ISSN: 2584-0495 Vol. 2, Issue 3, pp. 640-650

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

Statistical Analysis and Classification of EEG Signal

Shweta Suresh Pathak, Nilkanth H. Kulkarni

Cite as: Shweta, S. P., & Kulkarni, N. H. (2024). Statistical Analysis and Classification of EEG Signal. International Journal of Microsystems and IoT, 2(3), 640-650. https://doi.org/10.5281/zenodo.10884641

Published online: 11 March 2024

DOI: https://doi.org/10.5281/zenodo.10884641

International Journal of Microsystems and IoT
Vol. 2, Issue.3, pp. 640-650; DOI: https://doi.org/10.5281/zenodo.10884641

Statistical Analysis and Classification of EEG Signal

Shweta Suresh Pathak[1], Nilkanth H. Kulkarni[2]

[1]Department of Electronics & Telecommunications Engg., Shri Guru Gobind Singhji Institute of Engineering and Technology, Vishnupuri, Nanded, India.
[2]Department of Civil Engineering, Shri Guru Gobind Singhji Institute of Engineering and Technology, Vishnupuri, Nanded, India.

ABSTRACT

Recently vast research on EEG signal is being explored in the area of machine learning. In this research, online EEGBCI data set is taken to analyze EEG signal. XDAWN denoising technique is implemented for signal pre-processing to improve signal to noise ratio (SNR) of EEG data. Then, filtering is done to separate different brain waves of EEG signal. After filtering, statistical tests are performed to extract most significant features present in the signal also, hypothesis tests have been performed to find statistically significant features. By performing the T-test, prominent features are extracted. Further, machine learning algorithms are used. Highest classification accuracy of 98.88% is obtained using KNN (K=3) classifier. The recent study is compared to the previous research on the same data set in which highest accuracy found was 97.77%. So, it can be concluded that, this research has improved methodology as compared to the reference research.

KEYWORDS
ANOVA classification
denoising
EEG
feature extraction
filtering
FIR

1. INTRODUCTION

This is an experimental and empirical type of research. It has specific to general, bottom-up approach. Inductive method is being used. This paper consists of various statistical tests performed for feature extraction and classification techniques in machine learning used for EEG signals. In this paper, online EEGBCI data is used [1]. The Python programming is implemented to perform statistical analysis and classification. This paper involves denoising, FIR filtering, features extraction using statistical analysis and classification techniques applied on EEG signals which are discussed in the methodology section. Highest accuracy obtained using KNN classification, is discussed in results and discussion section and it is compared with Classification of Electroencephalogram (EEG) Signals Using Linear Discriminant Analysis [2].

2. LITERATURE REVIEW

Shweta Pathak et.al. (2023) presented the study on EEG signals using Linear Discriminant Analysis (LDA) supervised machine learning algorithm [2]. In the first step of signal pre-processing, Independent Component Analysis (ICA) technique was used. For features Extraction, time-frequency analysis (TFR) was done in which Gabor wavelet transform was implemented. Further, classification was done using LDA and highest accuracy obtained of 97.77%.

3. METHODOLOGY

The following flowchart shows step by step process of classification on EEG signals.

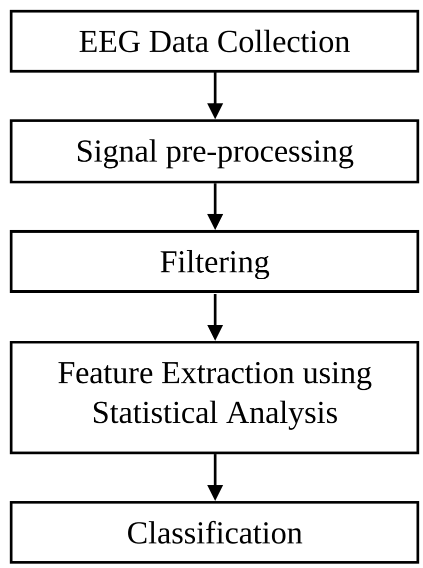

Fig. 1 Flowchart for EEG signal classification

3.1 Data Collection

EEG Data is collected online from BCI2000 system available on PhysioNet. A total of 109 subjects performed 14 different motor movement/ execution and motor imagery tasks. EEG signal recording time is one and two minutes respectively. One-minute for the tasks i.e. eyes open, eyes closed as relaxed state corresponds to T0. Two-minutes for motor movement and motor imagery tasks of left fist vs. feet corresponds to T1 and right fist vs. feet corresponds to T2. Over 1500 recordings are obtained.64

channels/electrodes are used. 160 samples per second are taken.

