rafferty, G. H. Walker, C. Baber, and D. P. Jenkins, *Human Factors Methods: a Practical Guide for Engineering and Design*. Boca Raton, FL, USA: CRC Press, 2017.

[29] E.-H. Jang, B. Park, S. Kim, M.-A. Chung, M. Park, and J. Sohn, "Classification of human emotions from physiological signals using machine learning algorithms," in *Proc. 6th Int. Conf. Adv. Comput.-Hum. Interact. (ACHI)*, 2013, pp. 395–400.

[30] B. Rim, N.-J. Sung, S. Min, and M. Hong, "Deep learning in physiological signal data: A survey," *Sensors*, vol. 20, no. 4, p. 969, Feb. 2020.

[31] D. Grimes, D. S. Tan, S. E. Hudson, P. Shenoy, and R. P. N. Rao, "Feasibility and pragmatics of classifying working memory load with an electroencephalograph," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2008, pp. 835–844.

[32] E. van Weelden, M. Alimardani, T. J. Wiltshire, and M. M. Louwerse, "Aviation and neurophysiology: A systematic review," *Appl. Ergonom.*, vol. 105, Nov. 2022, Art. no. 103838.

[33] P. Zhang, X. Wang, J. Chen, W. You, and W. Zhang, "Spectral and temporal feature learning with two-stream neural networks for mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1149–1159, Jun. 2019.

[34] I. Stancin, M. Cifrek, and A. Jovic, "A review of EEG signal features and their application in driver drowsiness detection systems," *Sensors*, vol. 21, no. 11, p. 3786, May 2021.

[35] R. Haloi, J. Hazarika, and D. Chanda, "Selection of appropriate statistical features of EEG signals for detection of Parkinson's disease," in *Proc. Int. Conf. Comput. Perform. Eval. (ComPE)*, Jul. 2020, pp. 761–764.

[36] Y. Liu, H. Ayaz, and P. A. Shewokis, "Multisubject 'learning' for mental workload classification using concurrent EEG, fNIRS, and physiological measures," *Frontiers human Neurosci.*, vol. 11, p. 389, Apr. 2017.

[37] V. Jusas and S. G. Samuvel, "Classification of motor imagery using a combination of user-specific band and subject-specific band for brain–computer interface," *Appl. Sci.*, vol. 9, no. 23, p. 4990, Nov. 2019.

[38] Z. Pei, H. Wang, A. Bezerianos, and J. Li, "EEG-based multiclass workload identification using feature fusion and selection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[39] M. Kaczorowska, M. Plechawska-Wójcik, and M. Tokovarov, "Interpretable machine learning models for three-way classification of cognitive workload levels for eye-tracking features," *Brain Sci.*, vol. 11, no. 2, p. 210, Feb. 2021.

[40] Y. Wu, Z. Liu, M. Jia, C. C. Tran, and S. Yan, "Using artificial neural networks for predicting mental workload in nuclear power plants based on eye tracking," *Nucl. Technol.*, vol. 206, no. 1, pp. 94–106, Jan. 2020.

[41] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 701–711, May 2005.

[42] G. R. J. Hockey, "Performance degradation," in *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks*, vol. 355. U.K: Oxford University Press, 2003, p. 8.

[43] M. Borys, M. Plechawska–Wójcik, M. Wawrzyk, and K. Wesołowska, "Classifying cognitive workload using eye activity and EEG features in arithmetic tasks," in *Proc. Inf. Softw. Technologies: 23rd Int. Conf.*, Druskininkai, Lithuania, Oct. 2017, pp. 90–105.

[44] M. Kaczorowska, M. Wawrzyk, and M. Plechawska–Wójcik, "Binary classification of cognitive workload levels with oculography features," in *Proc. 19th Int. Conf. Comput. Inf. Syst. Ind. Management*, Bialystok, Poland, Oct. 2020, pp. 243–254.

[45] H. Qu, Y. Shan, Y. Liu, L. Pang, Z. Fan, J. Zhang, and X. Wanyan, "Mental workload classification method based on EEG independent component features," *Appl. Sci.*, vol. 10, no. 9, p. 3036, Apr. 2020.

[46] W. L. Lim, O. Sourina, and L. P. Wang, "STEW: Simultaneous task EEG workload data set," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2106–2114, Nov. 2018.

[47] S. Ş. Karacan, H. M. Saraoğlu, S. C. Kabay, G. Akdağ, C. Keskinkılıç, and M. Tosun, "EEG-based mental workload estimation of multiple sclerosis patients," *Signal, Image Video Process.*, vol. 17, no. 7, pp. 3293–3301, Oct. 2023.

[48] M. Plechawska-Wójcik, M. Tokovarov, M. Kaczorowska, and D. Zapała, "A three-class classification of cognitive workload based on EEG spectral data," *Appl. Sci.*, vol. 9, no. 24, p. 5340, Dec. 2019.

