Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
ISSN: 2788-7421

CNN-PS: Electroencephalogram Classification of Brain States Using Hybrid Machine - Deep Learning Approach

Osama Mohsin Abdulaziz [1]*, **O.A. Saltykova** [2]

[1,2] Department of Mechanics and control processes-Peoples' Friendship University of Russia

*Corresponding Author: Osama Mohsin Abdulaziz

DOI: https://doi.org/10.52866/ijcsm.2023.04.04.006
Receive July 2023; Accepted September 2023; Available online October 2023

ABSTRACT: Electroencephalography (EEG) has been used for quite some time as a diagnostic technique in neurology. The goal of this publication is to serve as a resource for researchers interested in applying deep learning methods to EEG data. This paper proposes a unique Hybrid Machine-Deep Learning model that can learn and classify EEG signals on its own. This method allows the model to classify EEG signals of varied sampling frequencies and durations automatically. The proposed model used feature extraction methods from artificial design and performed extensive tests with EEG data collected at varying sample rates to determine how well our suggested model performed. The results show that the Hybrid Machine-Deep Learning strategy significantly improves performance, leading to a remarkable 99.97% classification accuracy. Notably, this method performs exceptionally well when labeling lower-frequency EEG signals (less than 4 Hz). The proposed model has improved consistency and robustness, as shown by this study.

Keywords: Electroencephalography, Deep Learning, Hybrid, Machine Learning, Classification.

1. INTRODUCTION

The human brain's cerebral cortex has a wonderful, flourishing spatiotemporal dynamics that is entirely works on its own. Millions of neurons in the brain communicate with one another via chemical and electrical signals (action potentials). Electroencephalography (EEG) analysis provides crucial insight into brain functions and can be used to diagnose neurological conditions like epilepsy. EEG includes waveforms with a wide range of frequencies, intensities, and spatial distributions. Delta waves are found in range (0.5 Hz to 4 Hz); between 4 and 7.5 Hz, theta is found; between 8 and 13 Hz, alpha is found; between 14 and 40 Hz, and finally, gamma is found over 40 Hz (as shown in table (1)). Abnormal electrical discharge can be seen on an EEG if the cause is a brain condition. To facilitate conversation, electrodes are implanted in the frontal pole (Fp), frontal (F), parietal (P), temporal (T), and occipital (O) regions of the brain [1], [2]. Even numbers represent the right hemisphere of the brain, odd numbers represent the left, and the letter Z represents the center of the brain. Figure (1) depicts the placement and naming of the electrodes.

Table 1. Comparison of brain waves [3].

Wave	Frequency Range	Frequency Level and Description
BETA	14 - 30 Hertz	Awake, normal levels of alertness. Also associated with overactive thinking patterns, stress, anxiety, frustration, and other undesired states. People spend most of their daily life operating at this level.
ALPHA	9 - 13 Hertz	Relaxed, calm levels of mental activity occur at this level. A peaceful state is associated with tranquility and relaxation, which people can achieve through effective relaxation exercises and meditation.
THETA	4 - 8 Hertz	A deeper state of mindfulness associated with creative insight, cognitive & memory enhancement, and feelings of deep connectedness. Also, the level at which people naturally, progress into a sleep state.
DELTA	1 - 3 Hertz	The deepest brainwave level associated with dreamless (non-REM) sleep. Essential for proper restoration of health and immune system. Difficult to achieve this level if overactive at the Beta level.

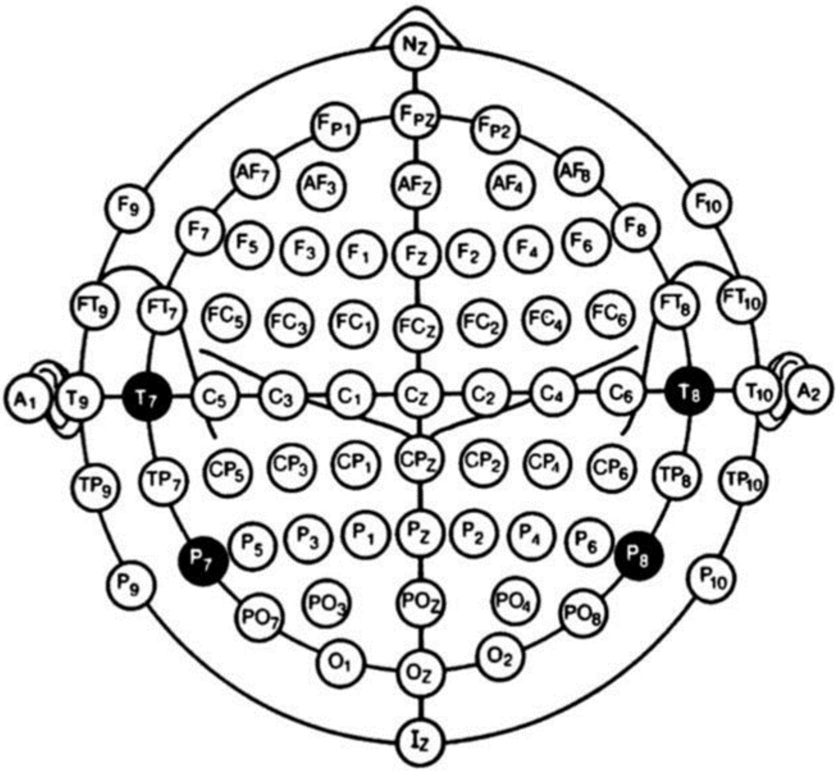

FIGURE.1 Location and identification of electrodes on the scalp surface [5].

EEG has numerous applications in the fields of neural engineering, neurology, and biomedical engineering, including sleep analysis, seizure detection, and BCI (brain computer interface) development. By eliminating the need for specialists to analyze the data, computerized classification of these signals will make EEG more widely available. Artifact elimination, feature extraction, and feature classification are typical steps in an EEG classification pipeline. Simplest of all, an EEG dataset is just a two-dimensional (time and channel) matrix of actual values representing scalp recordings of brain-generated potentials under certain task conditions. EEG data is well-suited to machine learning due to its high degree of organization. Several conventional machine learning and pattern recognition methods have been used to interpret the EEG data. As it relates to neural categorization.

It's worth noting that developing methods for recognizing EEG patterns associated with imaginary leg movements is crucially important for developing BCIs that would aid in the therapy of patients with varying motor disorders following trauma or stroke by means of prostheses, exoskeletons, or anthropomorphic robots [4].