3.2 Signal Pre-processing

Denoising method is used for Pre-processing. Noise removal in any signal is essential task this is done by denoising. This process is slightly better than ICA pre-processing technique as it improves signal to noise ratio (SNR). In this research, XDAWN denoising method is being used. Epochs are taken as hyperparameters as one epoch passes entire training data in one cycle of algorithm. In the following Figure, epochs before denoising are shown. In Fig. 2(a) epochs for hands and in Fig. 2(b) epochs for feet are displayed.

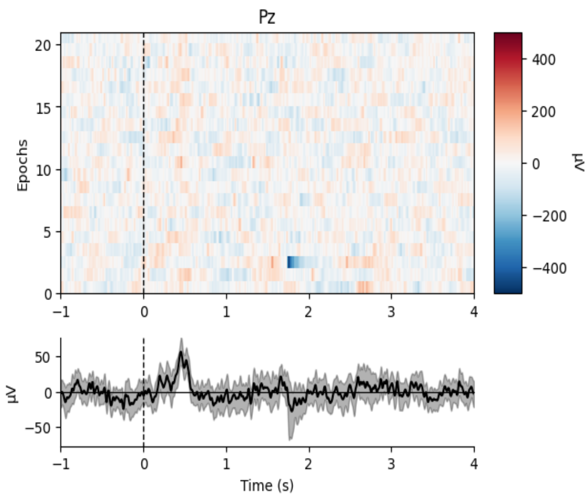

Fig. 2(a) Epochs for hands before denoising

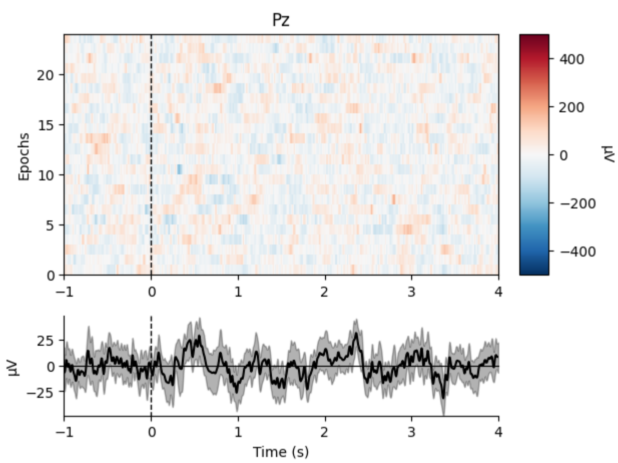

Fig. 2(b) Epochs for feet before denoising

For the denoising process, two XDAWN denoising components are taken here. After denoising, 45 matching events were found. Improved SNR in evoked response is shown in the following Figure. In Fig. 3(a) epochs for hands after denoising and in Fig. 3(b) epochs for feet after denoising are displayed.

