Innovations

Unveiling Human Mental States from EEG Signals: A Machine Learning Perspective

Swati Chowdhuri
Institute of engineering & management Kolkata, India

Dibyendu Mal
Heritage Institute of Technology, Kolkata, India

Abstract—*The present study investigates the efficacy of EEG signal analysis in gauging human mental stress across distinct attention stages. The primary objective is to discern EEG-based markers and employ appropriate classification methodologies capable of delineating brainwave patterns based on their intensity or frequency, thus facilitating the identification of varying mental states crucial for enhancing human-machine interaction. The research endeavors to classify three distinct mental states—relaxation, neutrality, and concentration—utilizing an Emotiv headset equipped with four EEG sensors (TP9, AF7, AF8, and TP10). A dataset comprising sessions of one-minute duration for each attention category was compiled, encompassing data from five individuals. To refine and assess various techniques, an array of feature selection algorithms was employed initially on a pool of 2100 features. Subsequent application of diverse classifiers, including Bayesian Networks, Support Vector Machines, and Random Forests, enabled the reduction of the feature set to 44 critical factors, resulting in an overall classification accuracy of 87%.*

Keywords*— Electroencephalogram signal; brain machine interface; mental state classification; machine learning; Emotiv sensor*

I. Introduction

THEsignificance of maintaining good mental health for overall physical and mental well-being is underscored by recent studies, which highlight its potential to mitigate the risk of cardiovascular events such as heart attacks and strokes. Conversely, poor mental health can precipitate hazardous behaviors and undermine physical health. Individuals grappling with severe mental health issues often experience a diminished quality of life characterized by distress, a sense of powerlessness, low self-esteem, social alienation, reduced activity, and feelings of hopelessness. It is imperative to recognize that mental health, akin to physical health, is intrinsic to one's holistic wellness, encompassing the state of one's mind, emotions, and feelings. The autonomous detection of cognitive or affective mental states holds considerable utility across

diverse domains, including robotics, medicine, education, and neurology. Among the array of options available for facilitating human-machine interaction, surface brain activity signals, colloquially known as electroencephalography (EEG), emerge as a viable choice. EEG captures the brain's electrical activity through electrodes positioned on the scalp, providing insights into neuronal dynamics. The current study adopts a short-time windowing technique to identify local discriminative features within EEG signals, crucial for their effective classification into distinct mental states. The human brain, comprising billions of neurons, generates intricate electrical signals manifesting as nonlinear brainwave patterns. These signals vary in significance across different brain regions, with the cerebral hemispheres, brainstem, and cerebellum constituting the principal divisions. The brain's lobes—frontal, temporal, parietal, and occipital—serve distinct functions, such as memory processing, sensory perception, and concentration, each characterized by unique electrical activity patterns discernible through EEG analysis. Over the years, extensive research has elucidated EEG signals and their correlation with mental states, facilitated by advancements in machine learning techniques for signal processing and classification. Utilizing EEG data, researchers have made strides in identifying abnormal brain activity associated with conditions like stroke, aiding in early detection and rehabilitation efforts. Additionally, EEG data has facilitated advancements in brain-machine interfaces, enabling motor recovery post-stroke. Notably, EEG analysis holds promise in distinguishing seizures in epilepsy patients, including newborns. However, prior research predominantly relied on server or high-end microcomputer setups, prompting the current study to propose an edge-level analysis and model deployment framework using readily accessible devices like the NvidiaJetson Nano or Raspberry Pi4 microcomputers. The proposed system entails real-time EEG data collection and analysis using four electrodes, with a focus on classifying emotional states. The subsequent sections of this paper delineate past research in EEG processing (Section 2), the proposed methodology encompassing preprocessing, feature extraction, and training processes (Section 3), and concluding insights along with limitations and future directions (Section 4).