In recent years, many different deep learning architectures have been created to decode EEG signals. In this research a comprehensive literature review on deep learning for EEG categorization. In these investigations, researchers considered a wide range of variables, such as task type, EEG pre-processing techniques, input type, and learning depth. This review compiles the state-of-the-art approaches and performance outcomes for deep learning-based EEG categorization. Recent EEG studies have concentrated on determining how to extract features from EEG signals. As a result, there is a growing need for flexible approaches to EEG classification. Several recent studies have used deep learning algorithms to successfully learn features and classify different types of data [5].

Artificial neural networks (ANNs) are among the most promising and effective technologies for classification of individual EEG sessions. Successfully implementing ANNs calls for thoughtful adjustment of their parameters, which can change dramatically between tasks and domains. One of the main challenges in creating effective ANN-based BCIs is optimizing the EEG input data (by means of dimensionality reduction, filtering, etc.) and channel selection. Dimensionality reduction is typically accomplished by statistical techniques like principal component analysis (PCA) and linear discriminant analysis (LDA), in which the original features are computationally projected onto a lower dimensional space[4].

There has been a significant uptick in the use of deep learning methods for classifying EEG signals. The biopotentials across the scalp that make up an EEG are typically recorded using a DL architecture throughout time. A recurrent neural network (RNN), often an LSTM, is used after a convolutional neural network (CNN) cascade. The earliest tiers of these cascade structures must behave as feature extractors for the latter layers to use [6]. Numerous research has employed a computer classification model to categorize features collected from EEG signals, thanks to advancements in computer science and technology. Typical procedures for this kind of study include collecting and preprocessing EEG data, developing a classification model, analyzing the data, and making predictions. One of the most crucial processes involves extracting features from EEG data. Different methods are used to extract EEG data, such as time-domain, frequency-domain, time-frequency analyses, and chaotic features. In addition, several studies have combined or altered these methods to extract other information, leading to effective classification. [7],[9].

The accuracy of medical EEG acquisition equipment has increased as a result of scientific and technological advancements. On top of that, there is now some portable EEG acquisition gear available. Emotive, for instance, has seen extensive application in brain-computer interface [11],[13] due to its low price, portability, and comparable performance to those of medical devices. While there is a wealth of EEG data that can be used for epilepsy research thanks to the proliferation of medical and portable EEG acquisition devices, this data is not standard in format due to differences in sample frequencies, signal durations, and sampling channels. The mismatch of data specifications often affects the features obtained by traditional feature extraction algorithms. This situation raises doubts about the flexibility of classification methods to incorporate new data into their analysis. The increased detection and recognition of EEG data necessitates a more generalizable approach to categorization [10],[12]. The primary contributions of the proposed study can be succinctly outlined as follows:

- This paper presents a novel Hybrid Machine-Deep learning model for autonomous learning and classification of EEG signals.
- This paper aims to devise a versatile methodology that can effectively classify electroencephalogram (EEG) signals with diverse sampling frequencies, signal durations, and sampling channels.
- This paper aims to examine the effects of portable electroencephalogram (EEG) acquisition devices on data heterogeneity and the consequent requirement for more adaptable classification techniques.
- This proposal presents a model that aims to decrease the dependence on manual feature design and extraction by utilizing the benefits of data-driven deep learning techniques.

At the moment, deep learning technology is a hot topic in academia. Because of its data-driven autonomy, this technology may forego the laborious steps involved in manually designing features and extracting information in conventional approaches, as well as the associated challenges and the need to manually change a plethora of parameters. The remainder of this work is structured as follows. Section 2 applies previously established feature designs to the data and reveals the limitations of the currently used categorization techniques are presented. In Section 3, details the proposed network model, training approaches, and data processing techniques. Section 4 focuses on the model's strengths, while Section 5 presents and discusses the model's results. Section 6 offers the final conclusions.

2. RELATED WORK

Many tasks that are challenging but not impossible to do using traditional methods can be completed with the help of in-depth learning technologies [13], [14]. Researchers have used deep neural networks to analyze EEG [15]. Tabar and Halici [16] used a short-time Fourier transform to translate brainwaves from one dimension (1D) to two-dimensional (2D) picture data, then accessed the deep network for categorization. Bashivan et al. [17] used spectral power to classify images into depth networks after converting frequency bands collected from brain waves into topographical maps (2D images). In order to present a solution for epilepsy prevention and control. All of the preceding examples demonstrate a wide variety of architectural approaches After investigating a variety of EEG characteristics and neural network topologies, Jirayucharoensak et al. [5] found an accuracy of 54%, suggesting that stacked auto encoders (SAEs) are not a good option for this task. They examined the effects of combining principal component analysis (PCA), covariate shift adaptation (CSA), and power spectrum density (PSD) traits, among others, but found that none of them improved upon the limitations of SAEs for this dataset. An SAE and many deep belief networks (DBNs) were evaluated for accuracy in [18], [19], and the DBN with 3 restricted boltzmann machines (RBMs) was found to be the most accurate.

Five of this research used models with convolutional layers. Two of the five used a hybrid architecture in which the CNN fed into LSTM RNN modules, however neither of these approaches achieved an accuracy of greater than 75%. Input formulation discrepancies are likely to blame for the discrepancy in accuracy observed across the three reference CNN experiments. While [20] fed signal values into the neural network, [21]and [22] processed the information into Fourier feature maps and 3D grids, respectively, and attained accuracies of 87% and 88%, respectively. For these CNN designs, and architectures with two convolutional layers, each with one or two dense layers. Used only the CNN architecture [20], which is on par with the performance of a typical CNN trained on signal values; nevertheless, the input formulation may have been the main contributor to the disparity. While the deep CNN just used signal values without any additional preprocessing, the EEG-based emotion recognition (EEG-ER) required extensive input preprocessing using PSD characteristics and pre-frontal asymmetry channel selection.

The solitary RNN-only (no convolutional layers) architecture in this set [23] was made up of two LSTM layers and a single dense layer. In this investigation, accuracy of 87% was achieved utilizing simply signal values as input. two hybrid convolutional recurrent models, will produce better classification accuracy. It was found that a deep learning recurrent architecture lacking convolutional layers outperformed those that included recurrent and convolutional layers for this dataset. Recording EEG from test subjects is standard procedure in experiments aiming to detect and categorize event-related potentials while presenting a visual display. The electroencephalograms of sleeping humans record the electrical activity of their brains. The sleep stage rating task has received the fewest number of studies. In order to classify the signals more easily, they were split into three groups: deep, light, and sleep. The ultimate aim of this research is to lessen the burden of sleep interpretation and analysis on medical professionals.