[49] A. S. Le, H. Aoki, F. Murase, and K. Ishida, "A novel method for classifying driver mental workload under naturalistic conditions with information from near-infrared spectroscopy," *Frontiers Human Neurosci.*, vol. 12, p. 431, Oct. 2018.

[50] S. B. Shafiei, S. Shadpour, and A. Shafqat, "Mental workload evaluation using weighted phase lag index and coherence features extracted from EEG data," *Brain Res. Bull.*, vol. 214, Aug. 2024, Art. no. 110992.

[51] S. Khalfallah, W. Puech, M. Tlija, and K. Bouallegue, "Exploring the effectiveness of machine learning and deep learning techniques for EEG signal classification in neurological disorders," *IEEE Access*, vol. 13, pp. 17002–17015, 2025.

[52] D. D. Chakladar, S. Dey, P. P. Roy, and D. P. Dogra, "EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm," *Biomed. Signal Process. Control*, vol. 60, Jan. 2020, Art. no. 101989.

[53] F. M. A. D. Heide, "A polynomial linear search algorithm for the n-dimensional knapsack problem," *J. ACM (JACM)*, pp. 668–676, 1983.

[54] *Emotiv Epoc+ 14-channel Wireless EEG Headset*, Emotiv, San Francisco, CA, USA, 2020.

[55] Y. Liu, H. Ayaz, and P. A. Shewokis, "Multisubject mental workload classification using random forest and EEG features," *Frontiers Human Neurosci.*, vol. 11, p. 389, May 2017.

[56] J. Zhang, J. Li, and Z. Y. Wang, "Mental workload classification using random forest and EEG data in a simulated driving task," *J. Neural Eng.*, vol. 17, no. 2, 2020, Art. no. 026013.

WALEED MANZOOR received the bachelor's degree in electrical engineering from Bahria University, Islamabad, Pakistan, in 2012, and the master's degree from UET Taxila, Pakistan, in 2017. His research interests include artificial intelligence, image processing, and signal processing.

NOMAN NASEER (Senior Member, IEEE) received the bachelor's degree in mechatronics engineering from the NUST College of EME, Rawalpindi, the master's degree from Air University, Islamabad, Pakistan, and the Ph.D. degree in mechatronics engineering from Pusan National University, Busan, South Korea. He is currently the Director of Neuroimaging Research Group and the Chair of the Department of Mechatronics and Biomedical Engineering, Air University. He has published more than 200 peer-reviewed papers. His name has been included in the list of top 2% scientist of The World by Stanford University and Elsevier for four consecutive year. His research interests include robotic rehabilitation, bio robotics, neurorobotics, artificial intelligence, and machine learning. He is the Vice President of the IEEE Robotics and Automation Society. He has served as a reviewer for more than 80 SCI (E)-indexed journals. He has been serving as an editorial board member for five SCI-indexed journals.

IMRAN FAREED NIZAMI received the B.Sc. degree (Hons.) in computer engineering from the University of Engineering and Technology (UET) Taxila, Pakistan, in 2005, the M.S. degree in electrical and electronic engineering from Yonsei University, Republic of Korea, in 2008, and the Ph.D. degree in electrical engineering from NUST, Pakistan, in 2019. He is currently Professor with the Department of Electrical Engineering, Bahria University, Islamabad, Pakistan. He was a recipient of the HEC Scholarship for the M.S. degree and the NUST Endowment Fund Scholarship for the Ph.D. degree. His research interests include IQA, pattern recognition, and machine learning.

SYED HAMMAD NAZEER (Senior Member, IEEE) received the bachelor's and master's degrees in mechatronics engineering from the NUST College of EME, Rawalpindi, and the Ph.D. degree in mechatronics engineering from Air University, Islamabad. He is currently an Associate Professor with the Department of Mechatronics and Biomedical Engineering and the Co-Director of the Neuroimaging Research Group, Air University. He has published more than 50 peer-reviewed research articles. His research interests include brain–computer interface, bio-robotics, neurorobotics, BCI-based exoskeletons, fNIRS-based cognitive workload assessment, and AI-driven biomedical applications. Beyond his other responsibilities, he has been actively involved in the IEEE Robotics and Automation Society, Pakistan, as the General Secretary, since 2018.

HUSAM A. NEAMAH (Member, IEEE) received the B.Sc. degree in computer engineering, in 2012, the M.Sc. degree in mechatronics engineering, in 2017, and the Ph.D. degree in engineering and technology informatics, in 2021. He is an Assistant Professor and the Head of the Autonomous and Assistive Research Laboratory, University of Debrecen, Hungary. His research areas focus on cognitive autonomous robotics, biomechatronic systems, interactive social robotics, and applied computational mathematical modeling. He is a member of the IEEE Industrial Electronics Society, the Vice Chair of the Technical Community on Robotics and Mechatronics, and a member of the IEEE Robotics and Automation Society. Additionally, he is a member of the editorial board, a reviewer, and part of the organization committee for various IEEE conferences.