Machine Learning Applications in Engineering Education and Management

*ISSN: **Applied***

*Volume **02** Issue **01 Year 2022** Page **11:19***

EEG Signal Classification with Machine Learning model using PCA feature selection with Modified Hilbert transformation for Brain-Computer Interface Application

[1]Prof. Dipannita Mondal, [2]Mrs. Sheetal S. Patil

[1]Dr. D.Y. Patil College of Engineering and Innovation Talegaon Varale Pune

[2]BVDUCOE Pune

[1]mondal.dipannita26@gmail.com

[2]sspatil@bvucoep.edu.in

Article History

Received: 20 January 2022
Revised: 10 April 2022
Accepted: 13 May 2022

Abstract

Brain Computer Interfaces (BCI's) computes the communication and control channels that rely on the output channel normal brain for the peripheral nerves. The present BCI advancement incorporates the technologies for augmentative communication between the users for the conventional technique for the effective control of muscles. The application of the Electroencephalography (EEG) activity estimates the brain activity for effective communication technology. This paper presented a Principal Component Analysis (PCA) integrated with the Modified Hilbert Transformation defined as the PCAHT. The developed model uses the PCA model for the estimation of the feature selection in the EEG signal for processing. The EEG signals inputs are pre-processed and evaluated for the features within the signal for processing. The proposed PCAHT model incorporates the modified Hilbert Transformation model for signal conversion and is applied over the machine learning model. The deployed PCAHT model perform the classification of the EEG signal those are normal and abnormal activity in the brain signals. The simulation analysis expressed that proposed PCAHT model achieves a higher accuracy value of 98%, recall value is measured as 98% and precision value is measured as the 98%. The comparative analysis observed that the proposed PCAHT model achieves the ~4% improved performance than the conventional classifier models.

Keywords: Brain-Computer Interface (BCI), Electroencephalography (EEG), Machine Learning, Hilbert Transformation, Principal Component Analysis (PCA)

1. Introduction

Brain-Computer Interface (BCI) is a class of communication system enabled to control the application of thoughts through the computer [1]. Generally, BCI comprises the neural input signal with the fundamental input signal in which the signal is processed, feature extraction for the classification, and conversion of the recognized signal in the signal commands to achieve desired actions [2]. The BCI comprises the activity of the brain through imaging technology for the detection of brain pattern characteristics with user control digital signal processing algorithm [3]. In BCI process electro-physiological generated signals are expressed with muscle contraction with the conventional communication technique. However, those BCI process requires the activity of brain pattern for the generation or control in the computer system. The electroencephalography (EEG) based BCI process comprises of the electrical signal pattern based on the human scalp [4]. The BCI-based approach comprises of the pattern for the menta task generation in the pattern of the brain for the identification and control through imaging-based hand movement.

Based on the evaluation of the mental task the EEG signal responses are computed for the translation of the user code book to ensure the BCI system for the EEG activity [5]. The BCI approach involved extensive training for the EEG signal with the self-regulatory activity of the brain stated as the mu rhythm [6]. The generated signal is distinguishable for the tandem brain activity to control the command for the recognition of brain activity in the BCI system through evaluation of the brain activity for the extraction of the brain signal activity through differentiable specific control activity [7]. The BCI activity is distinguishable for the generation of the control command in activity for the extraction of features and related activity. The BCI process comprises of the requirements for the control or communication of the attribute values of the EEG signal characteristics. With efficient BCI operation, the control activity is generated for the feature signal based on the intention of users [8].

Recently, BCI is an efficient technology for the restore, substituting, or augment of patient behaviour pattern based on neurological injuries such as traumatic brain injury, spinal cord, stoke and chronic gait function in the foot-drop [9]. The patients are provided with the BCI mounted on the wheel-chair, prosthetic limbs, robot mobile control and control of computer mouse for communication and environment control [10]. Some of the applications of the BCI process are lie-detection, monitoring, neurofeedback, neuromarketing and eye tracking. With the invasive BCI process the grey matter are directly implemented in the brain for the examination of neurosurgery [11]. The implementation of the invasive devices the BCI devices generates the high-quality signals whereas non-invasive BCI does not utilized for the surgery rather the signal is gathered from the surface of brain itself. Firstly, invasive BCI the brain interacts directly through the penetration in the brain cortex [12]. The resulted surgery for the brain signal is achieved for the specified time period. However, the non-invasive BCI techniques is highly expensive. Secondly, through non-invasive BCI the electrodes are placed in the brain scalp for the recording of the EEG signal [13]. The non-invasive process is less expensive compared with the invasive BCI system. The BCI non-invasive system records the EEG signal and comprises of the different types of P300 system and sensor motor system.

