APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 7, 2023, 178-189

Fatigue State Detection Through Multiple Machine Learning Classifiers Using EEG Signal

Md Mahmudul Hasan[1], Mirza Mahfuj Hossain[2] and Norizam Sulaiman[1*]

[1]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
[2]Department of Computer Science and Engineering, Jashore University of Science and Technology, Jashore-7408, Jashore, Bangladesh

[*]Corresponding author: norizam@umpsa.edu.my

Submitted 9 September 2023, Revised 19 November 2023, Accepted 2 December 2023, Available online 3 December 2023.

Abstract: Fatigued drivers can often cause long-distance accidents worldwide. Fatigue states are the primary cause of highway accidents. This study is conducted to provide a comprehensive and reliable fatigue state detection system to avoid accidents and make a good decision. Three machine learning algorithms were applied to seventy-six subjects' electroencephalogram (EEG) readings to test their performance. A preprocessing stage extracts relevant information before applying machine learning algorithms to the signal. Three analytical methods were employed in this study, specifically the Decision Tree, the K-Nearest Neighbors and the Random Forest. The study revealed that employing all the classifiers resulted in a satisfactory accuracy rate compared to existing state-of-the-art methods for detecting fatigue states. The classification accuracy using Decision Tree for four classes and two classes were achieved at 88.61% and 88.21% respectively, which can make this EEG-based technology a practical and dependable solution for real-time applications.

Keywords: Decision Tree; EEG signal; Fatigue detection; K-Nearest Neighbor; Machine learning; Random Forest.

1. INTRODUCTION

Approximately 15% to 20% of persons within the general population experience a condition known as excessive daytime sleepiness (EDS) [1-4], resulting in reduced work and driving efficacy. The primary etiological factors contributing to the development of EDS encompass socially induced sleep deprivation in persons without underlying medical conditions, medical diseases such as sleep apnea or narcolepsy, and the use of sedative medications [5-7]. The accurate evaluation of sleepiness holds significant importance in the areas of diagnosis, therapy, and the determination of driving capability. Despite the advancements in technology, the accurate evaluation of tiredness continues to pose a significant difficulty in the field of sleep-wake medicine [8]. Driver fatigue has been recognized as a prominent factor contributing to road accidents in numerous nations [9]. Globally, road accidents cause 1.17 million deaths annually, 70% of which take place in poor nations. Pedestrians account for 65% of these deaths, with children making up 35% of the victims [10]. Road crashes cause injuries to between 23 and 34 million people annually, according to estimates [10]. The reason behind the fatigue states and associated risks of fatigue state is illustrated in Figure 1.

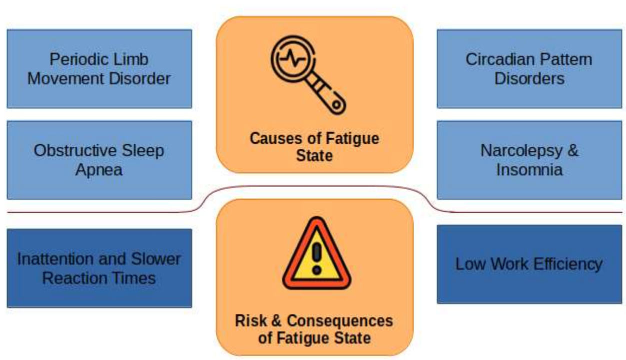

Figure 1. Fatigue states: causes and risks

Numerous research has been undertaken in the course of time with the aim of detecting drowsiness and alerting the driver, hence mitigating the frequency of accidents [11]. Another study on drowsiness and fatigue, the equipment that detects drowsiness and fatigue in drivers is variously known as a driver vigilance monitor, a drowsiness detection system, or a fatigue monitoring system [12]. A variety of valuable indicators can be employed to monitor and evaluate the state of driver drowsiness. Objective indicators Electro-oculogram (EOG), facial expressions and shifts, yawning, eye movement, pulse rate, breathing rate, skin conductance, and steering wheel grip are among the various manifestations of brain signals. Subjective approaches, such as the Karolinska Sleepiness Scale (KSS), are employed throughout the various stages of driving activity, including pre-driving, driving, and post-driving. Lane lateral deviation and steering wheel movement rates are additional measurements that assess the driving efficiency of a vehicle. The levels of drowsiness and alertness were assessed and categorized in several driving scenarios, including real-world driving, simulated driving tasks conducted in basic environments, and simulated driving in intricate environments within specialized laboratory settings [11]. Currently, the widely accepted method for objectively evaluating sleep and wakefulness is reliant on polysomnographic (PSG) data, with a specific focus on the electroencephalogram (EEG). An early report illustrated that the criteria for visually rating sleep were originally established by Rechtschaffen and Kales [13]. These criteria have since been adapted and revised by the American Academy of Sleep Medicine (AASM) and are currently utilized in their updated version [14-15]. The criteria utilized in this study are derived from 30-second epochs, which have been categorized into distinct stages of alertness, rapid eye movement sleep, and nonrapid eye movement sleep stages 1–3 (N1–N3). Clinically, the multiple sleep latency test (MSLT) [16] and the maintenance of wakefulness test (MWT) [17] are used to evaluate EDS [18-19]. The MWT is utilized to examine an individual's capacity to remain awake in the face of excessive daytime sleepiness (EDS). It is widely regarded as the primary measure of vigilance for evaluating a patient's suitability for driving [17, 20, 21]. It is debatable whether classifications of wakefulness and sleep based on 30-second intervals are still accurate [22], particularly within the realm of driving, brief instances of inattention can lead to severe and sometimes lethal outcomes. Therefore, the term microsleep is frequently used in contemporary scientific literature and mostly pertains to brief periods of "sleep" lasting less than 15 seconds, as determined using PSG data.

Microsleep can also be identified through behavioral observations captured by videography, such as eye lid closing, or through psychomotor performance assessments. Episodes of microsleep (MSEs) derived from EEG data are visually assessed as periods lasting from 3 to 15 seconds. These episodes are characterized by the prevalence of theta activity, which refers to the power of EEG signals within the frequency range of 4 to 8 Hz. During MSEs, the alpha activity, which represents the power within the frequency range of 8 to 12 Hz, is replaced by theta activity. Additionally, it is common for MSEs to be followed by the closure of the eyelids. Furthermore, less exact characterizations were employed for MSEs, such as referring to them as "brief episodes of typical stage 1 sleep" [23-26]. In addition to the absence of standardization and the various approaches used for MSE identification, visual scoring is time-consuming, requires training and experience, and is subjective.