I. Related works

There are numerous works that have been carried out in the field of human attention [Kishan P, 17] concentration detection as well as emotion measurement and detection. Rahaman et al. [Wang, 20] give a direction on the assessment of cognitive function as a pivotal part of e-healthcare [Chowdhuri. 22] Here the various classes of non-invasive sensor and their uses has been analyzed in the context of various disease categories like sleep apnea, and brain tumor patients. The state-of-the-art feature extraction mechanism has evolved to produce remarkable accuracy in this case. On the other hand, give a direction on personalized cognition-driven intelligent wayfinding techniques [Wang, 19]. The CNN-based model has been built to create the wayfinding method with the help of EEG signal data. A test has been carried out on a simulated platform. One of the scopes of this research is that more categories of EEG can be considered for wayfinding analysis. In another work, a vigilance measurement with defined EEG-Sub-bands is performed [Arjun, 22]. The experimental results reveal that the highest correlation can be applied for vigilance detection. A hybrid kinematic EEG signal processing methodology with wavelet decomposition has been implemented in this work. An independent emotion recognition system using EEG signals is proposed [Cui H, 20], the core concept employed here is an attention-driven neural network. The work primarily emphasizes the subject of independent emotion recognition. The two-fold architecture of unsupervised LSTM and CNN has been applied to EEG datasets. Multiple promising challenges can be addressed that can be solved using such a state-of-the-art methodology [Aziz, 17]. On the other hand, propose a study to determine the concentration and non-concentration based on Fourier Transform [Houssein, 22]. A feature extraction method has been implemented using Hilbert Huang Transform (HHT) [Ang K.K., 10] using EMD and HT. A multichannel EEG-based emotion detection has been studied [Jordan J., 19]. The study has been made on numerous feature extraction methodologies applied to EEG signals. Based on that the review suggests some of the major challenges and future works in the field of EEG-based emotion detection. A multi-channel EEG-based time-frequency analysis for the recognition of human emotion is proposed where DWT methodology [Rahman, 21] has been used for the decomposition of the EEG signals [Li W., 12] with a db6 wavelet function [Wagh, 22]. SVM, kNN and Decision tree model has been used to classify the model where 71% accuracy has been given by decision tree. A low-power wearable hardware device amalgamated with a convolutional neural network has been proposed [Gonzalez,20]. The test has been carried out on 5 healthy people with standard visual stimuli and the efficiency has been recorded as 11 GOps/W. the discriminative feature extraction from the EEG signal is quite challenging. An attention-based convolutional recurrent neural network to identify significantly discriminative features is invented [Tao W., 20]. Their experimental results give start-of-the-art results in attention-based emotion detection. Table I summarizes the implementation of the technologies involved.

TABLE I
SUMMARY OF THE IMPLEMENTED TECHNOLOGIES INVOLVED WITH THE RELATED RESEARCH

Proposed work	**Testbed creation**	**Energy efficiency**	**Attention-based approach**	**Intelligent modeling**
[6]			✔	✔
[7]	✔		✔	✔
[8]	✔		✔	✔
[9]			✔	✔
[10]				✔
[11]	✔		✔	✔
[12]	✔		✔	✔
[13]			✔	✔
[14]	✔	✔	✔	✔

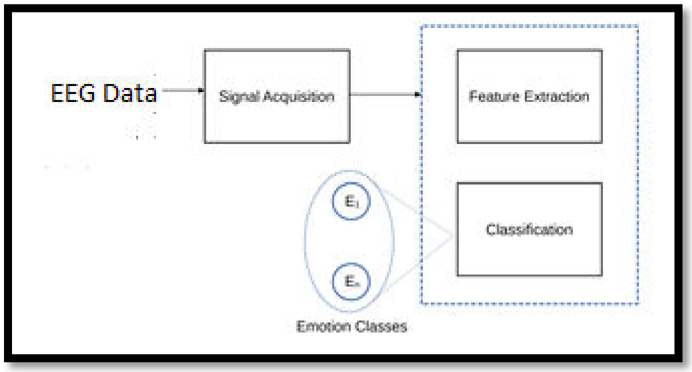

Fig. 1. Block diagram of signal acquisition and prediction using BCI Toolkit.

In this section we discuss the problem definition, preprocessing and overall working framework of our proposed edge level processing.

A. Problem Definition

The EEG signals captured from the scalp surface exhibit stochastic properties, which can be partitioned into frequency bands characterized by distinct amplitude and frequency attributes. These frequency bands, encompassing delta (d), theta (θ), alpha (α), beta (β), and gamma (γ) bands, convey valuable information regarding individuals' cognitive and emotional states. The amplitude of EEG signals typically falls within the range of 10 to 100 mV, with each frequency band associated with specific mental states such as deep sleep, relaxation, meditation, and various mental disorders.