The objects are extracted and characterized based on the measured identical values those are in similar category with consideration of the different objects [14]. The features in the BCI system are recorded based on the EEG values based on representatives. The feature characteristics are identified based on described EEG activity those are highly discriminated for the control classes [15]. The characterization of the EEG events is represented in time-domain process with Slow Cortical Potentials (SCPs), Movement related Potentials (MRPs), Visual Evoked Potentials (VEPs). The EEG signal features are extracted based on the analysis of the EEG waveform morphological characteristics through descriptive parameter amplitudes [16]. The EEG signal background activity are computed based on the low Signal to Noise Ratio (SNR) for the command-based EEG pattern, and feature extraction in time domain those are leaned with the parametric modelling methods for the computation of the EEG activity in time-series components.

In this paper proposed a PCAHT model for the classification in the EEG signal for the normal and abnormal activity in the signal. The proposed PCAHT model computes the features in the EEG signal for classification. Through the modified Hilbert Transformation, the signal performance is improved for the classification. The proposed PCAHT model achieves the higher accuracy value of 98% compared with the conventional classifier model. This paper is organized as follows: The related works associated with the machine learning signal is presented in section 2. The section 3 provides the proposed PCAHT model for the classification is presented with the results and discussion in the section 4. The overall derived conclusion for the proposed PCAHT model is presented in section 5.

2. Related works

In [16] presented a technique to integrates the EEG pattern related to the separability of the mental tasks 2-5. The BCI system comprises of the information transfer system for the subsets. The information transfer rates are ranges from 0.42 – 0.81 bits. Trial for the differentiation of the BCI mental tasks. The examination observed that integrated model exhibits the significant performance for the task three.

In [17] examined the brain computed for the interface system of computer with the analyses of the electrical activity of the brain. Through the analysis the least idea is computed for the generated subject intention for the output command control for the appropriate devices. Through the EEG based interface system the advantages and limitations are examined based on the consideration of the different aspect based on consideration of ranging EEG input, processing stage and output signal. The performance of the BCI model is effective for the EEG based signal processing system.

In [18] evaluated the different type of the EEG signal based on the class dependent ERD features based on the mu-rhythm for the motor imagery in the BCI sessions. The developed model is based on the localization of the Fourier Co-efficient with the single trial through Sparse Basis Field Expansions (SFLEX). The experimental analysis expressed that sensorimotor cortex expresses the BCI control signal through the neurophysiological feature process. With the sensorimotor analysis the extended SFLEX measurement are evaluated based on bias frequency with mu-band localized coherent.

In [19] examined and demonstrated the information transfer rate at the higher level that can be achieved through Readiness Potential (RP) for the laterality prediction for the hand movement of a person in right and left sides. The evaluation of RP similarity expressed the phantom movement in the amputees with the decreased signal strength for the loss limb in the longer side. Through the process of complementary approach with the discriminate imagined movement. The evaluation expressed that healthy subject are identified as 6 with minimal experience with the BCI control features of the 3 subjects are observed and the information transfer rate is observed as 35 bits per minute (bpm) for the 24 subjects are measured as 15bpm for the BCI control. The analysis of the results expresses that BCI based EEG system achieves the untrained subjects for the peripheral nervous system independent activity and not evoked the potential compared with the operation of the well-trained subject in the BCI system.

In [20] evaluated the classification model for the Mother Wavelets the features are extracted based on consideration of different datasets with 12 Mother Wavelets. The BCI model is classified based on the consideration of three classification algorithms such as a k-nearest neighbour, Support Vector Machine (SVM), and Linear Discriminant Analysis. The experimental analysis exhibits the Daubechies and Shannon for the effective wavelet for the extraction of the discriminative features are evaluated based on the BCI signals.

In [21] proposed a wavelet transform and Hilbert-Huang transforms (HHT) for the extraction of the features in the EEG signal for the efficient method performance characteristics. The HHT exhibits accurate distribution in the EEG signal based on the time and frequency component. HHT-based performance exhibits the self-adaptiveness for the signal data processing based on the local and instantaneous EEG signal.