3. METHODOLOGY

This part begins by outlining the six ML and CNN-based model structure, as well as the training methods for varying lengths of sample data. Statistical analysis has a rich history of utilizing feature reduction and its related themes of feature selection, feature extraction, and dimension reduction. The primary motivation for using feature reduction has been the need to lessen the volume and complexity of data, hence facilitating faster, cheaper, and less complicated analysis. Different techniques and concepts have been applied to the problem of feature reduction. In this category of algorithms, Principal component analysis (PCA) and singular value decomposition (SVD) are the most effect in machine analysis.

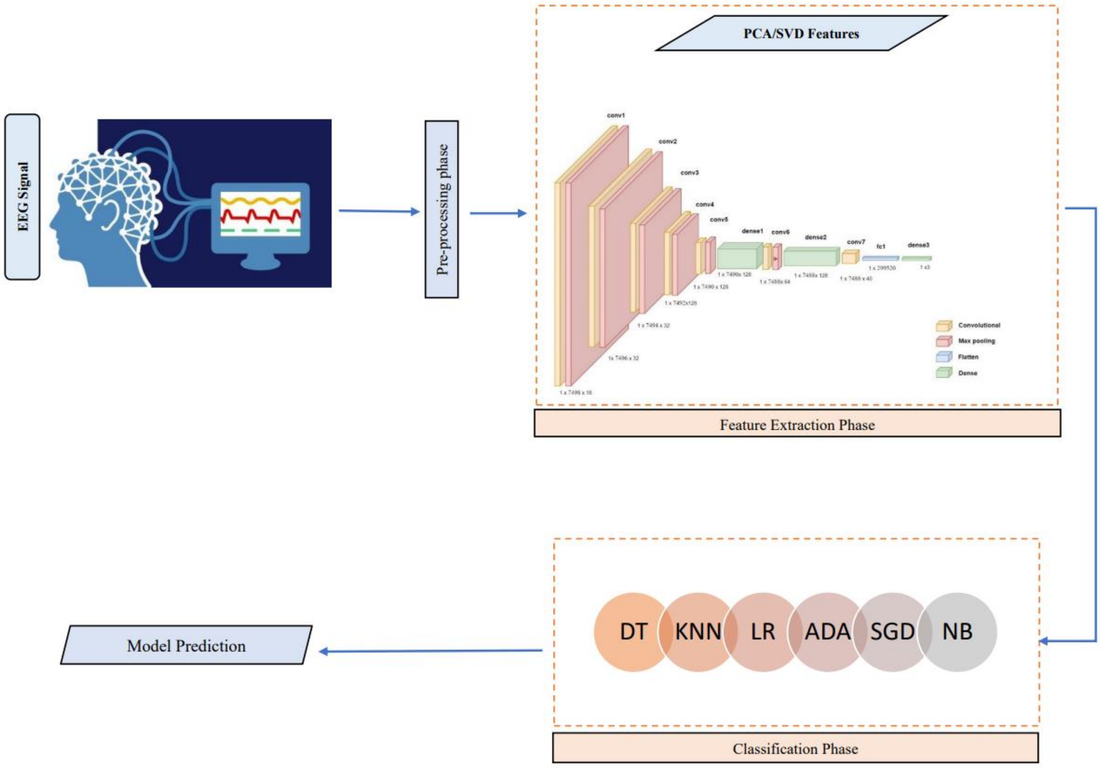

FIGURE. 2 The system's block diagram.

3.1 DATASET

A study conducted between 1987 and 1991 looked at the impact of aging on sleep in 153 healthy Caucasian adults ages 25 to 101 who did not use any form of sleep medicine. Two polysomnograms (PSGs), each lasting around 20 hours, were recorded at the patients' homes on two separate day-night cycles. The subjects carried on with their daily routines while also wearing a modified Walkman-like cassette-tape recorder. Subject numbers (ss) and night (N) are appended to filenames (SC4ssNEO-PSG.edf) to create unique filenames for each subject per night. Due to a malfunctioning cassette or laserdisc, the first nights of subjects 36 and 52, and the second night of topic 13, were erased. Each of the signals, EOG and EEG, were captured at 100 Hz. After digitally high pass filtering, rectifying, and low-pass filtering the submental-EMG signal, the resulting EMG envelope was recorded at 1Hz and represented in rms (root-mean-square). Furthermore, 1Hz samples were taken of or nasal airflow, rectal body temperature, and the event marker. As detailed in [24],[25].

3.2 FEATURES EXTRACTION

3.2.1 Principal Component Analysis

Principal component analysis (PCA) is commonly used to discover patterns in high dimensional data. The information theory method behind PCA's goal is that it uses a smaller group of typical feature pictures (called Eigenobject) to represent both known and unknown faces. The statistical data on PCA used in face recognition technology shows how useful it is for recognizing and verifying facial features. The PCA approach requires a transformation from the 2-Dimensional facial image matrices to a 1-Dimensional vector. The row or column orientation of the 1-dimensional vector is irrelevant [22, 23].

3.2.2 Singular Value Decomposition

As with PCA, Singular Value Decomposition (SVD) can be used to break down data. Feature extraction of a signal, matrix approximation, and pattern identification are just a few of its many uses in signal processing and statistics. However, PCA cannot extract features from a single signal, nor can it provide information about the features present in a signal of varying frequencies. It is often the case that the differences between physiological states are masked by differences in frequency, so extracting features using SVD rather than PCA can lead to a more complete data set as shown in Eq. (1)

$$X = LSR^T = [l_1 \dots \; l_p] \begin{bmatrix} \mathbf{s_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{s_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{s_q} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} -\mathbf{r_1^T} \\ -\mathbf{r_2^T} \\ \vdots \\ -\mathbf{r_q^T} \end{bmatrix} \tag{1}$$

3.2.3 Convolutional Neural Networks

The design of convolutional neural network (CNN) is the foundation for several feature extraction techniques. In contrast, the classification effect based on the general feature extraction method is unstable when the data changes. In this research, a CNN-based classification model was developed that could learn and categorize data characteristics on its own. This model performed both the feature extraction and classification processes on its own (see Figure (2)). It aims to achieve reliable classification results regardless of the size or frequency of the sample data. On the left, we see a two-stage classification procedure that uses artificial design aspects to make a determination. Figure (3) depicts the right side of the network model, where data is input and the classification results are produced without any intermediate steps.