In EEG data collection, headsets typically employ one to eight electrodes positioned on the scalp, adhering to the 10-20 standardization scheme. Let $P = [P_1\ P_2 \cdots P_k]$denote the set of connected probes, where EEG signals from each probe are amalgamated into a 2D vector P [n x k], with n representing the number of samples and k denoting the number of probes. The EEG dataset, comprising n samples of vectors akin to P, is represented as a matrix $\mathbf{F} = \begin{matrix} \mathbf{d_{11}} & \mathbf{d_{12}} & \mathbf{d_{1k}} \\ \mathbf{d_{21}} & \mathbf{d_{21}} & \mathbf{d_{2k}} \\ \mathbf{d_{n1}} & \mathbf{d_{n2}} & \mathbf{d_{nk}} \end{matrix}$. Subsequently, the correlation coefficient is computed across all features in matrix F, and a confusion matrix is constructed. To manage the increasing number of features while maintaining a constant sample size, feature engineering is employed to transform the input feature set into an [n x t] matrix

$$\mathbf{F_p} = \begin{matrix} \mathbf{f_{11}} & \mathbf{f_{12}} & \mathbf{f_{1t}} \\ \mathbf{d_{21}} & \mathbf{d_{22}} & \mathbf{d_{2t}} \\ \mathbf{d_{n1}} & \mathbf{d_{n2}} & \mathbf{d_{nt}} \end{matrix} \text{ where} t \geq k.$$

A classification model is then developed to map each of the n features to a corresponding label from the label set $\mathbf{Y} = [\mathbf{Y_1},\ \mathbf{Y_2}, \dots \mathbf{Y_i}]$, where i ranges from 1 to nt. The classification model M(X,Y) is formulated based on the evaluated features in the [n x t] matrix. The labels Y represent the set of emotions pertinent to the problem, with each Yi mapped to a subset e comprising distinct emotional states $\mathbf{e} = [\mathbf{e_1},\ \mathbf{e_2}, \dots, \mathbf{e_j}]$.

B. Preprocessing and Transformation

Brain-computer interface (BCI) applications face inherent challenges in feature extraction and classification of EEG signals due to their complex, nonlinear, non-stationary, and unpredictable nature. To address this complexity, short-time windowing techniques are commonly employed to capture transient signal characteristics. However, the non-stationarity of signals persists, influenced by factors such as eye blinking, changes in alertness, wakefulness, and mental state transitions. In this context, this subsection discusses the selection of features crucial for discriminating between different mental states, utilizing statistical methods including Shannon entropy, max-min features in temporal sequences, log-covariance, and time-frequency analysis via fast Fourier transform (FFT).

Each feature is computed based on the temporal distribution of signals over specific time spans, leveraging five signal frequencies (alpha, beta, gamma, delta, theta) to extract a comprehensive set of attributes. This results in the retrieval of 1656 feature values from the EEG signals. Classical statistical features are employed in conjunction with multiple features to capture patterns in time series and provide a compact representation of raw sensor data within defined time intervals.

The statistical features encompass various measures: (i) mean values, (ii) standard deviation, (iii) third and fourth-order statistical moments capturing skewness and kurtosis, (iv) autocorrelation of signals at each time window, and (v) maximum and minimum terms for each time frame, culminating in a total of 32 features per sample. Temporal features are further extracted by dividing one-second time windows into four batches of 0.25 seconds, evaluating mean, maximum, and minimum values for each batch, resulting in 18 features based on distance computation. Combining these with the previous statistical features yields a total of 30 features per window, with 150 temporal features computed per second for five signals. Log covariance analysis selects 144 values (12*12) for evaluation, encompassing upper and lower triangular elements. A total of 1200 features are evaluated for four samples. Additionally, FFT analysis is employed to analyze the spectrum of time-series data, resulting in the evaluation of 328 FFT features for each window. These diverse sets of features contribute to comprehensive feature representation essential for effective EEG signal classification and mental state discrimination in BCI applications.

C. Classification

Our study aims to classify different emotional states using preprocessed EEG data, represented as the feature matrix Fp. This task entails a multi-class classification problem, where each instance of EEG data may correspond to one of several potential emotional states. To achieve this, we have evaluated five prominent machine learning algorithms: Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, Neural Networks, and Support Vector Machine (SVM).