In [22] developed an EEG-based feature extraction signal through a motor imagery process with the HHT transformation. The BCI features are captured and processed with the EEG signal based on the consideration of the limb for the conversion of signal in the series for the control signals. The features are extracted based on the non-linear and non-stationary characteristics of the data captured through the EEG signal. The developed model uses the GA-based selection model in the frequency domain. The experimental analysis expressed that the proposed HHT-based GA model exhibits higher accuracy compared with the conventional feature extraction methods.

3. PCA Hilbert transformation in Machine Learning

To classify the EEG signal based on the brain signal activity BCI model is implemented using the PCA based Hilbert Transform based on the effective signal processing model. The proposed model PCAHT comprises of the different features such as analytical signal, bandpass sampling, network minimal phase and spectral analysis.

The PCAHT model function is presented in equation (1)

$$s(t) = L\{e(t)\} = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{e(t)}{t-\tau}d(\tau) \tag{1}$$

With the PCAHT based Fourier identities the Hilbert transform $s(t)$ in frequency domain are presented in equation (2)

$$s(t) \leftrightarrow S(f) = -j\, sgn(f)E(f) \tag{2}$$

The feature Fourier transform pair ⟺ stated as in equation (3)

$$sgn(f) = \begin{cases} 1 & f > 0 \\ 0 & f = 0 \\ -1 & f < 0 \end{cases} \tag{3}$$

The PCAHT model the Chebyshev response are computed based on the frequency ripple response. The ripple depth is estimated based on the range of 0.1dB to 3dB mode with the maximal flat value of Butterworth filter. The Chebyshev filter ripple peak-to-peak are determined using the equation (4)

$$d = 10\log(1 + \varepsilon^2) \tag{4}$$

The above equation (4) is solved based on ε defined in equation (5)

$$\varepsilon = \sqrt{10^{\,dB/10} - 1} \tag{5}$$

Through Chebyshev transfer function the parameter h is computed and measured using the equation (6)

$$h = tanh\left(\frac{1}{n} sinh^{-1}\frac{1}{\varepsilon}\right) \tag{6}$$

In above equation (6) n denoted as the low pass filter component. PCAHT computes the higher data dimensions through location findings based on the statistics for the big datasets for the analyse of the relationship between the individual point. The PCAHT model compute the dimensionality reduction based on the original dataset for the identification of the data pattern for the computation of difference similarities. The process involved in proposed PCAHT flow chart is presented in figure 1.

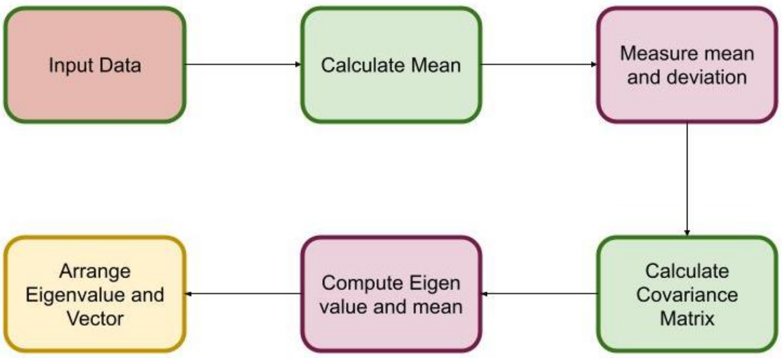


Figure 1: Flow of PCAHT

The PCAHT model uses the ranking based features for the decision tree process to achieve the effective classification performance. The PCAHT model uses the feature selection based on the filter approach for the selected features for the computation of the class information. Ideally, those features exhibit the discriminative model information gain in singe class based on the measured entropy. The entropy features are minimized based on the consideration of the specific attributes for the estimation of the information gain in the attributes data sets. Every feature information gain is determined based on the selection of features or not. With the threshold features the higher information gain are computed based on threshold for the selected features.