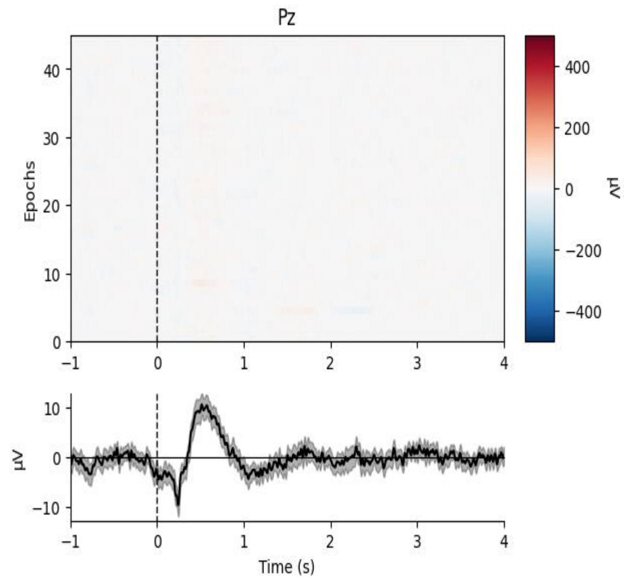

Fig. 3(a) Epochs for hands after denoising

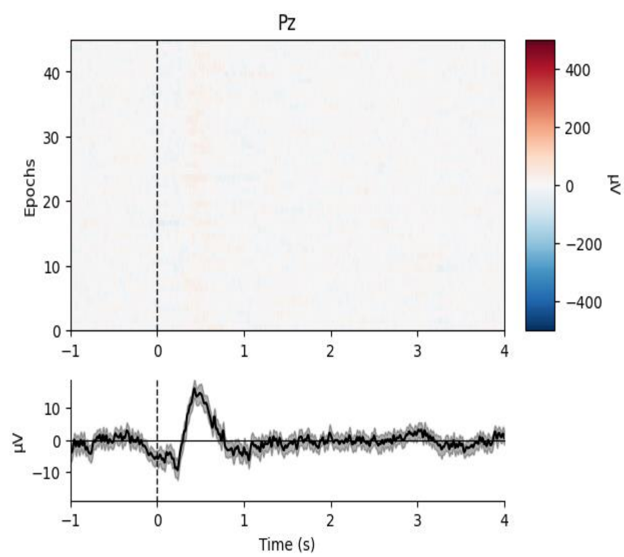

Fig. 3(b) Epochs for feet after denoising

3.3 Filtering

Further, EEG signals are filtered in different brain wave bands i.e. delta, theta, alpha, beta, gamma. To separate out these brain waves, a FIR filter for different frequency ranges is applied. This FIR filter has parameters like second order, bandpass, Butterworth, zero phase, non-causal, one pass filter. Hamming windowing method is taken. Filter length is 160 samples per second. First, frequency range is taken 1 to 4 Hz as Delta brain waves come under this frequency band. The following Fig. 4(a) shows delta waves in EEG for subject 1, run 1.

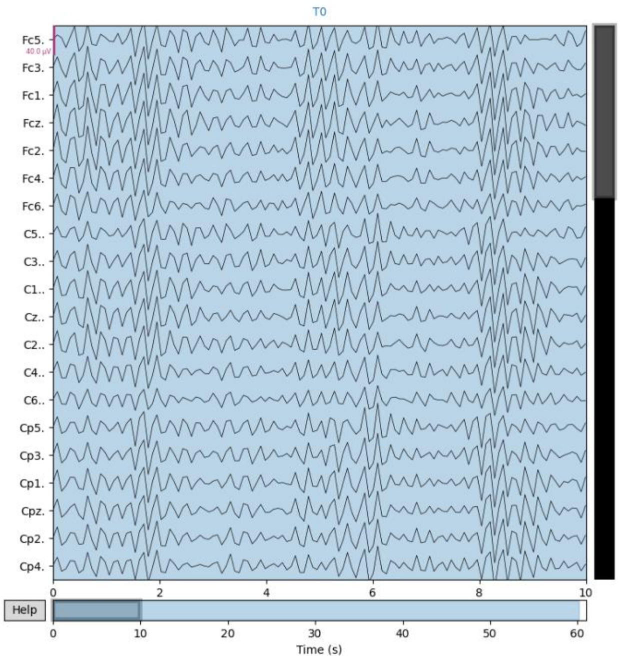

Fig. 4(a) Delta waves band for sub-1, run-1

For theta brain waves, 4 to 7 Hz frequency range is taken and it is shown in Fig. 4(b).

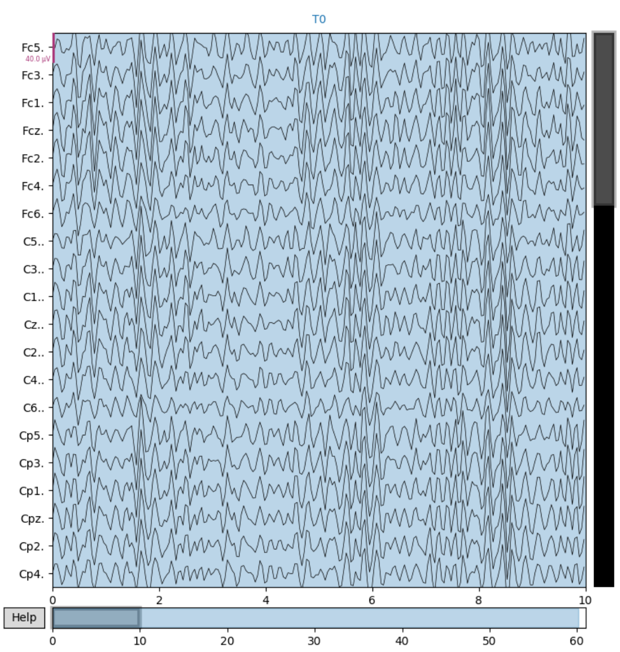

Fig. 4(b) Theta waves band for sub-1, run-1

Further, frequency range of 7 to 12 Hz is applied to separate alpha brain waves. In Fig. 4(c) alpha waves are shown.