The Decision Tree model operates by iteratively constructing a series of decision rules derived from the features, creating a hierarchical tree structure. Each decision rule divides the data based on the value of a particular feature, leading to branches that represent distinct decision paths culminating in predicted class labels. The tree construction process continues recursively until predefined stopping conditions, such as maximum tree depth or minimum samples per leaf, are met. In contrast, the KNN model classifies a new sample's emotional state by identifying its k nearest neighbors in the feature space, utilizing distance metrics like Euclidean or Manhattan distance. The predicted label for the new sample is determined by the majority class among its nearest neighbors.

The Naive Bayes classifier assumes conditional independence among features given the class label, leveraging Bayes' theorem to compute the probability of each class label for a given set of features. Despite its scalability and ability to handle high-dimensional data, the Naive Bayes model's accuracy may be constrained by the independence assumption's potential deviation from reality. Neural Networks consist of interconnected layers of neurons that learn to transform input data into increasingly abstract representations. This model, specifically Deep Artificial Neural Networks in our experimentation, can capture complex feature interactions and non-linear relationships with sufficient data and layers. SVM seeks to find the hyperplane in the feature space that best separates different class labels, and for non-linearly separable data, it employs a kernel trick to map the data into a higher-dimensional space where it becomes linearly separable. Effective for high-dimensional data, SVM can discern both linear and non-linear relationships between features, contributing to its versatility in classification tasks.

II. Results and Discussion

In this section, we provide an overview of the experimental data utilized for model training, along with the corresponding results obtained from the conducted experiments. The experimental dataset comprises six features, with preprocessing conducted to replace any missing data with 0. During experimentation, the features

"Right AUX" and "Timestamps" were omitted due to their characterization as noise. Additionally, a normalization process was implemented as part of the data preprocessing pipeline to ensure uniform scaling across all features. Figure 2 illustrates the correlation coefficient matrix computed between the selected set of features, representing the received data from four different probes. The correlation matrix was derived using Pearson's correlation coefficient. Analysis of this matrix reveals several significant findings: Moderate positive correlations are observed between TP9 and AF7 (0.45), as well as between TP10 and AF8 (0.46). Additionally, weaker positive correlations are noted between TP9 and TP10 (0.58), and between AF7 and AF8 (0.35). Conversely, a weak negative correlation is observed between AF7 and AF8 (-0.17). Interpreting these correlations, it can be inferred that signals from TP9 and AF7, as well as TP10 and AF8, exhibit stronger correlations compared to signals from other electrode pairs. This observation suggests that these electrode pairs may capture activity from proximal brain regions that are functionally connected. Furthermore, the weaker positive correlation between TP9 and TP10 implies a degree of symmetry in neural activity between these regions, while a similar correlation between AF7 and AF8 suggests potential functional differentiation between these regions.

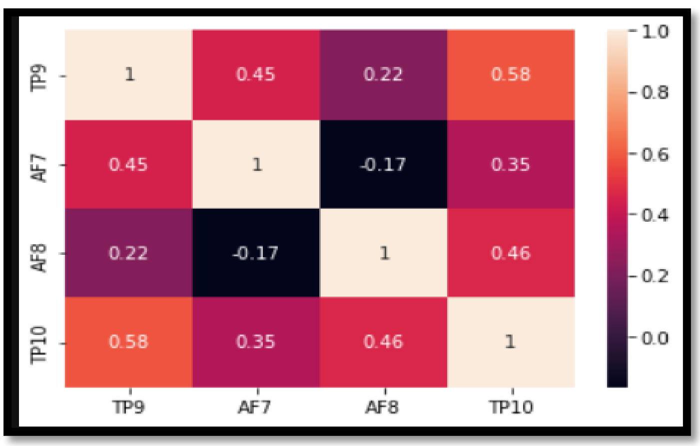

Fig. 2. Correlation Heat Map to showcase the correlation between features.

In this section, we explore the impact of clustering the data based on similarities in the feature space, employing the k-means clustering algorithm. By discerning patterns or groupings in the data that may not be readily discernible from raw data inspection, k-means clustering offers valuable insights into the data structure. The resulting visualization provides an overview of the data structure, elucidating relationships or dependencies among different features. Specifically, this visualization aids in identifying whether certain electrode pairs tend to cluster together, thereby shedding light on underlying neural activity patterns.