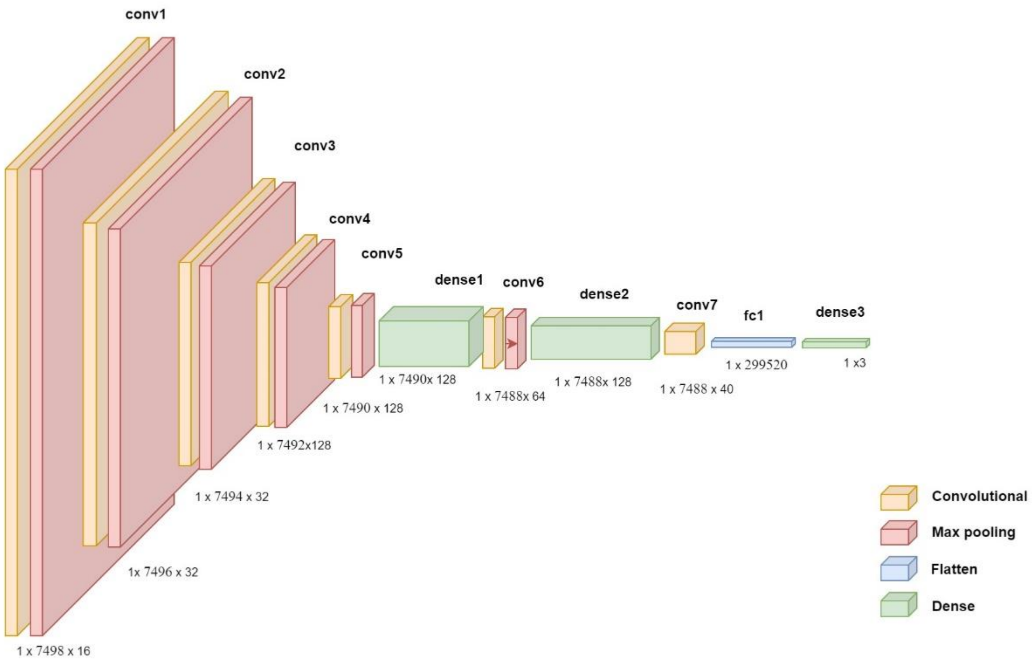

FIGURE 3. Layers of the proposed Hybrid CNN feature extractor model.

The CNN is a feed-forward neural network that enhances pattern categorization performance via posterior probability. Key components of the network include the convolutional, pooling, fully connected, and softmax layers. The feature map is generated by passing the input signal data through a series of convolution kernels in the convolution layer (i.e., number of convolution kernels equals the number of feature maps). The feature map generated by the convolution operation of the previous layer is down sampled in the pooling layer. Iterating the convolutional and pooling layers causes the network to grow in depth. Classification results are then output via the softmax layer (see Table 2), which is made possible by the fully connected layer, which links all feature maps from the previous layer to the hidden layer of a shared neural network. Classifying EEG data is the focus of this research, which recommends using a multilayer network with cubic iterative convolutional and pooling layers, a fully connected layer, and a softmax layer. To create its output sample data, the model categorizes one-dimensional EEG data from a single channel. In machine learning, the convolutional layer is analogous to the feature extractor. By convolving x with several convolution kernels, this layer generates many feature maps that can preserve the essential features of the input signal.

Table 2 CNN-PS Layers Parameters

Layer (type)	Output Shape	Param #
conv1d-1 (Conv1D)	(None, 7498, 16)	64
maxpooling1d-1 (MaxPooling1)	(None, 7498, 16)	0
conv1d-2 (Conv1D)	(None, 7496, 32)	1568
maxpooling1d-2 (MaxPooling1)	(None, 7496, 32)	0
conv1d-3 (Conv1D)	(None, 7494, 32)	3104
maxpooling1d-3 (MaxPooling1)	(None, 7494, 32)	0
conv1d-4 (Conv1D)	(None, 7492, 128)	12416
maxpooling1d-4 (MaxPooling1)	(None, 7492 128)	0
conv1d-5 (Conv1D)	(None, 7490, 128)	49280
maxpooling1d-5 (MaxPooling1)	(None, 7490, 128)	0
dense-1 (Dense)	(None, 7490, 128)	16512
conv1d-6 (Conv1D)	(None, 7488, 64)	24640
maxpooling1d-6 (MaxPooling1)	(None, 7488, 64)	0
dense-2 (Dense)	(None, 7488, 128)	8320
conv1d-7 (Conv1D)	(None, 7488, 40)	5160
flatten-1 (Flatten)	(None, 299520)	0
dense-3 (Dense)	(None, 3)	898563

3.3 CLASSIFICATION BASED MACHINE LEARNING ALGORITHMS

After converting the data from the flight data set to numeric values that are part of 12 different classes, six different classifiers were employed in sequence to predict the EEG signal. Each category in this classifier represents a different attribute. One approach for classifying a classification system's expected outcome is to use a numerical or binary array format. The classifiers were performed linearly as follows:

[1] The decision tree has a single root node, numerous branches, and numerous leaf nodes. This strategy begins with the user splitting the data into progressively smaller subsets until a decision tree with nodes and leaves is completed. Each branch represents a class, each leaf represents a specific attribute, and the test is used to obtain that class or attribute. The root node is located at the top of the tree [26], [27]. The rule for the decision tree is shown in Eq.(2).

$$Entropy(t) = -\sum p(i|t) log2\ p(i|t)\ c-1\ i=0 \quad (2)$$

Where c is the number of classes, $p\ (i\,|\,t)$ denotes the fraction of records belonging to class i at a given node t.

[2] Naive Bayes NB: It is an algorithm for learning under supervision. Using the data's frequency and collection sizes to derive a set of probabilities. Based on the Nave Bayes method, the posterior probability of a document d belonging to class c is provided as in Eq. (3) and (4) [24], [9].

$$P(c|d) = P(d|c)\ P(c)\ p(d) \quad (3)$$

$$P(c|d) = \quad P(w1, w2, \ldots wn|c)\ p(c)\ p(d) \quad (4)$$

Where: P(d|c) is the likelihood which is the probability of the predictor given class.