Consider $\boldsymbol{S}$ for the set of instances n set $\boldsymbol{C}$ is defined as the k classes. $\boldsymbol{Pr_B}$ denotes the fraction of the class $\boldsymbol{S}$. The information gain for the membership class function is defined as in equation (7)

$$\boldsymbol{Pr_B(X) = \sum_{i=1}^{k} S(C_i, X) \times log\big(P(C_i, X)\big)} \quad \textbf{(7)}$$

Based on the set of attributes $\boldsymbol{X}$ distinct values are computed based on the information achieved from the decision tree and weighted sum of information gain subset as $\boldsymbol{X_i}$. The label instances attributes are presented in equation (8)

$$\boldsymbol{Pr_B(X)} = -\sum_{i=1}^{v} \frac{|X_i|}{|X|} \times \boldsymbol{Pr(X_i)} \quad (8)$$

The variation in the information gain is measured using the equation represented in equation (9) denoted as $Pr\ (X)$ and $\boldsymbol{Pr_B(X)}$

$$Gain\ (P) = Pr\ (X) - \boldsymbol{Pr_B(X)} \quad (9)$$

Through the information gain the higher value of likelihood estimation is computed for the target classes in the EEG signal. The classification in the EEG signal are applied and processed with the Multi-Layer Perceptron Algorithm steps defined as follows:

1. Initialize the network and compute the weights random number ranges between -1 and +1.
2. Through training pattern achieve the output
3. Compare the network output and targeted output
4. Propagate the backward error values in the network.

The PCAHT weights in the correct layer is computed using the equation (10)

$$b_{io} = b_{io} + ({}^{\eta}\delta_o J_n) \quad (10)$$

The connected hidden layer b with the learning rate of δ_o and the hidden later is computed as in equation (11)

$$\delta_o = J_n(1 - J_n)(t_n - J_n) \quad (11)$$

The network transfer function for the proposed PCAHT for the adjusted learning rate for the establishment of relationship between neuron are computed using the sigmoidal function defined in equation (12)

$$S(t) = \frac{1}{1 + e^{-s}} \quad (12)$$

The tanh function in the equation are represented as in equation (13)

$$\tanh = \frac{e^{2s} - 1}{1 + e^{2s}} \quad (13)$$

The proposed PCAHT model for the transfer function model for the different learning rate are presented in table 1.

Table 1: Function in PCAHT

Function	Rate of Learning	Momentum	Accuracy (%)	RMSE
Sigmoidal	0.2	0.1	93.89	0.6841
tanh	0.1	0.1	90.78	0.8483
Sigmoidal	0.3	0.3	91.98	0.5942
Sigmoidal	0.2	0.2	89.48	0.6742
tanh	0.1	0.1	88.74	0.6839
Sigmoidal	0.2	0.3	89.63	0.3985
Sigmoidal	0.2	0.3	91.48	0.4679
tanh	0.1	0.3	92.75	0.5943

The overall process involved in proposed PCAHT model is shown in figure 2.