Moreover, the application of Principal Component Analysis (PCA) further simplifies the interpretation of complex data by reducing its dimensionality. Figure 3 presents the visualization of k-means clustering with 3 clusters following the application of the PCA process. The k-Means algorithm operates via a straightforward iterative approach, initially selecting data points as centroids and subsequently computing new centroids to minimize the net distance between all points in a cluster and its centroid. Given the unsupervised nature of the approach, the number of clusters is fixed at 3, aligning with the three emotion classes labeled in our dataset.

Observing the visualization, it becomes apparent that the clusters are well-defined, with minimal overlap between them. This suggests the effectiveness of the k-means clustering algorithm in partitioning the data into distinct groups based on underlying similarities in the feature space.

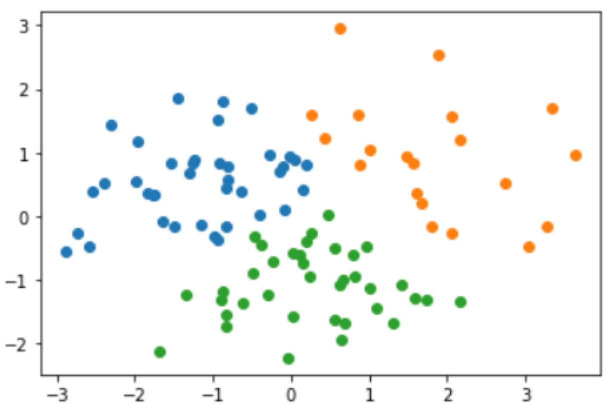

Fig. 3. Visualization of KMeans Clustering with n_cluster = 3 is shown below (data compressed using PCA).

An analysis on Silhouette's score is given in Table II. The Silhouette score is a measure of how well each data point in a clustering algorithm fits its assigned cluster relative to the other clusters. It is usually used to evaluate the quality of the clustering solution, that is with higher Silhouette scores indicating better clustering results. It can be observed that k-means clustering with 3 clusters has the highest Silhouette score of 0.40513, indicating that it may be the best clustering approach. This aligns well as the number of classes in our dataset is also three; i.e. the 3 emotion classes.

TABLE II
Summarized Results From Selected Clustering Algorithms

Algorithm	**Number of Clusters**	**Silhouette's Score**
KMeans Clustering	2	0.3478
KMeans Clustering	**3**	**0.40513**
Agglomerative Clustering	3	0.3718
Agglomerative Clustering	2	0.3382

TABLE III
Model Accuracy Comparison

Algorithm	**Test Accuracy**
KNN	94.8
Decision Tree (without Bagging)	90.5
Naive Bayes (without Bagging)	81.7
Decision Tree (with Bagging)	92.7
Naive Bayes (with Bagging)	82.3
Deep Artificial Neural Networks	**97.4**

TABLE IV
Precision and Recall For All The Classifiers Used

Algorithm	**Precision (class 0)**	**Precision (class 1)**	**Precision (class 2)**	**Recall (class 0)**	**Recall (class 1)**	**Recall (class 2)**
KNN	0.920	0.935	0.988	0.972	0.884	0.977
Decision Tree (withou	0.920	0.825	0.964	0.910	0.870	0.930

t Bagging)						
Naive Bayes (without Bagging)	0.741	0.790	0.922	0.899	0.541	0.965
Decision Tree (with Bagging)	0.970	0.872	0.933	0.921	0.890	0.965
Naive Bayes (with Bagging)	0.764	0.770	0.922	0.871	0.596	0.965
Deep Artificial Neural Networks	**0.977**	**0.946**	**0.994**	**0.966**	**0.966**	**0.988**

Table III and Table IV provide insights into the classification process conducted on our dataset, as depicted in Figure 4. Employing a range of classification models including KNN, Decision Tree, Naïve Bayes, Random Forest, Naïve Bayes with bagging, and Neural Network, we evaluated their respective performances. Notably, Table 3 outlines the parameters utilized for the Neural Network model.
The results indicate that Naïve Bayes classification exhibited the lowest accuracy at 82.05%, while the Deep Neural Network, with the specified parameter settings, achieved the highest accuracy exceeding 98%. These findings suggest the superiority of neural networks over traditional machine learning methods such as KNN, Naïve Bayes, and Decision Tree. Furthermore, we observed enhanced accuracy by employing ensemble techniques, particularly through majority voting with a combination of Naïve Bayes and Decision Tree models. The combined model achieved an accuracy of 93.95%, surpassing the individual model accuracies.