P (wi |c) is the qualified probability of term wi taking place in document d of class c. P (wi |c) denotes to a measure of how much wi contributes, and c is the accurate class. (w1, w2, ..., wn) are the tokens in document d and are part of the vocabulary used for classification, and n is the number of such tokens in document d.

[3] Logistic regression used to sort information into categories. Depending on the values of the input variables, it determines the likelihood of an event occurring (in terms of 0 and 1). Logistic regression can be used to make predictions with a binomial outcome, such as whether or not an email is spam. Also, categorical dependent variables can be predicted with logistic regression .Linear regression is used to predict the values of continuous variables, such as the price of real estate over three years [28], [26]. Logistic regression's equation is displayed in Eq. (5).

$$logy = 1 / (1 + e\text{^}(-(b0 + b1*x1 + b2*x2)) \quad (5)$$

Where b0, b1, and b2 are the coefficients, X1 and X2 are the features (input value) [27].

[4] The K-nearest neighbor algorithm is a classification algorithm. It uses a database of already-classified data points, and treats the categorization of the sample data point as a classification problem. KNN is considered non-parametric due to the fact that it does not presume anything about the distribution of the underlying data. The KNN algorithm has a number of advantages, including the fact that it is a straightforward method that is easy to put into practice. The model's creation didn't break the bank. Because of its adaptability, this classification method works particularly well with groups that share characteristics across multiple media types. When it comes to predicting a function based on expression profiles, this method can be the most accurate [29], [30]. Eq. (6) represents the standard human conceptualization of distance in the physical world, it has been used to represent the Euclidean distance.

$$D\ euclidean(x,y) = \sqrt{\sum}\ (xi-yi)2\ m\ i=1 \quad (6)$$

Where m is the total number of distinct words in the set of documents, xi represents the importance of the term i in document x, and yi represents its importance in document y.

[5] Adaptive boosting of an adaptive boosting ML method was used to improve the classification results. Meta-algorithms are useful when used in combination with other learning algorithms because they can boost the overall performance of learning algorithms. It is adaptive in that subsequent classifiers constructed have been modified to examples that have been misclassified by earlier classifiers. In other words, the fundamental principle of ADA is to repeatedly use a weak classifier, and then tweak the weight given to each example with each call. In this way, Eq. (7) the misclassified examples will be weighted higher than the correctly classified ones, causing the new classifier to favor the misclassified examples [31], [32].

$$H(x) = \sum_{i=1}^{T} \alpha_i h_i(x) \quad (7)$$

That is, H uses a linear combination of the decisions of each of the hi hypotheses in the ensemble. The AdaBoost algorithm sequentially chooses hi from H and assigns this hypothesis a weight αi
.

[6] The stochastic gradient descent model is a highly effective learning algorithm for linear classifiers. Simply replace the real gradient obtained from the full dataset with an approximation based on a randomly selected portion. A stochastic (or “operational”) gradient descent algorithm assigns a gradient to each learning element, approximating the gradient of the cost function. The settings were adjusted to reflect estimated gradients. The model parameters were recalculated after each learning object. The stochastic gradient descent technique is

much faster than the normal gradient descent for large datasets [33],[34]. This model is an effective form of facilitation. Easily digestible SGD updates as follows in Eq. (8) :

$$\theta^{(t+1)} = \theta^{(t)} - \alpha_t \nabla l_i(\theta^{(t)}) \tag{8}$$

Where: t represents the iteration and $\boldsymbol{\theta}$ represents the parameter updates, where $\boldsymbol{\alpha}$ is the learning size. In this instance, the value of the index I will be picked at random before each iteration. In practice, we frequently shuffle the samples in a random manner and then proceed to go through them in order [26].

4 EVALUATION METRICS

At this point, the model's classification accuracy was determined. By contrasting the true class labels with the anticipated ones, the classifier's accuracy was measured. Correctly classified instances (true positives), correctly classified but irrelevant examples (true negatives), incorrectly classed examples (false positives), and unclassified examples can all be used to calculate a classification system's accuracy (false negatives)[35]. Quantitative analysis measurements were as follows:

• Accuracy, or the degree to which a model is likely to accurately predict outcomes, is defined by the proportion of correct predictions relative to the total number of predictions, as shown in Eq. (9):

$$Accurcy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

• Precision refers to how accurately a group of documents describes its subject, and thus how precisely they were classified. Class ci, symbolized by the symbol (Pi), has an accuracy that can be quantified as follows, as shown in Eq. (10):

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{10}$$

• Recall measures how well a classifier can identify documents as belonging to a given class (as demonstrated by Eq (11). Class ci recall (Ri) can be calculated using the formula:

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{11}$$

In this case, $\boldsymbol{TP_i}$ points to a true-positive value. $\boldsymbol{FP_i}$ stands for false positives and $\boldsymbol{FN_i}$ represents false negatives.

• F1 is the precision-recall synchronization rate. Overall system performance is good if F1 is high. Given Eqs. (12) and (13), the following is a description of F1:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{12}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{13}$$

5 RESULT AND DISCUSSION

This section, will first have a look at the procedures that were followed to get the data for this study ready. After that, we'll get into the big-picture classifications of tasks, input formulations, and architectural movements. An application employing a freely available dataset for evaluating multiple deep learning platforms rounds out the findings. Hybrid Machine - Deep Learning models, like the one seen in Figure (1). Hybrid Machine - Deep Learning uses numerous thin layers in conjunction with a deep learning technique. Following fully linked layers in terms of study count was the convolutional neural network (CNN).

Comparison of Various Classification Strategies Utilizing Constructed Features. Using an artificial design feature extraction strategy, select features or create new features for classification. Following the feature extraction process, many well-known classifiers are selected from the scikit-learn library. K-Nearest Neighbors, Stochastic Gradient

Descent, Decision Trees, Logistic Regression, and Naive Bayes are all examples of classifiers. The parameters of these classifiers are kept in a place outside from the library data. Tables (3) and (4) use the aforementioned features and classifiers, with the average accuracy of each classifier's classifications shown in the final column of AVG. Popular classifier KNN has high levels of classification accuracy, demonstrating the efficacy of the feature extraction methods. Traditional classification approaches based on artificial design feature produce varying classification outcomes in various classifiers, as shown in Table (3). At low sampling rates, classification stability is compromised. Compared to other methods, the KNN has a greater average accuracy (see Table 3). This finding demonstrates that the artificial design feature-based classification approach yields the best classification outcomes. However, there is still considerable variation in the categorization accuracy of data collected using a variety of sampling frequencies.