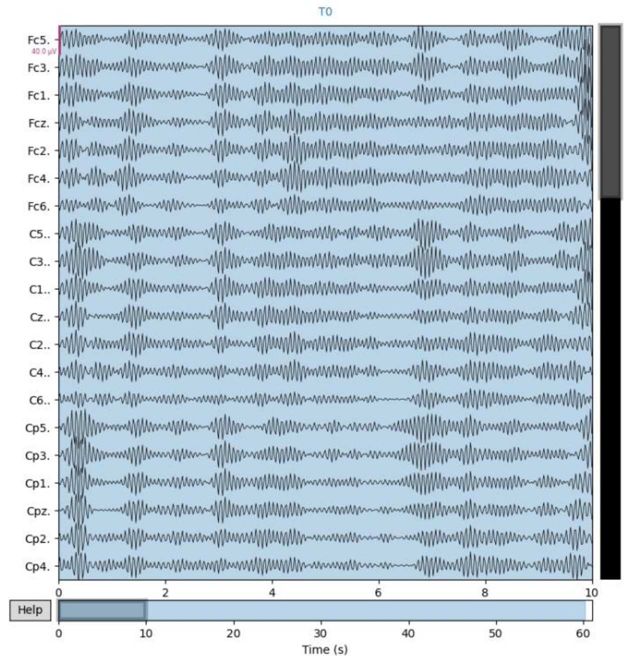

Fig. 4(c) Alpha waves band for sub-1, run-1

Next, beta brain waves are separated. Frequency range 12 to 30 Hz is used. Fig. 4(d) shows beta waves band.

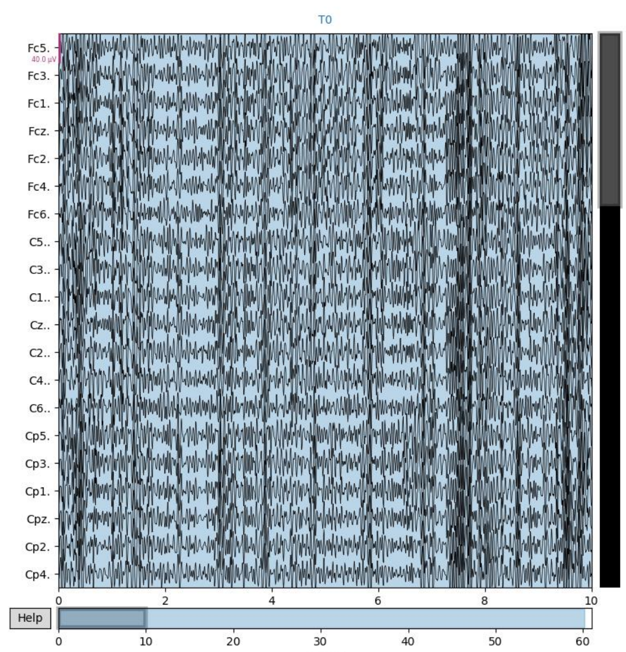

Fig. 4(d) Beta waves band for sub-1, run-1

At last, frequency range of 30 to 60 Hz is taken to filter out gamma brain waves. In Fig. 4(e) gamma waves are displayed.

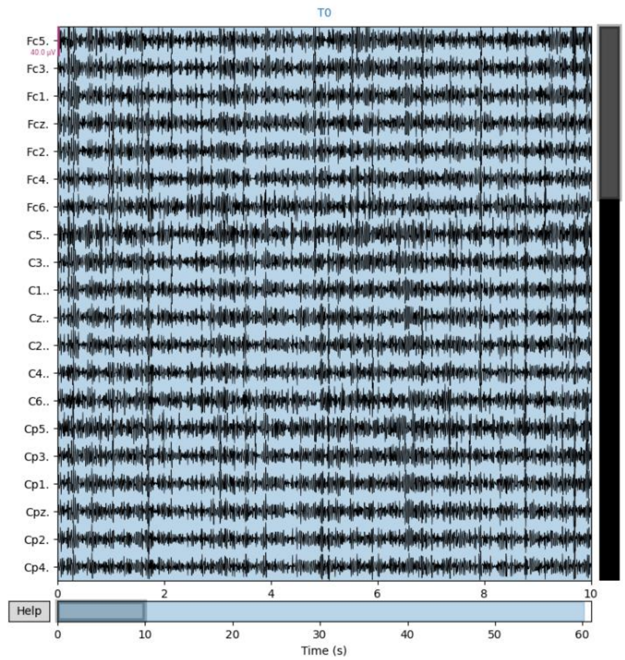

Fig. 4(e) Gamma waves band for sub-1, run-1

3.4 Feature Extraction

Statistical analysis is done for the features extraction. Various statistical tests i.e. mean, median, mode, standard deviation, variance, skewness, kurtosis are performed on epochs and events present in EEG signal. In the following table 1, different statistical tests and observations after calculations are displayed.