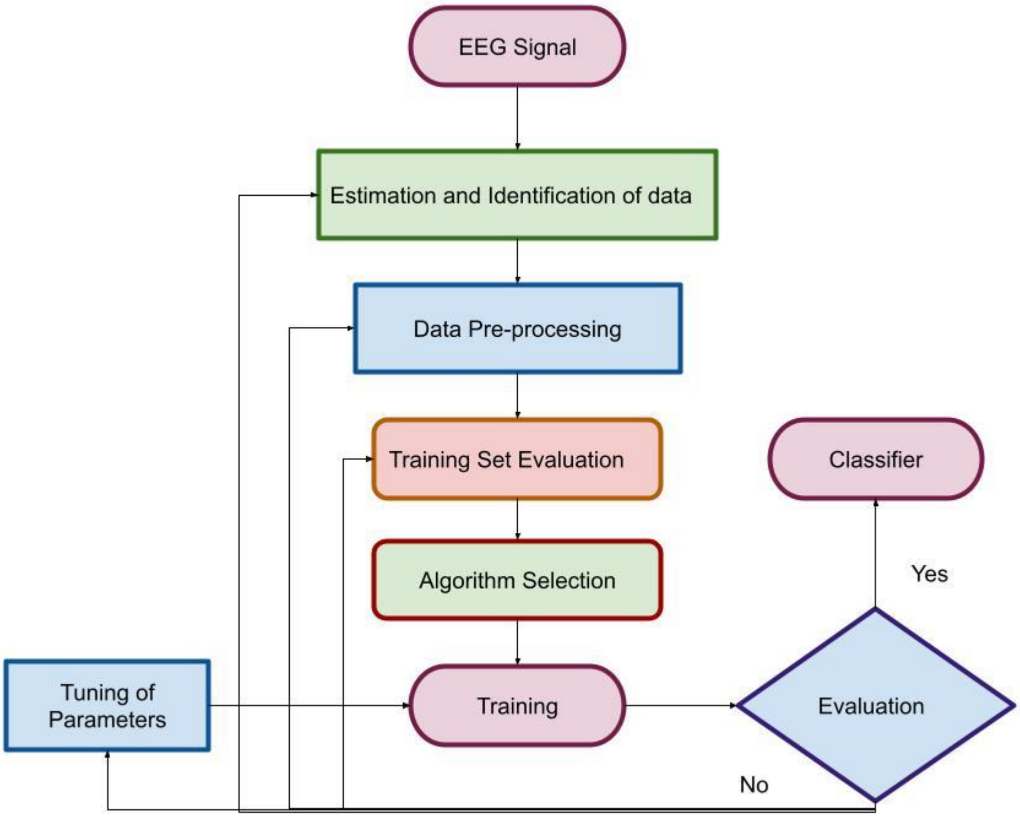


Figure 2: Flow Chart of PCAHT

The proposed PCAHT model comprises of the pre-processing of the EEG signal with the estimation and the identification of the features in the signal. Initially, the EEG signal is pre-processed for the evaluation for the training and testing of the EEG signal. With the algorithm selection model, the PCAHT model the classification is performed for the EEG signal. With the incorporated PCAHT model the signal is classified for the normal and abnormal signal in the EEG signal.

4. Results and Discussion

The proposed PCAHT exhibits the higher accuracy for compared with the existing classifier model. The performance of the proposed PCAHT model is comparatively analysed with the PCA and HT model. In table 2 the performance of the proposed PCAHT model accuracy is presented.

Table 2: Comparison of Performance Metrices

	Accuracy		**Recall**		**Precision**	
Classifier	**PCA**	**HT**	**PCA**	**HT**	**PCA**	**HT**
Linear Kernel	65.47	69.84	68.72	70.37	73.45	77.86
Poly Kernel	68.73	71.43	71.64	74.72	80.35	82.34
RBF Kernel	70.57	73.47	78.94	80.34	84.62	88.93
Naive Bayes	78.93	79.59	85.62	87.76	86.73	89.93
KNN	83.59	88.67	89.84	90.43	91.34	93.42
PCAHT	96.78	98.63	97.54	98.93	97.65	98.73

The EEG signal is classified based on proposed PCAHT model compared with the existing classifier model such as linear kernel, poly, RBF, Naïve Bayes and KNN model. The computation of the recall value expressed that proposed PCAHT model achieves the higher recall value with the HT with 98.93%. The existing linear, poly, RBF, Naïve Bayes and KNN achieves the recall value of 70.37, 74.72,80.34,87.76 and 90.43 respectively. The comparative analysis of the precision expressed that proposed PCAHT model achieves the precision value of 98.73% those are significantly higher than the conventional classifier models. Similarly, in table 4 the comparative analysis of the RMSE for the different classifier are presented.

Table 3: Comparison of RMSE

Classifier	PCA	HT
Linear Kernel	0.21	0.34
Poly Kernel	0.23	0.28
RBF Kernel	0.18	0.14
Naive Bayes	0.13	0.16
KNN	0.11	0.12
PCAHT	0.08	0.09

In figure 3 the comparative analysis of the RMSE value for the different classifier with the proposed PCAHT mode is provided. The comparative examination expressed that proposed PCAHT model achieves the minimal RMSE compared with the existing classifier such as linear, poly, RBF, Naïve Bayes and KNN.