The utilization of multichannel EEG analysis enhances the precision of sleep stage recognition, a critical component in the precise identification of driver fatigue states. Furthermore, we have employed two distinct methodologies in our dataset to enhance the quality of our study. The initial classification consisted of four categories: wake, microsleep episode, microsleep episode candidate, and episode of drowsiness. An alternative method involved the adaptation of the four categories into two distinct groups. Normal state was assigned to the category of Wakefulness. MSE, microsleep episode candidate (MSEc) and episode of drowsiness (ED) have been designated to the fatigue state category, which provides greater classification precision and detects the fatigue state with ease. Thirdly, our method employs multiple machine learning classifiers to provide a comprehensive and accurate method for detecting driver fatigue.

This study introduces an approach that aims to accurately classify different levels of driver fatigue states by employing various machine learning classifiers, such as K-Nearest Neighbors (KNN) and Decision Tree (DT). Following the process of classification, we proceeded to examine the significance of Cohen's Kappa, Sensitivity, Specificity, and Precision. Cohen's Kappa coefficient considers both the observed level of agreement among raters and the level of agreement that would be anticipated due to chance. Sensitivity is a quantitative assessment of the accuracy with which a test or diagnostic method can correctly detect persons who truly possess a specific ailment or trait. The concept of specificity refers to the capacity of a diagnostic test to accurately classify persons who do not possess a certain condition as "negative" or "non-affected". Precision is a quantitative metric used to assess the degree of correctness in positive predictions generated by a model or a test.

2. RELATED WORKS

Numerous methodologies have been proposed to ascertain the fundamental processes of weariness in EEG signals. One approach involves the computation of various forms of entropies as feature sets using a single channel [27]. Quintero-Rincon has proposed an approach that is both direct and effective in detecting driver fatigue in real-time systems [28]. This method utilizes a single-channel EEG signal. The technique employs a selection process to identify the most prominent channel and subsequently extracts four feature parameters. These parameters are utilized in the detection of fatigue through the implementation of a combination notched decision trees classifier. The proposed methodology attains a precision rate of 92.7% and incurs a time delay of 1.8 seconds, utilizing data acquired from the database of Jiangxi University of Technology. Nevertheless, it is crucial to acknowledge that the investigation assessed the approach within the confines of a particular dataset, therefore necessitating additional research to ascertain its efficacy across diverse datasets and varying circumstances. Furthermore, it is worth noting that the time delay of 1.8 seconds may pose practical limitations in certain scenarios when real-time monitoring is required. It is crucial to take into account the potential consequences for driver safety that could arise from any delay in recognizing weariness.

In a separate investigation, Jing et al. conducted a study with the objective of identifying driving fatigue in circumstances characterized by low-voltage and hypoxic plateau conditions. This was achieved by the utilization of both subjective and objective monitoring techniques [29]. The EEG data acquired from live driving experiments were examined using both linear and nonlinear methodologies to assess the patterns manifested by the signals throughout periods of wakefulness, criticality, and weariness. However, the scope of the investigation was restricted to conducting tiredness assessments during field driving in a particular setting. Consequently, additional research is imperative to corroborate the outcomes in diverse locations and various driving circumstances.

Furthermore, Zhang et al. introduced a novel methodology called clustering on brain networks (CBNs) in order to enhance the efficacy of driver tiredness detection [30]. The methodology employed by the CBNs is the use of a clustering algorithm to identify spatial nodes with unique connection properties from EEG data. The wavelet entropy data obtained from these nodes is further transformed into spatiotemporal images and subjected to analysis using an image edge detection technique, with the aim of discerning varying degrees of fatigue. The utilization of this technique effectively mitigates signal interference and enables the detection of fatigue prior to the emergence of subjective sensations. Consequently, it holds significant promise as a valuable instrument for the purpose of early warning and accident avoidance. The study presented findings that highlight the constraints associated with utilizing EEG indicators in both the temporal and frequency domains for the purpose of accurately identifying driver weariness. These limitations mostly arise from the complexities of signal mixing and the restricted sample size employed in the research. Moreover, the study lacks a comprehensive comparison with established methodologies and fails to validate its findings in real-world driving situations. Subsequently, the preceding researcher put out a sophisticated method designed to identify driver fatigue through the use of EEG signals [31]. The proposed system consists of a feature generating network that incorporates texture descriptors and a hybrid feature selection method in order to improve the accuracy of detection. The framework that was proposed demonstrated a notable classification accuracy of 97.29% in the identification of exhaustion through the analysis of EEG data. This outcome underscores the framework's potential in effectively detecting driver fatigue. Nevertheless, the framework that was suggested employed conventional machine learning techniques, potentially constraining its capacity to accommodate intricate and ever-changing driving conditions.

A new methodology for effectively identifying driver weariness through the analysis of EEG data [32]. The methodology employs an innovative channel selection technique that is based on correlation coefficients. Additionally, the proposed approach integrates an ensemble classifier that utilizes the random subspace KNN algorithm. Additionally, power spectral density (PSD) is employed for feature extraction. The methodology successfully attained a notable level of precision, reaching 99.99% accuracy in the detection of driver fatigue through the analysis of EEG data within a time frame of 0.5 seconds. The approach described in this study exhibits robust performance and efficiently identifies driver weariness using EEG-based measurements. Nevertheless, the utilization of a KNN-based ensemble classifier in real-time applications may be impractical due to its substantial processing complexity. In a separate study, Hwang et al. presented a model for classifying driver fatigue states using EEG data that is not reliant on the individual being studied [33]. This model also takes into consideration the variations in performance levels among individuals. The adversarial training methodology was utilized by the researchers to intentionally induce misperception of object categories within their classification model. In addition, the researchers employed the Inter-subject Feature Distance Minimization (IFDM) method to address performance discrepancies among individuals. The approach utilized in their study facilitated the training of EEG datasets that had a restricted number of subject labels. The evaluation of this approach was conducted on the SEED-VIG dataset, which led to improved accuracy and reduced variability in individual performance when classifying drowsiness. Nevertheless, a significant limitation lies in the substantial inter-individual variability present in EEG data, posing a formidable obstacle in the development of a comprehensive model capable of achieving optimal performance across all individuals.

The literature examined presents a range of methodologies for identifying driver weariness through the analysis of EEG data, encompassing both simple single-channel feature extraction techniques and more intricate machine learning models. One prevalent methodology is employing power spectral density and diverse entropy measurements as feature sets, whilst alternative approaches include clustering algorithms and picture edge detection to differentiate between distinct states of weariness. Numerous studies have also investigated the issue of individual performance gaps and subject variability through the utilization of adversarial training methodologies and component-specific batch normalization techniques. These researches illustrate the promise of utilizing EEG-based driver fatigue identification as a means of early warning and accident prevention. They have achieved high levels of accuracy and have opened up new avenues for extracting more information from intricate EEG data. Nevertheless, there are differences in the computing complexity, the number of channels needed, and the degree of subject independence attained among the various ways. This indicates the necessity for additional research to determine the most optimal and successful approach for real-world implementations.