TABLE 3. PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS .

Classifiers		Precision	Recall	F1-score
NB	weighted avg	0.81	0.75	0.78
SGD	weighted avg	0.87	0.85	0.85
ADA	weighted avg	0.90	0.85	0.84
DT	weighted avg	0.89	0.89	0.89
LR	weighted avg	0.86	0.83	0.85
KNN	weighted avg	0.99	0.98	0.97

The model provides more consistent classification results and higher classification accuracy, as seen in Tables (4).

Table 4. Applied Ml Algorithms Accuracy Results.

Classifiers	Accuracy
NB	0.85
SGD	0.88
ADA	0.85
DT	0.89
LR	0.83
KNN	0.99

According to what researchers generally know, in ML, there are two crucial factors to consider: the time it takes to create the model and the quality of the ML algorithm that is being used to develop it. In this study, the accuracy was first evaluated as a metric for evaluating and analyzing the ML method, and then the second metric is considered, which is the amount of time it takes to construct the model. Because computational complexity is now the most important and demanding challenge in ML, the time required to develop a model seems to be the most significant and critical problem in ML. However, it is critical to keep track of the amount of time required to construct the model. As a result, when the time to construct the model is considered, the DT ML method is more effective than the other ML techniques. However. According to the results of the experimental research, KNN algorithm is the most successful of the six ML algorithms tested for EGG signal prediction.

The results of the various methodologies were compared and confirmed based on characteristics such as class precision and recall. Tables (3) , (4) and Figure (4) demonstrate the results of all the approaches used. Regarding the trustworthiness of outcomes, class recall is a crucial factor to consider. The findings revealed that the proposed model was capable of predicting EEG signals with an accuracy of 99.97%, which is a reasonable degree of accuracy.

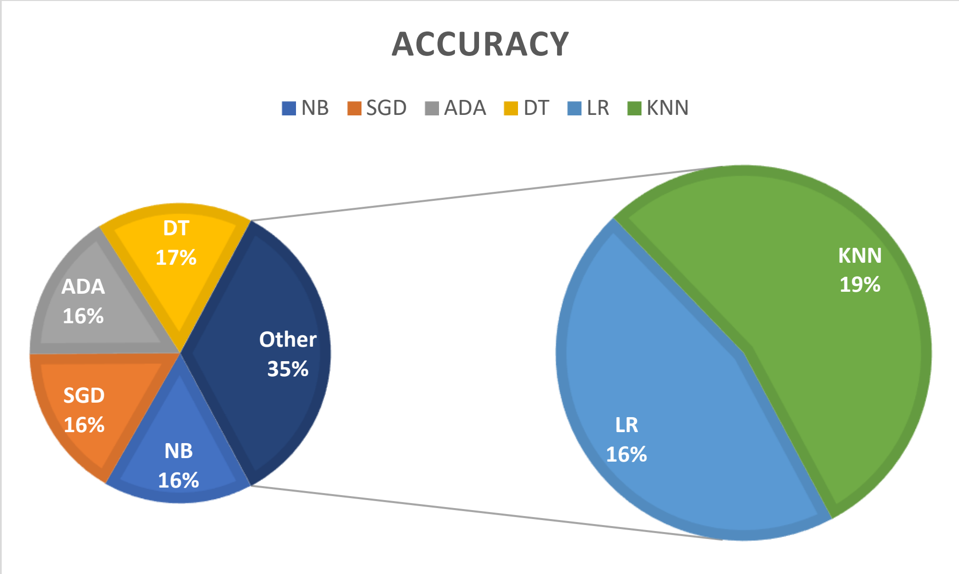

FIGURE 4. applied ml algorithms accuracy percentage.

This is useful for boosting the machine performance. However, the overall performance of the ML algorithm used is quite efficient in terms of accuracy, precision, recall, and time required to build the model. Our proposed techniques yield promising results, but an in-depth study shows some insightful and important information that may be deployed to efficient ML choice and the effectiveness of ML algorithm. These insights and valuable information are provided in conclusion.

6 COMPARISON RESULTS

In table 5 shows the results of a thorough comparison of several classifiers used for EEG categorization. The accuracy gained by Incremental Attribute Learning was 76%, while that of Convolutional Neural Network (CNN) was 90% and that of Artificial Neural Network (ANN) was 87%. K-Star and K-Nearest Neighbor (KNN) both had accuracies of 94%, while the combination of Factorization Machine (FM) and Long Short-Term Memory (LSTM) got to 93%. Accuracy rates of 80% to 85% were attained by other methods, such as a combination of KNN and ANN. The accuracy of the Support Vector Machine (SVM) ranged from 80.9 to 91.1 percent. An astounding 99% accuracy sets apart the proposed solution, a Hybrid CNN-KNN strategy. Combining deep learning methods (CNN) with the knowledge-based neural network (KNN) approach is shown to be beneficial in this method for EEG categorization. The suggested technique outperforms all competitive classifiers in terms of accuracy, demonstrating its promise for reliable and precise EEG data classification. This study's findings demonstrate the efficacy of the suggested Hybrid CNN-KNN model in classifying EEG data, making a significant contribution to the study of EEG-based brain state categorization.

Table 5. Accuracy comparison .

$F1$	$F1$	$F1$
[40]	Incremental Attribute Learning	76%
[41]	Convolutional Neural Network	90%
[36]	Artificial Neural Network	87%
[38]	Factorization Machine (FM) and Long Short-Term Memory (LSTM)	93%
[42]	K- Star	94%
[37]	K-Nearest Neighbor, Artificial Neural Network	80%-85%
[43]	Support Vector Machine	80%-91%
[39]	K-Nearest Neighbor	98%
Proposed Method	Hybrid CNN-KNN	99%

The proposed model using the KNN algorithm gives the highest classification accuracy compared to the related studies.