Additionally, the utilization of bagging methods notably improved the individual accuracies of Naïve Bayes and Decision Tree models. In the case of Neural Networks, optimization of hyperparameters within the Adam Optimizer framework and fine-tuning of the model architecture significantly influenced accuracy outcomes. Figure 5 presents the confusion matrix depicting the prediction outcomes of the Neural Network model, while Figure 6 illustrates the ROC curve visualization using the same model. The ROC curve offers insights into the model's ability to distinguish between true positive and false positive cases across various label values. The substantial area under the ROC curve for Neural Networks in a one-vs-all classification scenario signifies robust performance and accurate classification of the target class amidst other emotion classes.

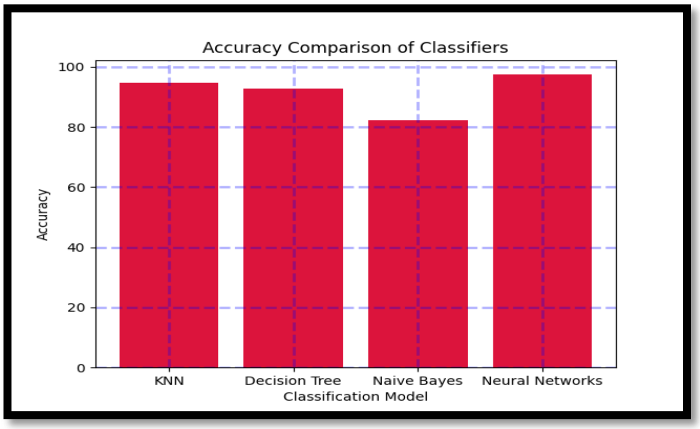

Fig. 4. Accuracy Comparison between the selected set of classifiers

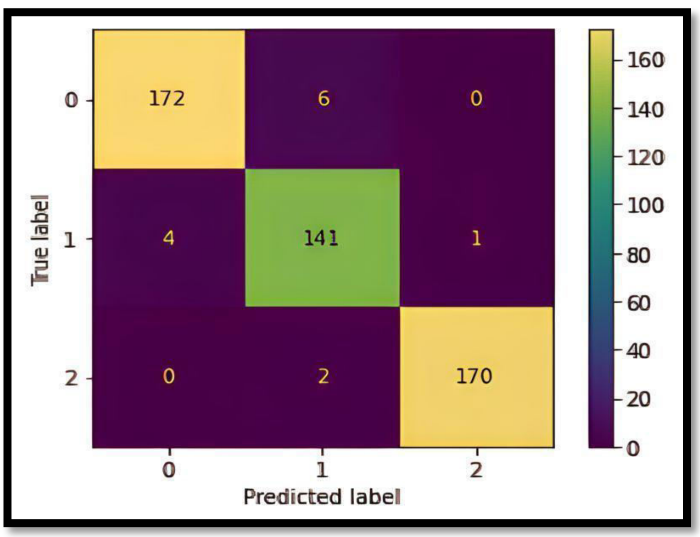

Fig. 5. Confusion Matrix of Neural Network on the dataset

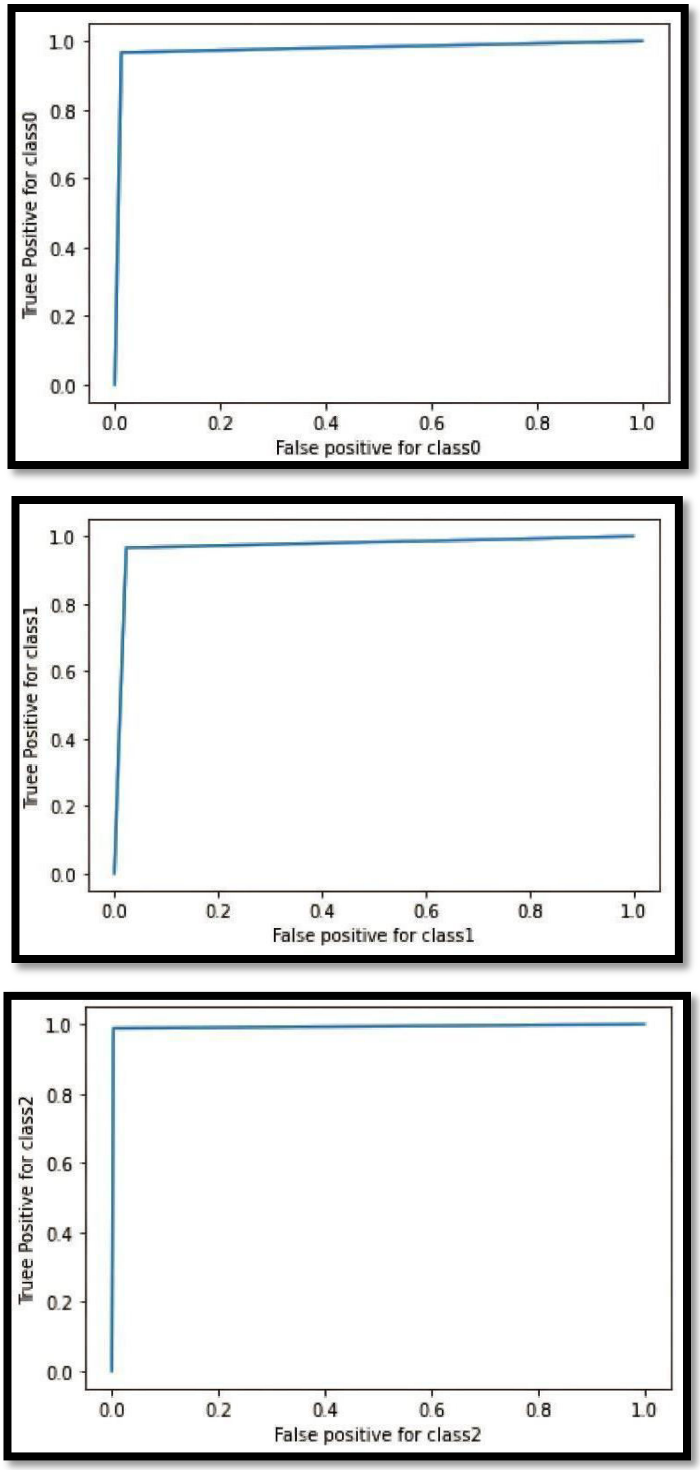

Fig. 6.ROC curve visualization of Neural Network

All the experimentation was conducted on an edge level device. The edge level analysis of the CPU utilization of model training is showcased in figure 7. It can be observed that a Neural network utilized the maximum memory during the training process.

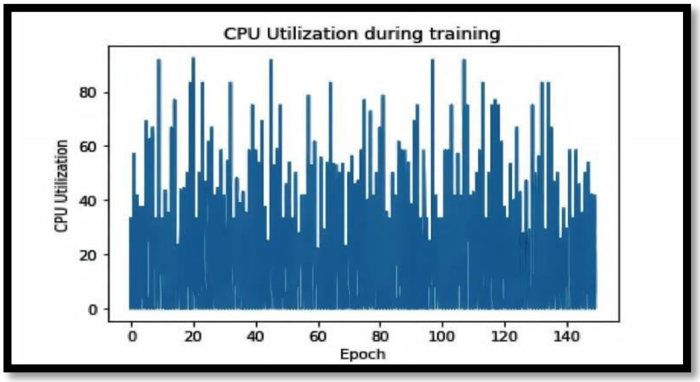

Fig. 7.CPU utilization during Training of Neural Network