Wilapiprasitporn et al. put forth a deep learning methodology that integrates Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for the purpose of recognizing persons based on affective EEG data [34]. The research employed the Database for Analysis of Emotions using Physiological Events (DEAP) dataset and revealed that the proposed approach exhibits greater performance in comparison to a Support Vector Machine (SVM) baseline system. The Correct Recognition Rate (CRR) achieved by the proposed methodology ranged from 99.90% to 100%. Recent research showed that CNN-GRU models outperform CNN- Long Short-Term Memory (LSTM) models in the domain of person recognition utilizing EEG data collected particularly from the frontal region of the brain. Additionally, these models have demonstrated effectiveness in mitigating the influence of affective states. Nevertheless, the proposed methodology is dependent on EEG signals, which necessitate the use of specialized equipment and experience in data collecting and interpretation.

Table 1. The summary of traditional machine learning algorithms applicated in EEG-based fatigue detection

Reference	Feature Extraction Method	Classification Algorithm	Accuracy (%)
Zhao et al. [38]	MVAR[1]	KPCA-SVM[2]	81.64
Chai et al. [39]	ERBM-ICA[3], AR	BNN[4]	84.3/83.0

[1]MVAR: Multivariate Autoregressive; [2]KPCA-SVM: Kernel Principal Component Analysis - SVM; [3]ERBM-ICA: Entropy Rate Bound Minimization Analysis – Independent Component Analysis; [4]BNN: Bayesian Neural Network.

Table 2. The summary of deep learning algorithms used in EEG-based fatigue detection

Reference	Classification Algorithm	Accuracy (%)
Reddy et al. [41]	CNN	85.42
Paulo et al. [42]	CNN	75.87
Liu et al. [43]	MIDA[1], TCA[2]	73.01

[1]MIDA: Maximum Independence Domain Adaptation; [2]TCA: Transfer Component Analysis.

In their study, Qin et al. introduced a novel deep learning model that integrates CNN and LSTM networks. The purpose of this model is to effectively extract vein features from raw photos, specifically for the application of finger-vein biometrics [35]. The model under consideration employs supervised encoding techniques to reduce binary vein texture, leading to a notable enhancement in verification accuracy when assessed on a finger-vein database that is publicly accessible. Nevertheless, it is worth noting that deep learning models exhibit a susceptibility to overfitting, a phenomenon in which they become too proficient at learning from the training data, hence impeding their ability to effectively generalize to novel data. Methods like regularization and dropout can be employed to mitigate the issue of overfitting. One of the earlier attempts to apply deep learning approaches to solve the drowsiness detection problem was a deep neural technique that was proposed in [36]. The driver's RGB video is used to extract facial features using this method. Feature fused architecture (FFA) was created by combining three CNN models: Residual Network (ResNet50), InceptionV3, and Visual Geometry Group 16 (VGG16). This technique's low accuracy of 78% is its key drawback.

2.1 Classification using Conventional Machine Learning Algorithms

Feature extraction methods are commonly utilized in the application of traditional machine learning classification algorithms. During the initial stages, researchers employed conventional feature extraction algorithms in conjunction with machine learning algorithms to enhance the precision of classification [37]. Furthermore, researchers have put forth several network architectures that incorporate different feature extraction techniques in order to autonomously extract profound and impactful information. Table 1 presents an overview of the machine learning articles, including information on the author, feature extraction method, classification algorithm, and accuracy. The table presents a comparison of feature extraction techniques, where the initial four items employ conventional methods for feature extraction, while the last two items utilize a feature extraction network to extract features.

2.2 Classification using Conventional Deep Learning Algorithms

When contrasting machine learning and deep learning, it becomes evident that deep learning possesses the capability to effectively use substantial volumes of data, hence resulting in enhanced categorization performance [40]. In recent years, there has been a notable utilization of various deep networks for the purpose of detecting weariness based on EEG data. In order to take into consideration for the substantial inter-individual variability in EEG measurements, it is imperative to develop a robust model that is trained on data acquired from the same individual. Moreover, this particular model has limited suitability for novice users, while the initial model demonstrates a comparatively lower level of robustness. Hence, it is imperative to develop a robust model for real-world applications that can effectively detect weariness across different individuals. In recent academic investigations, a limited cohort of scholars employed deep learning techniques to detect exhaustion based on EEG data across different individuals. Specifically, they sought to enhance the CNN model in order to enhance the accuracy of fatigue identification among subjects [37]. Numerous researchers have endeavored to employ transfer learning as a means to mitigate the variability observed between subjects. This approach involves using the knowledge and information acquired from a source domain and applying it to a target domain. Several techniques for domain adaptation have been utilized in this sector, including Maximum Independence Domain Adaptation (MIDA), Domain-Adversarial Neural Network (DANN), and Easy Transfer Learning (EasyTL). Table 2 provides a comprehensive overview of the deep learning studies, encompassing details such as the authors, classification algorithms employed, and corresponding accuracy measures.

The literature reviewed in this study explores various deep learning techniques that have been utilized in a wide range of applications. These applications include emotional EEG-based human recognition, finger-vein biometrics, operative workflow investigation, improvement of speech, and structural shape design. The models that were proposed exhibited notable enhancements in terms of accuracy, efficiency, and application when compared to previous methodologies. In summary, the utilization of EEG-based methods for detecting fatigue states has demonstrated significant promise in terms of providing timely alerts and preventing accidents. This has been accomplished through the application of diverse machine learning techniques, which have yielded impressive levels of accuracy and facilitated the extraction of more comprehensive insights from intricate EEG data. Furthermore, the field of machine learning has made substantial advancements in terms of enhancing precision, effectiveness, and practicality across a wide range of applications. The applications encompass a range of areas, such as emotional EEG-based human recognition, finger-vein biometrics technology, surgical process analysis, voice advancement, and structural layout design. Further research is necessary to determine the most effective and efficient approach for the practical use of machine learning.