Table. 1 Statistical tests on epochs and events

Statistical test	Epochs	Events
Mean	3.16e-08	0.53
Median	1.05 e-07	1.0
Mode	3.8e-05	1
Mode count	Array [1111]	24
Standard Deviation	1.65e-05	0.498
Variance	2.72e-10	0.248
Skewness	-0.01265	-0.138
Kurtosis	0.531	-2.075

The mean, median, mode can be visualized using probability density function graph. Following Fig. 5(a) shows Probability Density Function (PDF) for epochs and Fig. 5(b) shows Probability Density Function (PDF) for events.

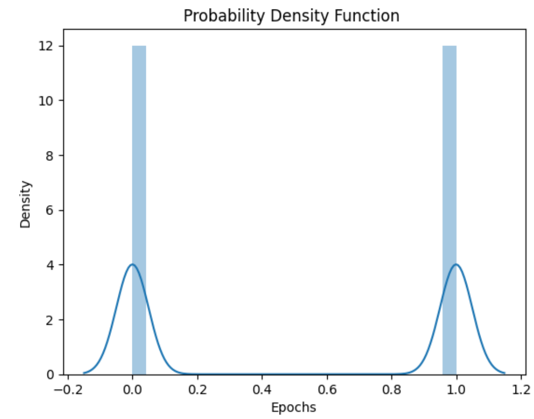

Fig. 5(a) Probability Density Function for epochs

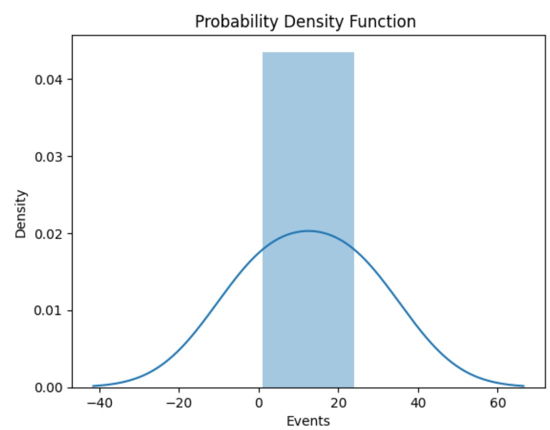

Fig. 5(b) Probability Density Function for events

Histogram for epochs is displayed in Fig. 6.

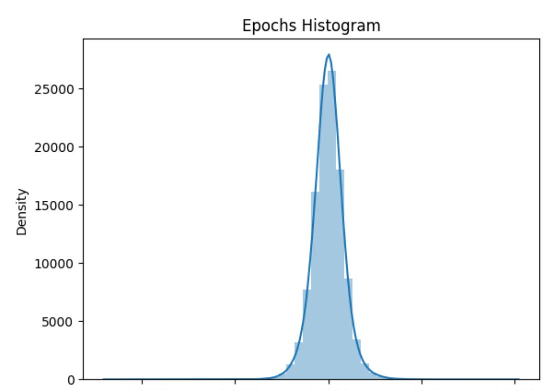

Fig. 6 Epochs Histogram

Normal distribution curve for epochs is shown in the following Fig. 7.

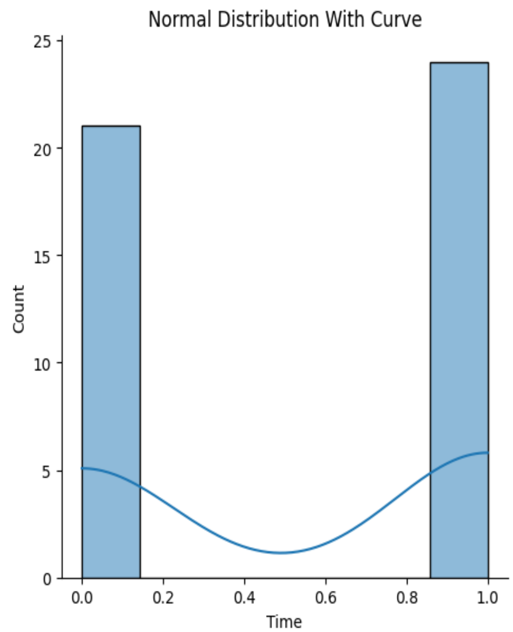

Fig. 7 Normal distribution curve for epochs

Empirical Cumulative Distribution Function (ECDF) for events is shown in Fig. 8.