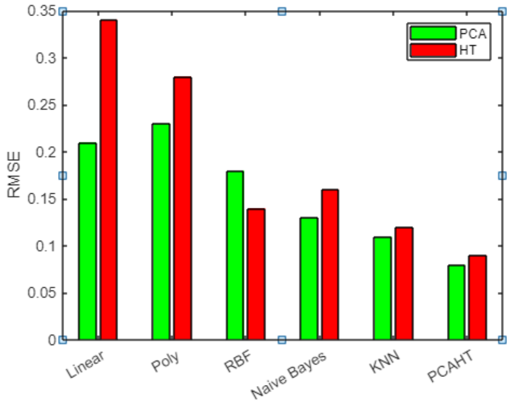
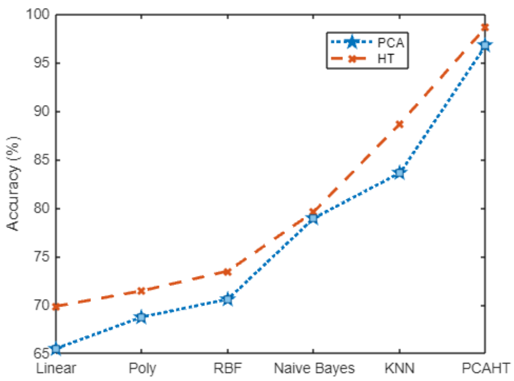
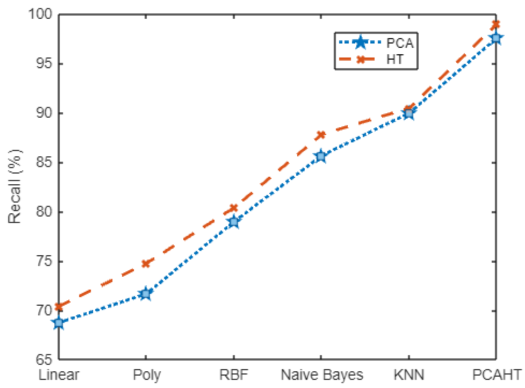


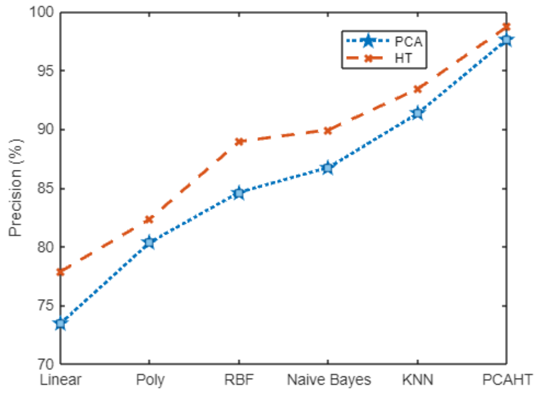
Figure 3: Comparison of RMSE

The comparative analysis of the performance of the proposed PCAHT model with the existing classifier achieves the higher accuracy, recall and precision value as illustrated in figure 4.





(a) **(b)**



(c)

Figure 4: Comparison of (a) Accuracy (b) Recall (c) Precision

The proposed PCAHT model performance is evaluated for the consideration of the different parameters such as Accuracy, recall and precision value. The proposed PCAHT model achieves the higher value of 98% which is significantly higher than the existing classifier. The table 3 provides the higher accuracy compared with the other classifiers.

Table 3: Comparison of Accuracy

Classifiers	**Accuracy (%)**
CSP	85.67
Non-Normalized Transductive Learning Model	88.54
TSLDA	96.45
Proposed PCAHT	98.63

The proposed PCAHT model achieves the higher accuracy value of the 98.63% while the conventional classifier model such as CSP model achieves the accuracy of 85.67%, Transductive Learning model accuracy is measured as the 88.54% and TSLDA model achieves the accuracy value of 96.45%. The comparative analysis expressed that proposed PCAHT model achieves the higher performance compared with the existing classifier model.

5. Conclusion

This paper presented a PCAHT model for the extraction and classification of the features in the EEG signal for the time series conversion with feature reduction. The PCAHT model examine the EEG signal those are classified for the normal and abnormal signal in the EEG signal. The EEG signal are pre-processed and features are extracted and with the implementation of the machine learning model the EEG signal is classified. The developed PCAHT model achieves the effective classification of the EEG signal for the accuracy of 98% which is significantly higher for the Kernel, Poly, RBF, Naïve Bayes and KNN classifier. The proposed PCAHT model achieves the higher accuracy of 98%, recall value of 98% and precision value of 98% which is ~4% higher than the conventional classifier models.

References

[1] Qureshi, M. B., Afzaal, M., Qureshi, M. S., & Fayaz, M. (2021). Machine learning-based EEG signals classification model for epileptic seizure detection. *Multimedia Tools and Applications*, *80*(12), 17849-17877.

[2] Kokate, P., Pancholi, S., & Joshi, A. M. (2021). Classification of upper arm movements from EEG signals using machine learning with ica analysis. *arXiv preprint arXiv:2107.08514*.

[3] Popescu, A. B., Taca, I. A., Nita, C. I., Vizitiu, A., Demeter, R., Suciu, C., & Itu, L. M. (2021). Privacy preserving classification of eeg data using machine learning and homomorphic encryption. *Applied Sciences*, *11*(16), 7360.