7 CONCLUSION

Movement imaging and seizure detection are just two of the experiments where deep learning for EEG classification has proven its worth. This work sheds light on how deep learning can be applied to EEG datasets in the future. Designs for deep network research have varied as a result of the many possible input formulations and network topologies. Classification accuracy and transfer learning are two areas where PCA/SVD and convolutional layers with recurrent layers show improvement over more traditional approaches. Although most studies on EEG labeling have concentrated on increasing precision rather than transferability, it is essential to recognize that the latter merits additional exploration. To overcome the difficulty of identifying EEG signals with different sample rates and durations, this study develops a CNN-PS classification model. The suggested model can efficiently process a wide variety of EEG data with a outstanding 99% accuracy rate. We also investigated issues with applying a single classification strategy based on feature extraction to EEG signals of varying sampling frequencies. Especially when working with small samples, our findings show the limitations of conventional methods that place a premium on feature design and selection. The CNN-PS model, however, gets over these restrictions by automatically learning the features of the sample data and adjusting to different data lengths with the help of effective data completion techniques. The study's shortcomings and strengths should be considered alongside its useful insights. First, the CNN-PS model evaluation may have a limited scope because it was conducted on only two datasets. It is recommended that future studies expand the number and variety of EEG datasets used to verify the model. Improving the CNN network's ability to display its learned characteristics may also help the CNN-PS model perform better. Finally, integrating new factors and researching alternative machine learning approaches should be investigated to further improve the robustness and application of the suggested strategy. In conclusion, this research makes a significant contribution to the field of EEG categorization by recommending a Hybrid Machine-Deep Learning model that can successfully process EEG data with a wide range of properties. The results show that the model is remarkably accurate and has broad application potential. Deep learning approaches have the potential to greatly improve our understanding of and practical use of EEG categorization, but only if we address the mentioned constraints and pursue future research areas.

Funding

None

ACKNOWLEDGEMENT

The accessibility of datasets played a pivotal role in facilitating the experimentation process and verifying the efficacy of the Hybrid Machine-Deep Learning framework.

CONFLICTS OF INTEREST

None

REFERENCES

[1] L. Boubchir, B. Daachi, and V. Pangracious, “A review of feature extraction for EEG epileptic seizure detection and classification,” in 2017 40th International Conference on Telecommunications and Signal Processing, TSP 2017, 2017, vol. 2017-January. doi: 10.1109/TSP.2017.8076027.

[2] R. Jenke, A. Peer, and M. Buss, “Feature extraction and selection for emotion recognition from EEG,” IEEE Trans Affect Comput, vol. 5, no. 3, 2014, doi: 10.1109/TAFFC.2014.2339834.

[3] W. Srimaharaj, R. Chaisricharoen, S. Chaising, and P. Sittiprapaporn, “Classification of human brain attention focused on meditation, effected by L-theanine acid in Oolong tea,” in 3rd International Conference on Digital Arts, Media and Technology, ICDAMT 2018, 2018. doi: 10.1109/ICDAMT.2018.8376536.

[4] V. A. Maksimenko et al., “Artificial neural network classification of motor-related EEG: An increase in classification accuracy by reducing signal complexity,” Complexity, vol. 2018, 2018, doi: 10.1155/2018/9385947.

[5] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, “EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation,” Scientific World Journal, vol. 2014, 2014, doi: 10.1155/2014/627892.

[6] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. W. Koh, “Application of entropies for automated diagnosis of epilepsy using EEG signals: A review,” Knowledge-Based Systems, vol. 88. 2015. doi: 10.1016/j.knosys.2015.08.004.

[7] R. Sharma and R. B. Pachori, “Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions,” Expert Syst Appl, vol. 42, no. 3, 2015, doi: 10.1016/j.eswa.2014.08.030.

[8] T. Wen, Z. Zhang, M. Qiu, M. Zeng, and W. Luo, “A two-dimensional matrix image based feature extraction method for classification of sEMG: A comparative analysis based on SVM, KNN and RBF-NN,” J Xray Sci Technol, vol. 25, no. 2, 2017, doi: 10.3233/XST-17260.

[9] T. Wen and Z. Zhang, "Effective and extensible feature extraction method using genetic algorithm-based frequency-domain feature search for epileptic EEG multiclassification," Medicine (United States), vol. 96, no. 19, 2017, doi: 10.1097/MD.0000000000006879.

[10] R. Abiri, X. Zhao, and Y. Jiang, "A Real Time EEG-Based Neurofeedback platform for Attention Training," Biomedical Engineering Society Annual Meeting, 2016.

[11] H. H. Kha, V. A. Kha, and D. Q. Hung, "Brainwave-controlled applications with the Emotiv EPOC using support vector machine," in Proceedings - 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2016, 2017. doi: 10.1109/ICITACEE.2016.7892420.

[12] M. Duvinage et al., "A P300-based quantitative comparison between the emotiv epoc headset and a medical EEG device," in Proceedings of the 9th IASTED International Conference on Biomedical Engineering, BioMed 2012, 2012. doi: 10.2316/P.2012.764-071.

[13] M. W. Nadeem et al., "Brain tumor analysis empowered with deep learning: A review, taxonomy, and future challenges," Brain Sciences, vol. 10, no. 2. 2020. doi: 10.3390/brainsci10020118.

[14] H. Li, Y. Pan, J. Zhao, and L. Zhang, "Skin disease diagnosis with deep learning: A review," Neurocomputing, vol. 464, 2021, doi: 10.1016/j.neucom.2021.08.096.

[15] Z. Tang, G. Zhao, and T. Ouyang, "Two-phase deep learning model for short-term wind direction forecasting," Renew Energy, vol. 173, 2021, doi: 10.1016/j.renene.2021.04.041.

[16] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," J Neural Eng, vol. 14, no. 1, 2017, doi: 10.1088/1741-2560/14/1/016003.

[17] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, 2016.

[18] X. Jia, K. Li, X. Li, and A. Zhang, "A novel semi-supervised deep learning framework for affective state recognition on EEG signals," in Proceedings - IEEE 14th International Conference on Bioinformatics and Bioengineering, BIBE 2014, 2014. doi: 10.1109/BIBE.2014.26.

[19] H. Xu and K. N. Plataniotis, "Affective states classification using EEG and semi-supervised deep learning approaches," in 2016 IEEE 18th International Workshop on Multimedia Signal Processing, MMSP 2016, 2017. doi: 10.1109/MMSP.2016.7813351.

[20] M. Yanagimoto and C. Sugimoto, "Recognition of persisting emotional valence from EEG using convolutional neural networks," in 2016 IEEE 9th International Workshop on Computational Intelligence and Applications, IWCIA 2016 - Proceedings, 2017. doi: 10.1109/IWCIA.2016.7805744.