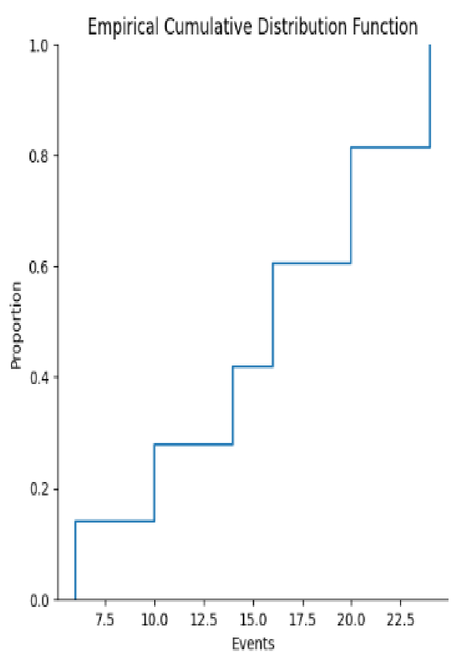

Fig. 8 Empirical Cumulative Distribution Function (ECDF) for events

The Kernel Density Estimation (KDE) is performed on events which is shown in the following Fig. 9.

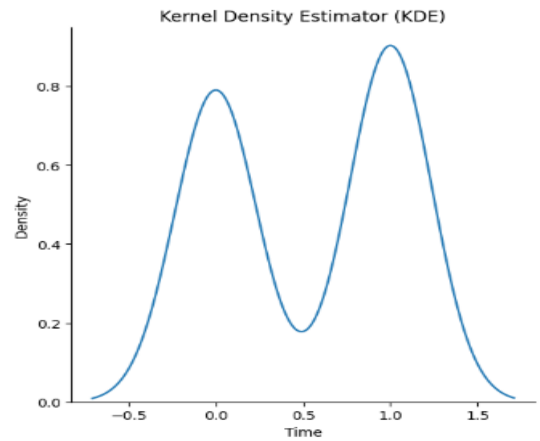

Fig. 9 Kernel Density Estimation function (KDE) for events

Karl Pearson's coefficient method is used to calculate skewness. The skewness in curve for events is -0.138 which is smaller as compared to the skewness in curve for epochs as mentioned in Table 1, so, it's not plotted. Skewness in normal distribution curve for epochs is -0.01265 which is nearer to zero so, it can be said that curve is nearly symmetrical and it is shown in the following Fig. 10.

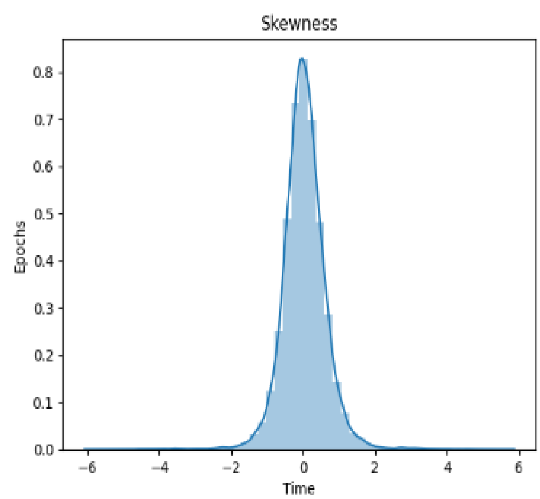

Fig. 10 Skewness in normal distribution curve for epochs

Further, kurtosis for events is -2.07 having negative value so, it is said that negative tailed kurtosis as compared to the kurtosis for epochs so, it's not plotted. In case of, kurtosis for epochs is 0.531 has positive value so, it is positive tailed platykurtic kurtosis as it has thin curve which is displayed in Fig. 11.