3. METHODOLOGY

Block diagram and flowchart of the suggested system are shown in Figures 2 and 3. Figure 2 illustrates the sequential processing steps performed on the EEG data. Initially, the input stage involves loading the EEG data. Subsequently, the data undergoes filtering using the Fast Fourier Transform (FFT) technique. Following this, the data is further subjected to band filtering, resulting in segmentation into distinct frequency bands. Finally, a normalization process is used to the filtered data. During the processing stage, the training data undergoes training using three algorithms, namely DT, KNN and Random Forest (RF). Subsequently, the outcome is analyzed and the classifiers are further refined based on this analysis. The specified classification algorithms were trained using the Python-based scikit-learn module within the Jupyter notebook environment. The training utilized preprocessed data obtained from the records of 60 participants. The models that were obtained as a result of the training process were carefully maintained for future use. The stored models were utilized in a sequential manner to analyze the test data, which consisted of information gathered from a total of 16 people. The careful examination enabled the calculation and determination of the achieved level of accuracy inside the assessment procedure.

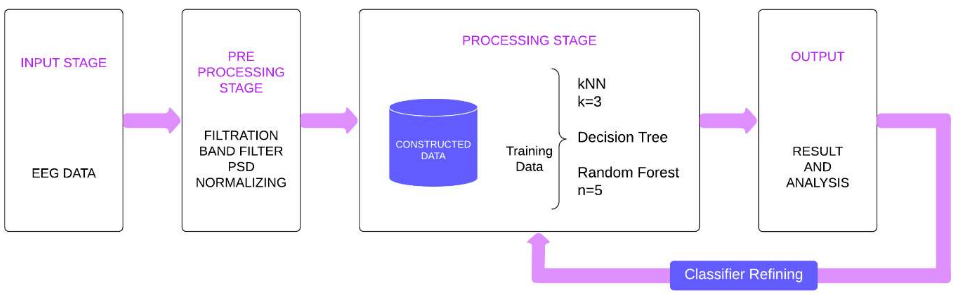

Figure 2. Comprehensive design of the current investigation

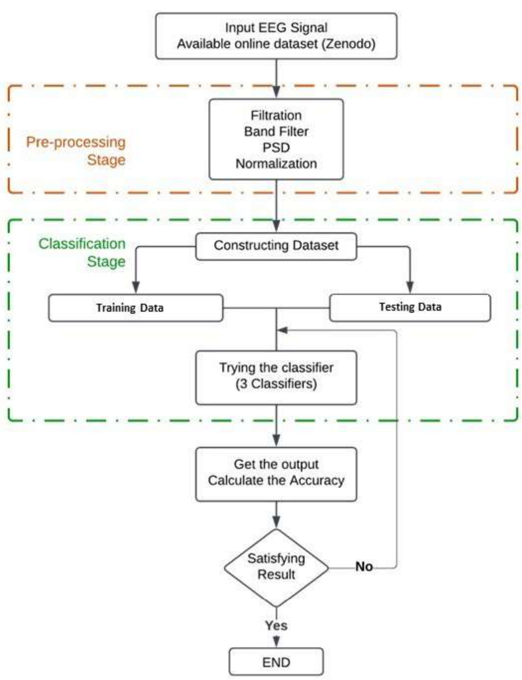

Figure 3. System flowchart for EEG-based fatigue detection

The flowchart in Figure 3 provides a comprehensive overview of the experimental procedure. The EEG data was initially loaded. Subsequently, the data underwent filtration using a band limit ranging from 0.5 Hz to 45 Hz. Then PSD was calculated using Equation (1),

$$PSD = |X(k)|^2 = |\sum_{n=0}^{N-1} x(nTs) e^{-j2\pi nk/N}|^2 \tag{1}$$

where the variable k is defined as an integer ranging from 0 to $N-1$. It is worth noting that for a specific value of k, there are N complex multiplications involved. The multiplication of w^{kn} and $x(n)$ was performed N times, where n ranges from 0 to $N-1$.

Subsequently, the data was partitioned into four distinct frequency bands, namely delta, theta, alpha, and beta, and subsequently subjected to normalization in preparation for the training procedure. The collected data was partitioned into two subsets, namely the training data and the testing data. The training dataset was subjected to the three classification algorithms, (KNN, DT, RF). The accuracy results were derived by computing the output provided by the algorithms. If the results did not meet state-of-the-art criteria, the classifiers were re-executed with some adjustments. Upon obtaining satisfactory outcomes, we deemed the algorithms to be optimal and concluded our experimental endeavors.

3.1 Data Acquisition

We employed an online EEG dataset consisting of 76 participants for our experiments [44] and data (https://zenodo.org/record/3251716) are available. These records encompass 2 EEG channels. The first EEG channel O1-M2 where M2 is the mastoid electrode on the opposite side and the second EEG channel O2-M1 where M1 is the mastoid electrode on the opposite side. This EEG channels were recorded over a sampling rate of 200 Hz. These EEG channels are labeled for each sample which are further classified into four states which are Wake, MSE, MSEc, and ED states. Our approach involved using data from 60 participants for model training, while the remaining 16 participants' data was reserved for testing and evaluation. This division allowed us to construct and fine-tune our predictive models effectively. By training on a substantial portion of the dataset and testing on unseen samples, we aimed to assess the generalization and performance of our models accurately.

This dataset's diverse patient profiles and distinct EEG states provided a robust foundation for our investigation, enabling us to develop models that can potentially contribute to improved understanding and analysis of neural signals in various clinical contexts.

3.2 Data Pre-processing

The unprocessed EEG signal underwent signal processing techniques to eliminate undesired interferences and non-signal artifacts [45-46]. The presence of high frequency components, such as power-line interferences, significantly contributes to the introduction of noise in the EEG signal [47]. The presence of noise into the system has the potential to interfere with and distort the signal, hence diminishing the precision of feature extraction and classification [45]. The signals were subjected to bandpass filtration within the frequency range of 0.5–45 Hz using a Fourier filter. Equation (2) represents the FFT:

$$\mathrm{X}(k) = \sum_{k=1}^{N-1} X(n) W_N^{kn};\ K = 0 \ldots\ldots N-1 \tag{2}$$

where $X(k)$ is the FFT coefficient at frequency k, $x(n)$ is the discrete-time signal, W is the twiddle factor, and N is the length of the signal.