[21] R. Qiao, C. Qing, T. Zhang, X. Xing, and X. Xu, "A novel deep-learning based framework for multi-subject emotion recognition," in ICCSS 2017 - 2017 International Conference on Information, Cybernetics, and Computational Social Systems, 2017. doi: 10.1109/ICCSS.2017.8091408.

[22] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby Shalaby, "EEG-based emotion recognition using 3D convolutional neural networks," International Journal of Advanced Computer Science and Applications, vol. 9, no. 8, 2018, doi: 10.14569/ijacsa.2018.090843.

[23] S. Alhagry, A. Aly, and R. A., "Emotion Recognition based on EEG using LSTM Recurrent Neural Network," International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, 2017, doi: 10.14569/ijacsa.2017.081046.

[24] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberyé, "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," IEEE Trans Biomed Eng, vol. 47, no. 9, 2000, doi: 10.1109/10.867928.

[25] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," Circulation, vol. 101, no. 23, 2000, doi: 10.1161/01.cir.101.23.e215.

[26] S. K. Singh, R. W. Taylor, M. M. Rahman, and B. Pradhan, "Developing robust arsenic awareness prediction models using machine learning algorithms," J Environ Manage, vol. 211, pp. 125–137, Apr. 2018, doi: 10.1016/j.jenvman.2018.01.044.

[27] P. Sokkhey and T. Okazaki, "Hybrid machine learning algorithms for predicting academic performance," International Journal of Advanced Computer Science and Applications, vol. 11, no. 1, pp. 32–41, 2020, doi: 10.14569/ijacsa.2020.0110104.

[28] H. R. Pourghasemi, A. Gayen, S. Park, C.-W. Lee, and S. Lee, "Assessment of Landslide-Prone Areas and Their Zonation Using Logistic Regression, LogitBoost, and NaïveBayes Machine-Learning Algorithms," Sustainability (Switzerland), 2018, doi: 10.3390/su10103697.

[29] G. Alimjan, T. Sun, Y. Liang, H. Jumahun, and Y. Guan, "A New Technique for Remote Sensing Image Classification Based on Combinatorial Algorithm of SVM and KNN," Intern J Pattern Recognit Artif Intell, vol. 32, no. 7, Jul. 2018, doi: 10.1142/S0218001418590127.

[30] M. M. Hasan, M. T. Zahara, M. M. Sykot, A. U. Nur, M. Saifuzzaman, and R. Hafiz, "Ascertaining the Fluctuation of Rice Price in Bangladesh Using Machine Learning Approach," in 2020 11th International

Conference on Computing, Communication and Networking Technologies, ICCCNT 2020, 2020. doi: 10.1109/ICCCNT49239.2020.9225468.

[31] A. Taherkhani, G. Cosma, and T. M. McGinnity, “AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning,” Neurocomputing, vol. 404, pp. 351–366, Sep. 2020, doi: 10.1016/j.neucom.2020.03.064.

[32] A. Shahraki, M. Abbasi, and Ø. Haugen, “Engineering Applications of Artificial Intelligence Boosting algorithms for network intrusion detection : A comparative evaluation of Real AdaBoost , Gentle AdaBoost and Modest AdaBoost,” Eng Appl Artif Intell, vol. 94, no. February, p. 103770, 2020, doi: 10.1016/j.engappai.2020.103770.

[33] Y. Zhang, A. M. Saxe, M. S. Advani, and A. A. Lee, “Energy–entropy competition and the effectiveness of stochastic gradient descent in machine learning,” Mol Phys, vol. 116, no. 21–22, pp. 3214–3223, Nov. 2018, doi: 10.1080/00268976.2018.1483535.

[34] H. M. Fadhil, M. N. Abdullah, and M. I. Younis, “A Framework for Predicting Airfare Prices Using Machine Learning,” Iraqi Journal of Computers, Communications, Control and Systems Engineering, vol. 22, no. 3, 2022, doi: 10.33103/uot.ijccce.22.3.8

[35] A. Bibi et al., “Spam Mail Scanning Using Machine Learning Algorithm,” J Comput (Taipei), vol. 73, no. 2, 2020, doi: 10.17706/jcp.15.2.73-84.

[36] S. S. Poorna, V. v. Arsha, P. T. A. Aparna, P. Gopal, and G. J. Nair, “Drowsiness detection for safe driving using PCA EEG signals,” in Advances in Intelligent Systems and Computing, 2018, vol. 710. doi: 10.1007/978-981-10-
7871-2_40.

[37] P. Ma and Q. Gao, “EEG Signal and Feature Interaction Modeling-Based Eye Behavior Prediction Research,” Comput Math Methods Med, vol. 2020, 2020, doi: 10.1155/2020/2801015.

[38] V. V. Priya and M. Uma, “EEG based Drowsiness Prediction Using Machine Learning Approach,” Webology, vol. 18, no. 2, 2021, doi: 10.14704/web/v18i2/web18351.

[39] T. Wang, S. U. Guan, K. L. Man, and T. O. Ting, “EEG eye state identification using incremental attribute learning with time-series classification,” Math Probl Eng, vol. 2014, 2014, doi: 10.1155/2014/365101.

[40] S. Chaabene, B. Bouaziz, A. Boudaya, A. Hökelmann, A. Ammar, and L. Chaari, “Convolutional neural network for drowsiness detection using eeg signals,” Sensors, vol. 21, no. 5, 2021, doi: 10.3390/s21051734.

[41] D. S. Benitez, S. Toscano, and A. Silva, “On the use of the Emotiv EPOC neuroheadset as a low cost alternative for EEG signal acquisition,” in 2016 IEEE Colombian Conference on Communications and Computing, COLCOM 2016 - Conference Proceedings, 2016. doi: 10.1109/ColComCon.2016.7516380.

[42] R. Oliver and D. Suendermann, “A First Step towards Eye State Prediction Using EEG,” Proc. of the AIHLS, 2013.

[43] D. I. Sotelo, J. A. P. Benitez, and J. H. E. Hernandez, “Identification and classification of eyes movement using EEG signals,” in 2018 28th International Conference on Electronics, Communications and Computers, CONIELECOMP 2018, 2018, vol. 2018-January. doi: 10.1109/CONIELECOMP.2018.8327171.