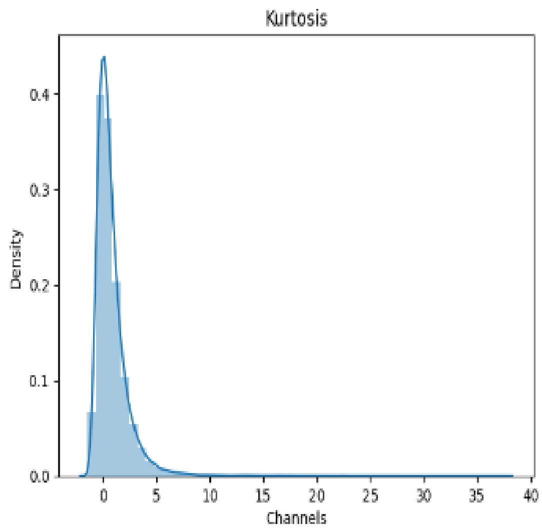

Fig. 11 Positive tailed platykurtic kurtosis for epochs

In the following Fig. 12 both skewness and kurtosis for epochs is displayed.

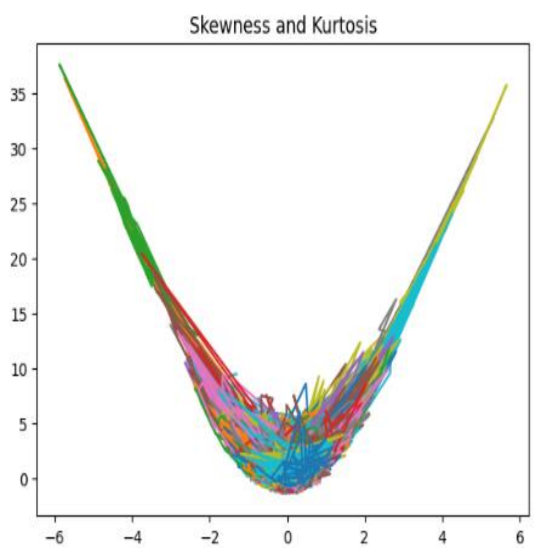

Fig. 12 Skewness and kurtosis for epochs

3.4.1 Cross-correlation

Cross-correlation is also an important measure in statistical analysis. This method is used to compare relation between the two time series variables. This research attempts to find cross-correlation between epochs and events with respect to runs of EEG data. In the following Fig. 13(a) the cross-correlation between events and runs is dispayed.

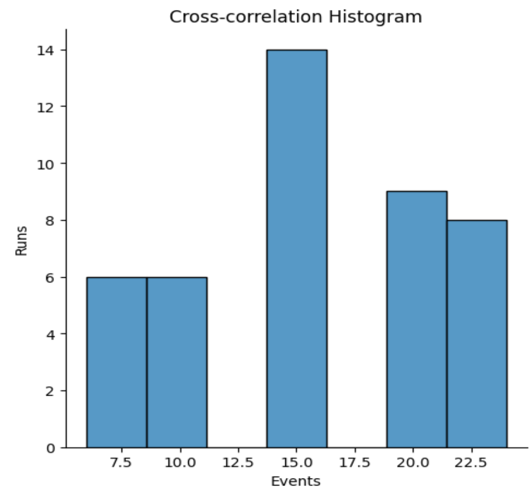

Fig. 13(a) Cross-correlation between events and runs

In the Fig. 13(b) Cross-correlation between labels of epochs and runs in channels is shown.

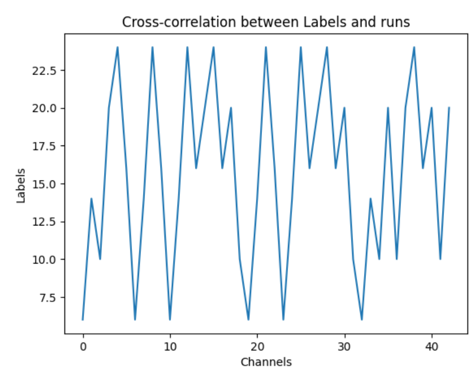

Fig. 13(b) Cross-correlation between labels and runs

3.4.2 Hypothesis test

Hypothesis test is an important part in statistical analysis. Different types of hypothesis tests, i.e. T-test, Z-test, Chi-square test, ANOVA are performed on epochs and events in EEG data. Alpha value is taken as 0.05. The following Table 2 shows observations computed in these tests.

Table. 2 Hypothesis test on epochs and events

Test	Statistic	P-value	Degree of
ANOVA	124.069	0.934	2.86
Chi-square	21.0	0.9	3.27
Z-test	-125.86	0.5	1.64
T-test	-4.097	0.054	2.0

From the above table 2, only p-value in T-test (0.054) is closer to alpha value 0.05, so, only T-test is statistically significant and null hypothesis can be rejected. So, features are extracted using only T-test and these features are further used for the next process of classification. In the following Fig. 14(a) to 14(d) statistic and p-value graph of hypothesis tests are shown.