The procedure involved conducting a FFT on the EEG signals, selectively eliminating frequencies below 0.5 Hz and above 45 Hz, and afterwards reversing the FFT process. Signal conditioning is widely acknowledged as a crucial phase in the process of responding to future data collected from various devices. Notwithstanding this processing, the data retains its raw identifier as no discernible attributes were retrieved for the purpose of classification. Subsequently, the Fourier filter bandpass is employed to process the data, resulting in the filtration of the raw data into four distinct signal bands, namely Theta, Delta, Alpha, and Beta. EEG signals are typically examined throughout diverse frequency bands, including Delta waves. Delta waves are the EEG patterns that exhibit the slowest patterns and are distinguished by the highest amplitude EEG waveforms. These waveforms are normally measured at around 250-325 microvolts and encompass frequencies that are below 4 Hz [48-51]. Theta waves are characterized by larger amplitudes and frequencies, ranging from 4 Hz to 8 Hz. On the other hand, alpha waves are regular rhythms occurring during states of relaxation and alertness, with frequencies ranging from 8 Hz to 12 Hz. Beta waves, which exhibit smaller voltage and higher frequency rhythms, normally range from 14 Hz to 32 Hz, occasionally reaching up to 50 Hz [48-51]. The figure for the raw EEG signal, filtered signal and different bands after filtration for the first participant (0ncr) is illustrated in Figure 4. The figure also represents the graphical representation of the unprocessed EEG signal, the filtered EEG signal, PSD, and the signal partitioned into four distinct frequency bands, namely delta, theta, alpha, and beta. The EEG signal segment was extracted from the EEG dataset of the first subject (0ncr).

This method ensures the signal data's compatibility with diverse recording equipment in the future, preserving the data's utility and facilitating robust analysis in varying experimental settings. Then, we normalized the features by removing the mean and scaling to unit variance using the normalization procedure. The standard score of a sample x is computed in accordance with Equation (3) where x^i is the i-th original value of the feature, μ is the mean of all feature values and σ is the standard deviation of all feature values.

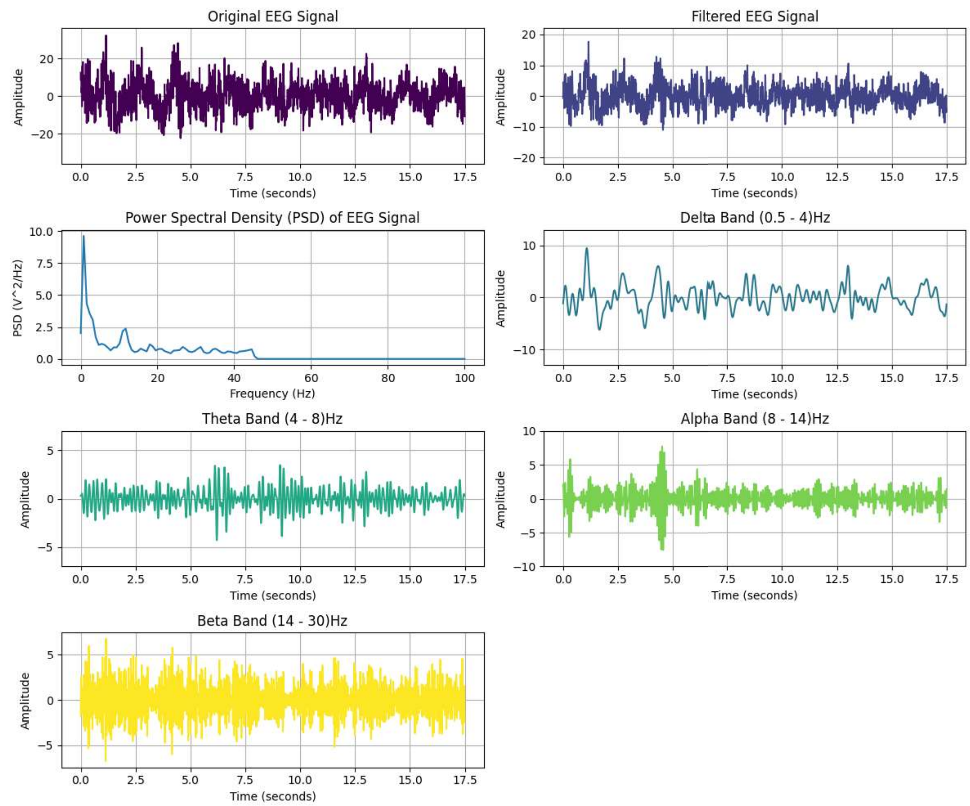

Figure 4. Raw EEG signal, filtered signal, PSD, delta, theta, alpha and beta bands

$$Z^i = \frac{X^i - \mu}{\sigma}\ ;\ i = 0,1 \ldots\ldots N \tag{3}$$

3.4 Performance Evaluation

In order to assess the quality of classification, the evaluation of classification results and classifier performance is conducted using various metrics. These metrics include classification accuracy, specificity, sensitivity, recall, F1 score, precision, Matthews correlation coefficient (MCC), and area under the curve. These metrics are stated in Equations (4) – (9). Precision shows how well it can separate true positives and true negatives from the rest of the data. Besides, the recall value can be seen as the ratio of accurately identified positive samples in relation to the total number of positive samples. The F1 score is a metric that may be regarded as a weighted mean of precision and recall, where a value of one represents the optimal performance and a value of zero represents the worst performance.

The terms TP, TN, FP and FN represent the abbreviations for true positive, true negative, false positive and false negative, respectively. MCC is a metric commonly employed in the field of machine learning to evaluate the performance of binary classification models [32]. The MCC can be understood as a numerical measure representing the correlation coefficient, which ranges from -1 to +1.

$$\text{Sensivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} + 100\% \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} + 100\% \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} + 100\% \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{F1} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{9}$$

Table 3. Performance of the classifiers in percentages for two algorithms while considering two types of conditions

Algorithm	Sensitivity (%)	Precision (%)	Specificity (%)	Cohen's Kappa	MCC	Accuracy (%)
Decision Tree-4	86.05	51.25	73.13	0.56	0.92	**88.61**
KNN-4	30.48	77.92	72.72	0.29	0.21	88.04
Random Forest-4	25.01	25.01	75.01	0.06	0.19	86.63
Decision Tree-2	99.96	34.64	88.95	0.59	0.16	88.21
KNN-2	98.71	52.86	90.75	0.31	0.34	87.02
Random Forest-2	86.55	96.77	89.08	0.16	0.26	86.55

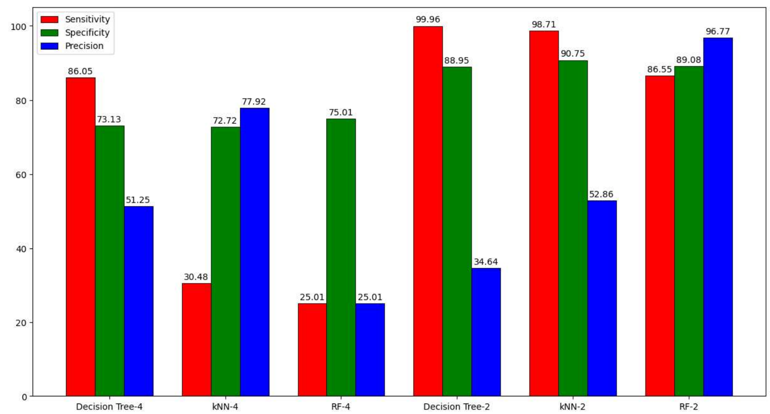

Figure 5. Performance evaluation of DT, KNN and RF using sensitivity, specificity and precision