TABLE V
Deep Ann Parameter Setting.

Parameters	**Setting**
Hidden Layers and Activation Function	Layer 1: 988 (Dense), Layer 2: 800(Dense), Layer 3: 800(LeakyReLU), Layer 4:400(Dense), Layer 5: 400(LeakyReLU), Layer 6:200(Dense), Layer 7: 200(LeakyReLU) Layer 8: 100(Dense), Layer 9: 100(LeakyReLU), Layer 9: 100(Flatten), Layer 10: 3(Dense)
Learning rate	0.0017
Epochs	50
Error Metric	Sparse Categorical Crossentropy

III. CONCLUSION

The research conducted has laid a promising groundwork for future investigations into EEG-based feature extraction and classification for recognizing mental states in human-machine interactions. Utilizing the Emotiv BCI headband with four EEG sensors (TP9, AF7, AF8, TP10), a dataset comprising five individuals across three distinct mental states - relaxation, neutrality, and concentration - was successfully established. From an extensive pool of over 2100 potential features, a rigorous selection process coupled with the application of multiple classifiers such as Bayesian Networks, Support Vector Machines, and Random Forests resulted in the identification of 44 critical factors. Despite the significant reduction in feature space, the streamlined model achieved an impressive overall recognition accuracy exceeding 87%. Furthermore, employing Deep Neural Network (DNN) techniques led to an outstanding accuracy rate of 97.8% on the collected dataset, corroborated by the validation through ROC curve estimation.