[4] Vivaldi, N., Caiola, M., Solarana, K., & Ye, M. (2021). Evaluating performance of eeg data-driven machine learning for traumatic brain injury classification. *IEEE Transactions on Biomedical Engineering*, *68*(11), 3205-3216.

[5] Rabby, M. K. M., Islam, A. K., Belkasim, S., & Bikdash, M. U. (2021, April). Epileptic seizures classification in EEG using PCA based genetic algorithm through machine learning. In *Proceedings of the 2021 ACM southeast conference* (pp. 17-24).

[6] Ramírez-Arias, F. J., García-Guerrero, E. E., Tlelo-Cuautle, E., Colores-Vargas, J. M., García-Canseco, E., López-Bonilla, O. R., ... & Inzunza-González, E. (2022). Evaluation of Machine Learning Algorithms for Classification of EEG Signals. *Technologies*, *10*(4), 79.

[7] Satapathy, S. K., & Loganathan, D. (2021). A study of human sleep stage classification based on dual channels of EEG signal using machine learning techniques. *SN Computer Science*, *2*(3), 1-16.

[8] Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C. K., Li, X., & Guan, C. (2021). An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 809-818.

[9] Rahman, A., Chowdhury, M. E., Khandakar, A., Kiranyaz, S., Zaman, K. S., Reaz, M. B. I., ... & Kadir, M. A. (2021). Multimodal EEG and keystroke dynamics based biometric system using machine learning algorithms. *IEEE Access*, *9*, 94625-94643.

[10] Mohutsiwa, L. O., & Jamisola, R. S. (2021, May). EEG-Based human emotion classification using combined computational techniques for feature extraction and selection in six machine learning models. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1095-1102). IEEE.

[11] Gonçales, L. J., Farias, K., Kupssinskü, L. S., & Segalotto, M. (2022). An empirical evaluation of machine learning techniques to classify code comprehension based on EEG data. *Expert Systems with Applications*, *203*, 117354.

[12] Srinath, R., & Gayathri, R. (2021). Detection and classification of electroencephalogram signals for epilepsy disease using machine learning methods. *international Journal of imaging Systems and technology*, *31*(2), 729-740.

[13] Sethy, P. K., Panigrahi, M., Vijayakumar, K., & Behera, S. K. (2021). Machine learning based classification of EEG signal for detection of child epileptic seizure without snipping. *International Journal of Speech Technology*, 1-12.

[14] Dadebayev, D., Goh, W. W., & Tan, E. X. (2021). EEG-based emotion recognition: Review of commercial EEG devices and machine learning techniques. *Journal of King Saud University-Computer and Information Sciences*.

[15] Saeidi, M., Karwowski, W., Farahani, F. V., Fiok, K., Taiar, R., Hancock, P. A., & Al-Juaid, A. (2021). Neural decoding of EEG signals with machine learning: a systematic review. *Brain Sciences*, *11*(11), 1525.

[16] Qaisar, S. M., & Hussain, S. F. (2021). Effective epileptic seizure detection by using level-crossing EEG sampling sub-bands statistical features selection and machine learning for mobile healthcare. *Computer Methods and Programs in Biomedicine*, *203*, 106034.

[17] Khan, S., Khan, A., Hameed, N., Taufiq, M. A., & Riaz, S. (2021). Localizing Epileptogenic Zone from High Density EEG Data Using Machine Learning. *Int. J. online Biomed. Eng*, *17*(6), 73-86.

[18] Majoros, T., & Oniga, S. (2022). Overview of the EEG-Based Classification of Motor Imagery Activities Using Machine Learning Methods and Inference Acceleration with FPGA-Based Cards. *Electronics*, *11*(15), 2293.

[19] Satapathy, S. K., Bhoi, A. K., Loganathan, D., Khandelwal, B., & Barsocchi, P. (2021). Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal. *Biomedical Signal Processing and Control*, *69*, 102898.

[20] Rajashekhar, U., & Harish, H. M. (2022). Automatic diseases detection and classification of EEG signal with pervasive computing using machine learning. *International Journal of Pervasive Computing and Communications*.

[21] Nandi, A., Xhafa, F., Subirats, L., & Fort, S. (2021). Real-time emotion classification using EEG data stream in e-learning contexts. *Sensors*, *21*(5), 1589.