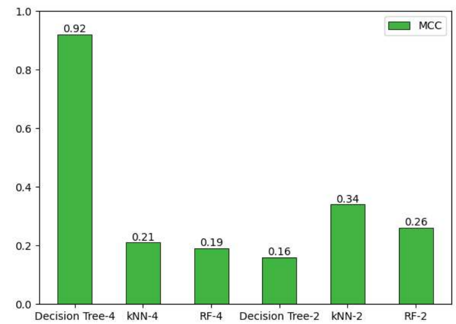

Figure 6. Performance evaluation of DT and KNN using MCC

4. RESULT

4.1 Performance Evaluation

Initially, this study conducted an analysis of the dataset, taking into account the four categories classified as waking, MSE, MSEc, and ED. The investigation focused on the alpha (8-13 Hz), delta (0.5-4 Hz), and theta (4-8 Hz) categories of frequencies, that have been linked to diverse cognitive and emotional activities. The objective of this work was to assess the efficacy of two widely used machine learning models, namely DT and KNN, in the context of categorization tasks. The DT exhibited superior performance in terms of accuracy, specificity, and precision compared to the alternative classifier, achieving values of 88.61%, 75.79%, and 94.04%, respectively (Table 3).

Nevertheless, the sensitivity value of 86.05% shown a satisfactory value compared to the KNN and DT algorithms for four classes classification. Figure 5 illustrates the sensitivity, specificity, and precision metrices for the algorithms employed in the study, namely DT-4, KNN-4, RF-4, DT-2, KNN-2 and RF-2. Meanwhile, RF-2 gave the best result in case of precision, which was 96.77% with a satisfactory accuracy of 86.55%. The figure illustrates that the inclusion of all four states results in a very low sensitivity. However, when only two states, namely normal and fatigue, are considered, the sensitivity significantly increases. In the context of specificity and precision, an inverse relationship can be observed, whereby both metrics exhibit high values when four states are taken into account, and low values when just two states are included.

The DT algorithm for four classes classification yielded MCC value of 0.92, which was higher compared to the MCC value of 0.21 obtained by the KNN algorithm and 0.19 obtained by the RF algorithm (Figure 6). This figure provides an illustration of the MCC values associated with various algorithms. The algorithm DT achieved the highest MCC value of 0.92 for both types of classification. However, while considering two states, the DT algorithm has the lowest MCC value among all other algorithms.

The dataset was further examined, with waking being classified as the Normal condition, and MSE, MSEc, and ED being classified as indicators of the fatigue state. This study also assessed the performance of DT and KNN algorithms in order to determine the most efficient classifier. In this instance, the DT algorithm exhibited superior performance in terms of accuracy and sensitivity, achieving values of 88.21% and 99.96% respectively. The accuracy comparison can be observed in Figure 7. The DT method achieved the highest accuracy measurement of 88.61% when all four states were taken into account. However, it is worth noting that the accuracy measures for alternative algorithms are also rather satisfactory, with all of them surpassing a value of 86%. The accuracy measurement of the DT algorithm demonstrates superior performance in both scenarios, encompassing both four-state and two-state classifications. The accuracy of the KNN algorithm was found to be 87.02%, which was the lowest among the two states under consideration. The value of Cohens Kappa for RF classifier for both approaches did not show a better result. On the other hand, Cohen's Kappa for DT for both approaches gave best moderate results (Table 4).

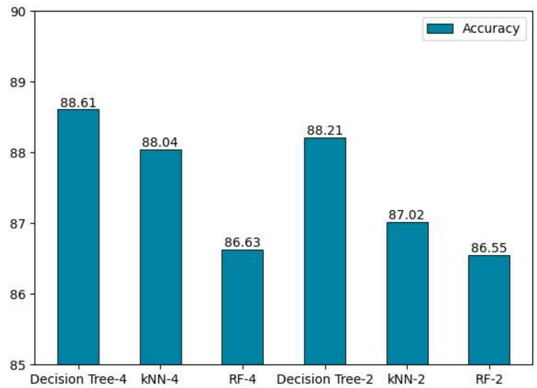

Figure 7. Comparison of accuracy for different algorithms

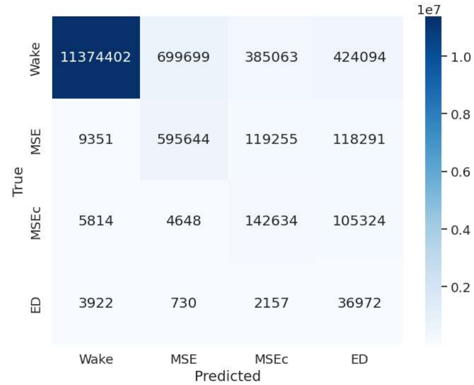

Figure 8. Confusion matrix for the proposed DT-4 model

However, the specificity, precision, and MCC of the aforementioned method were shown to be inferior to those of the KNN algorithm (Figures 5 and 6). Based on the findings, it was determined that the DT model exhibited more efficiency in discerning fatigue states from EEG signals compared to the KNN and RF models. The results indicate that the DT classifier exhibits considerable promise in reliably identifying human fatigue stages based on EEG signals. The findings of this study provide significant significance in terms of mitigating the hazards linked to drowsy driving and enhancing overall safety.

4.2 Performance Assessment

Confusion matrices are valuable tools for assessing a classification model's performance, which makes them important in the field of machine learning. Classifying predictions into four distinct categories—true positives, true negatives, false positives, and false negatives—allows for the evaluation of the model's accuracy. The information mentioned above is essential for evaluating the model's benefits and shortcomings, identifying possible improvement areas, and fine-tuning algorithms to maximize overall performance.

Figure 8 illustrates the 3D confusion matrix of DT model for four class classifications. It is also visible that DT-4 model predicted all four classes in a very satisfactory amount, while it gave a small portion of wrong prediction. DT-4 gave significant performance. True positive values and True negative values were much higher compared to False positive and False negative values. The findings indicate that the DT-4 model exhibits the highest level of performance, notably in accurately recognizing positive situations while also keeping a relatively low percentage of false positives. This makes it the most notable pick among other models.