Additionally, the research showcased the feasibility of edge CPU utilization through experimentation on Raspberry Pi 4 devices, affirming the potential use of affordable consumer-grade EEG devices for practical applications in mental state recognition and human-machine interactions.
However, one limitation of the approach lies in categorizing mental states into only three categories - relaxation, neutrality, and concentration. Human cognitive states are multifaceted and extend beyond these predefined categories. Future endeavors aim to address this limitation by exploring further subdivisions of cognitive states to gain a more nuanced understanding of EEG signal variations. Overall, this research underscores the significance of EEG-based methodologies in advancing the field of mental state recognition, with implications for enhancing human-machine interaction systems and paving the way for future investigations into cognitive state categorization.

REFERENCES

1. *Diego R. Faria (2019) Mental Emotional Sentiment Classification with an EEG-based Brain-Machine Interface, The International Conference on Digital Image & Signal Processing (DISP'19), Oxford University, UK.*
2. *Kishan P, Harit S, Malcolm D, and Dvijesh S (2017) Evaluating NeuroSky's Single-Channel EEG Sensor for Drowsiness Detection, International Conference on Human-Computer Interaction.*
3. *Li W., Jaramillo C. and Li Y. (2012) Development of mind control system for humanoid robot through a brain computer interface, International Conference on Intelligent System Design and Engineering Application, 679-682 IEEE.*
4. *Ang K.K., Guan C., Chua K.S.G., Ang B.T., Kuah, C. and Wang C. (2010) Clinical study of neurorehabilitation in stroke using EEG-based motor*

imagery brain computer interface with robotic feedback. Annual International Conference of the IEEE, 5549-5552.

5. *Chowdhuri S. and Mal D. (2022) Concentration Level Detection for Left/Right Brain Dominance using Electroencephalogram Signal, Romanian Journal of Information Technology and Automatic Control, Vol. 32, No. 3, 109-118.*
6. *Rahman, Mustafizur M, Sarkar A, MdHossain A, MdHossain S and Mohammad Ali M (2021) Recognition of human emotions using EEG signals: A review, Computers in Biology and Medicine 136: 104696.*
7. *Wang Y, Yangming S, Jing D, Yingzi L and Qi W (2020) A CNN-based personalized system for attention detection in wayfinding tasks, Advanced Engineering Informatics 46:101180.*
8. *Wang D., Heng L. and Jiayu C (2019) Detecting and measuring construction workers' vigilance through hybrid kinematic-EEG signals, Automation in Construction 100: 11-23.*
9. *Arjun R., Aniket S. and Mahesh R. (2022) Subject independent emotion recognition using EEG signals employing attention driven neural networks, Biomedical Signal Processing and Control 75: 103547.*
10. *Cui H, Aiping L, Xu Z, Xiang C, Kongqiao W and Xun C (2020) EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network, Knowledge-Based Systems 205: 106243.*
11. *Aziz, Fathrul A, Ibrahim S, Noor Akhmad S and Yasue M (2017) Classification of Human Concentration in EEG Signals using Hilbert Huang Transform, International Journal of Simulation--Systems, Science & Technology 18(1).*
12. *Houssein, Essam H, AsmaaHammad and Abdelmgeid A. Ali (2022) Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review, Neural Computing and Applications 34(15): 12527-12557.*
13. *Wagh, Kalyani P and K. Vasanth (2022) Performance evaluation of multi-channel electroencephalogram signal (EEG) based time frequency analysis for human emotion recognition Biomedical Signal Processing and Control 78: 103966.*
14. *Gonzalez, Hector A., Shahzad M, Jerald Y and Ibrahim M (2020) BioCNN: A hardware inference engine for EEG-based emotion detection, IEEE Access 8: 140896-140914.*
15. *Tao W., Chang L., Rencheng S., Juan C., Yu L. and Xun C (2020) EEG-based emotion recognition via channel-wise attention and self-attention, IEEE Transactions on Affective Computing.*