5. DISCUSSION

In summary, the proposed framework for detecting fatigue states via different machine learning architectures has demonstrated significant impact in reliably discerning fatigue states from EEG signals. The findings of our investigation indicate that the framework exhibited a notable level of accuracy across all classifiers. Specifically, the Decision Tree classifier had the best accuracy of 88.61% when all four states were taken into account. In an alternative methodology, it was noted that the DT algorithm exhibited a level of accuracy up to 88.21%. This level of accuracy was achieved through the adjustment of the waking state to signify the standard condition, while the state of weariness was denoted by the metrics MSE, MSEc, and ED. The KNN algorithm demonstrated notable accuracy, with a rate of 88.04% when all four states were taken into account. In contrast, the KNN algorithm demonstrated an accuracy rate of 87.02% when considering the two states. Meantime, the RF classifier showed the accuracy of 86.63% when all four states were taken into account and the accuracy of 86.55% when two states were taken into account. The contribution of our study is centered on the presentation of a comprehensive and efficacious framework for the precise detection of fatigue states from EEG signals. This framework surpasses the performance of prior approaches. Moreover, the suggested framework has the potential to be used in real tiredness detection systems, hence enhancing driving safety and mitigating the occurrence of road accidents resulting from driver exhaustion. Subsequent investigations may extend the present framework by examining its efficacy within practical contexts and delving into strategies for enhancing its capacity to detect nuanced variations in EEG data.

6. CONCLUSION

The rapid pace of contemporary culture can readily result in weariness, which can have direct or indirect detrimental effects on the human body. Undoubtedly, the development of high-precision, real-time, and universally applicable tiredness detection holds immense importance. This investigation proposed a strategy within an EEG-based framework for the objective detection of fatigue states. Three machine learning classifiers (KNN, DT and RF) were employed to train and test the data. The study documented notable performance results, with an accuracy rate of approximately 85-90% achieved by all the classifiers. The findings indicated that the utilization of this particular technology holds potential in the identification of tiredness conditions. Fatigue detection, being a rapidly evolving area of research, necessitates the integration of multidisciplinary knowledge in order to enhance its efficacy. In addition to enhancing the detection algorithm, it is important to develop novel and efficient paradigms and portable equipment, as well as gather reliable standard data. It is imperative to do a replication of the study using a substantial sample size and real-world driving scenarios to ascertain the accuracy of EEG measurements. Future research should focus on investigating the dependability and comfort of various frequency bands while utilizing additional features for the purpose of real-time monitoring of tiredness levels.

ACKNOWLEDGEMENT AND FUNDING

The authors would like to acknowledge the great support to Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah for providing research grant, RDU210318 and laboratory facilities to conduct the research.

DECLARATION OF CONFLICTING INTERESTS

The authors declare no potential conflicts of interest with respect to the research and publication of this article.

REFERENCES

[1] E. S. Ford, T. J. Cunningham, W. H. Giles and J. B. Croft, Trends in insomnia and excessive daytime sleepiness among US adults from 2002 to 2012, *Sleep Medicine*, 16(3), 2015, 372-378.
[2] A. C. Hayley *et al.*, Excessive daytime sleepiness and falls among older men and women: cross-sectional examination of a population-based sample, *BMC Geriatrics*, 15(1), 2015, 74.
[3] C. Hara, F. Lopes Rocha and M. F. F. Lima-Costa, Prevalence of excessive daytime sleepiness and associated factors in a Brazilian community: the Bambuí study, *Sleep Medicine*, 5(1), 2004, 31-36.
[4] T. B. Young, Epidemiology of daytime sleepiness: definitions, symptomatology, and prevalence, *The Journal of Clinical Psychiatry*, 65(16), 2004, 12-16.
[5] J. Mathis and C. W. Hess, Sleepiness and vigilance tests, *Swiss Medical Weekly*, 19(15-16), 2009, 214-219.
[6] T. Akerstedt, C. Bassetti, F. Cirignotta and D. Garc'ia-Borreguero, Sleepiness at the wheel: white paper, *European Sleep Research Society, French Motorway Companies (ASFA), National Institute of Sleep and Vigilance (INSV)*, 2013.
[7] D. J. Gottlieb, J. M. Ellenbogen, M. T. Bianchi and C. A. Czeisler, Sleep deficiency and motor vehicle crash risk in the general population: a prospective cohort study, *BMC Medicine*, 16(1), 2018, 44.
[8] J. Skorucak, A. Hertig-Godeschalk, D. R. Schreier, A. Malafeev, J. Mathis and P. Achermann, Automatic detection of microsleep episodes with feature-based machine learning, *Sleep*, 43(1), 2020, zsz225.
[9] Commercial Motor Vehicle Driver Fatigue, Long-Term Health, and Highway Safety: Research Needs, National Academies Press, Washington, USA, 2016.
[10] A. Kareem, Review of global menace of road accidents with special reference to malaysia- a social perspective, *Malaysian Journal of Medical Sciences*, 10(2), 2003, 31-39.
[11] I. A. Fouad, A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms, *Ain Shams Engineering Journal*, 14(3), 2023, 101895.
[12] Fatigue European Road Safety Observator Report, http://www.erso.eu, 2006 ((accessed 08.09.2023).
[13] A. Rechtschaffen and A. Kales, A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. *Public Health Service*, US Government Printing Office, Washington DC, 1968.
[14] R. B. Berri *et al.*, The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications, version 2.5. Darien, CT: *American Association of Sleep Medicine*, 2018.
[15] C. Iber, The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specification, https://cir.nii.ac.jp/crid/1370004237604151044, 2007 (accessed 08.09.2023).
[16] M. A. Carskadon, Guidelines for the multiple sleep latency test (MSLT): A standard measure of sleepiness, *Sleep*, 9(4), 1986, 519-524.
[17] M. Mitler, K. S. Gujavarty and C. P. Browman, Maintenance of wakefulness test: A polysomnographic technique for evaluating treatment efficacy in patients with excessive somnolence, *Electroencephalography and Clinical Neurophysiology*, 53(6), 1982, 658-661.
[18] J. Mathis, S. de Lacy, C. Roth and E. A. Hill, Measuring-monitoring sleep and wakefulness, *ESRS European Sleep Medicine Textbook. Regensburg: European Sleep Research Society*, 2014, 125-143.
[19] M. R. Littner *et al.*, Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test, *Sleep*, 28(1), 2005, 113-121.
[20] J. Mathis and D. Schreier, Daytime sleepiness and driving behaviour, *Therapeutische Umschau*, 71(11), 2014, 679-686.
[21] Y. Harrison and J. A. Horne, Occurrence of 'microsleeps' during daytime sleep onset in normal subjects, *Electroencephalography and Clinical Neurophysiology*, 98(5), 1996, 411-416.
[22] C. Guilleminault, M. Billiard, J. Montplaisir and W. C. Dement, Altered states of consciousness in disorders of daytime sleepiness, *Journal of the Neurological Sciences*, 26(3), 1975, 377-393.
[23] V. L. Tirunahari, S. A. Zaidi, R. Sharma, J. Skurnick and H. Ashtyani, Microsleep and sleepiness: a comparison of multiple sleep latency test and scoring of microsleep as a diagnostic test for excessive daytime sleepiness, *Sleep Medicine*, 4(1), 2003, 63-67.

[24] H. J. Moller, L. Kayumov, E. L. Bulmash, J. Nhan and C. M. Shapiro, Simulator performance, microsleep episodes, and subjective sleepiness: normative data using convergent methodologies to assess driver drowsiness, *Journal of Psychosomatic Research*, 61(3), 2006, 335-342.

[25] L. N. Boyle, J. Tippin, A. Paul and M. Rizzo, Driver performance in the moments surrounding a microsleep, *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(2), 2008, 126-136.

[26] U. S. Herrmann, C. W. Hess, A. G. Guggisberg, C. Roth, M. Gugger and J. Mathis, Sleepiness is not always perceived before falling asleep in healthy, sleep-deprived subjects, *Sleep Medicine*, 11(8), 2010, 747-751.

[27] M. -P. Hosseini, A. Hosseini and K. Ahi, A review on machine learning for EEG signal processing in bioengineering, *IEEE Reviews in Biomedical Engineering*, 14, 2021, 204-218.

[28] A. Quintero-Rincon, M. E. Fontecha and C. D'Giano, Driver fatigue EEG signals detection by using robust univariate analysis. *arXiv*, 2019 (accessed 08.09.2023).

[29] D. Jing, D. Liu, S. Zhang and Z. Guo, Fatigue driving detection method based on EEG analysis in low-voltage and hypoxia plateau environment, *International Journal of Transportation Science and Technology*, 9(4), 2020, 366-376.

[30] C. Zhang, L. Sun, F. Cong, T. Kujala, T. Ristaniemi and T. Parviainen, Optimal imaging of multi-channel EEG features based on a novel clustering technique for driver fatigue detection, *Biomedical Signal Processing and Control*, 62, 2020, 102103.

[31] T. Tuncer, S. Dogan, F. Ertam and A. Subasi, A dynamic center and multi threshold point based stable feature extraction network for driver fatigue detection utilizing EEG signals, *Cognitive Neurodynamics*, 15(2), 2021, 223-237.

[32] M. Rashid, M. Mustafa, N. Sulaiman, N. R. H. Abdullah and R. Samad, Random subspace K-NN based ensemble classifier for driver fatigue detection utilizing selected EEG channels, *Traitement du Signal*, 38(5), 2021, 1259-1270.

[33] S. Hwang, S. Park, D. Kim, J. Lee and H. Byun, Mitigating inter-subject brain signal variability for EEG-based driver fatigue state classification, *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, 2021, 990-994.

[34] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul and E. Chuangsuwanich, Affective EEG-based person identification using the deep learning approach, *IEEE Transactions on Cognitive and Developmental Systems*, 12(3), 2020, 486-496.

[35] H. Qin and P. Wang, Finger-vein verification based on LSTM recurrent neural networks, *Applied Sciences*, 9(8), 2019, 1687.

[36] V. Vijayan and E. Sherly, Real time detection system of driver drowsiness based on representation learning using deep neural networks, *Journal of Intelligent & Fuzzy Systems*, 36(3), 2019, 1977-1985.

[37] F. Wang, Y. Wan, M. Li, H. Huang, L. Li, X. Hou, J. Pan, Z. Wen and J. Li, Recent advances in fatigue detection algorithm based on EEG, *Intelligent Automation & Soft Computing*, 35(3), 2023, 3573-3586.

[38] C. Zhao, C. Zheng, M. Zhao, Y. Tu and J. Liu, Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic, *Expert Systems with Applications*, 38(3), 2011, 1859-1865.

[39] R. Chai, G. R. Naik, T. N. Nguyen, S. H. Ling, Y. Tran, A. Craig and H.T. Nguyen, Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system, *IEEE Journal of Biomedical and Health Informatics*, 21(3), 2017, 715-724.

[40] Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, 521(7553), 2015, 436-444.

[41] T. K. Reddy, V. Arora, V. Gupta, R. Biswas and L. Behera, EEG-based drowsiness detection with fuzzy independent phase-locking value representations using Lagrangian-based deep neural networks, *IEEE Transactions on Systems, Man, and Cybernetics*, 52(1), 2022, 101-111.

[42] J. R. Paulo, G. Pires and U. J. Nunes, Cross-subject zero calibration driver's drowsiness detection: exploring spatiotemporal image encoding of EEG signals for convolutional neural network classification, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, 2021, 905-915.

[43] Y. Liu, Z. Lan, J. Cui, O. Sourina and W. Müller-Wittig, Inter-subject transfer learning for EEG-based mental fatigue recognition, *Advanced Engineering Informatics*, 46, 2020, 101157.

[44] H. -G. Anneke, S. Jelena, M. Alexander, A. Peter, M. Johannes and R. Schreier David, Maintenance of wakefulness test (MWT) recordings, *Zenodo*, Sep. 27, 2019.

[45] J. McNames, T. Thong and M. Aboy, Impulse rejection filter for artifact removal in spectral analysis of biomedical signals, *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, CA, USA, 2004, 145-148.

[46] M. Misiti, Y. Michel, G. Oppenheim and J. -M. Poggi, *Wavelet Toolbox 4 - User's Guide*, The MathWorks Inc., Massachusetts, USA, 2007.

[47] J. Wagner, J. Kim and E. Andre, From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification, *IEEE International Conference on Multimedia and Expo*, Amsterdam, The Netherlands, 2005, 940-943.

[48] L. Sun and M. Shen, Analysis of non-stationary electroencephalogram using the wavelet transformation, *6th International Conference on Signal Processing,* Beijing, China, 2002, 1520-1523.

[49] S. I. Hjelm, S. Ilstedt and C. Browall, Brainball-using brain activity for cool competition, *NordiCHI2000 Proceedings*, 7(9), 2000, 19.
[50] M. Murugappan, S. Murugappan, Balaganapathy and C. Gerard, Wireless EEG signals based neuromarketing system using Fast Fourier Transform (FFT), *IEEE 10th International Colloquium on Signal Processing and its Applications*, Kuala Lumpur, Malaysia, 2014, 25-30.
[51] Anuradha and G. Gupta, A self explanatory review of decision tree classifiers, *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*, Jaipur, India, 2014, 1-7.