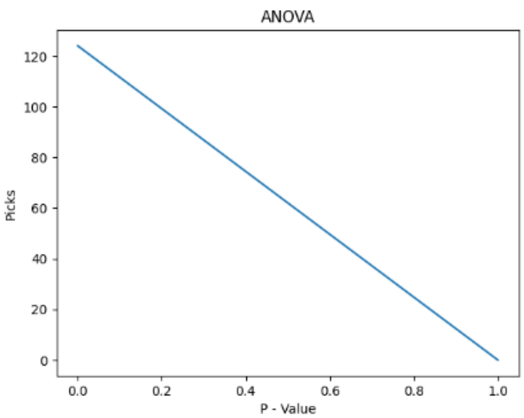

Fig. 14(a) statistic and p-value graph

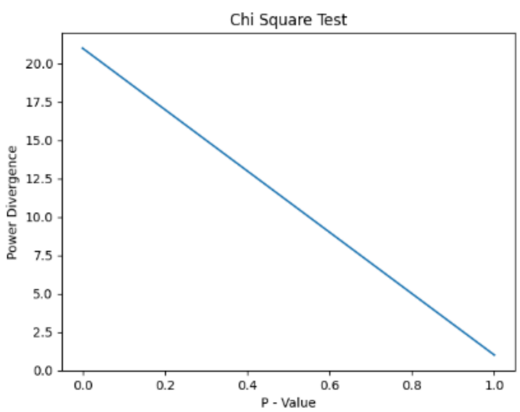

Fig. 14(b) statistic and p-value graph

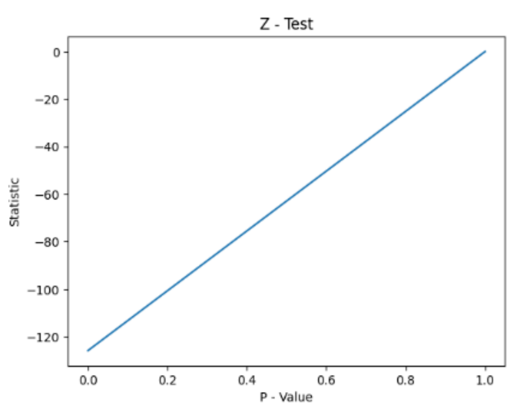

Fig. 14(c) statistic and p-value graph

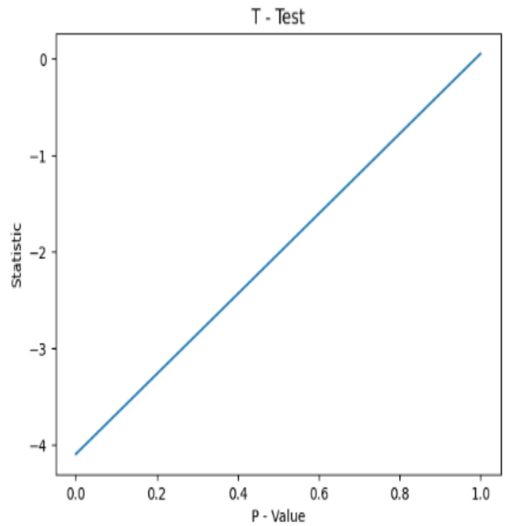

Fig. 14(d) statistic and p-value graph

3.5 Classification

Further, classification is done using various machine learning algorithms i.e. logistic regression, SVM, KNN, decision tree, Naïve Bayes, Random Forest. Training data and testing data is split in 90-10 %. Monte Carlo cross validation method is used which is also known as shuffle split method. 42 Random states are taken. The following Table 3 shows accuracies obtained in these machine learning techniques.

Table. 3 Classifier vs. accuracy

Sr. No.	Classifier	Accuracy
1	Logistic regression	66.66 %
2	KNN =12, 9, 5, 4	88.11 %
3	KNN=3	98.88 %
4	SVM	66.66 %
5	Naive Bayes	66.66 %
6	Decision tree	77.77 %
7	Random forest	77.77 %

In case of, KNN classification 2 to 14 values are taken. Highest accuracy of 98.88% is achieved only for k = 3. The following Fig. 15 shows graph of k-value range (2 to 15) vs. accuracies.

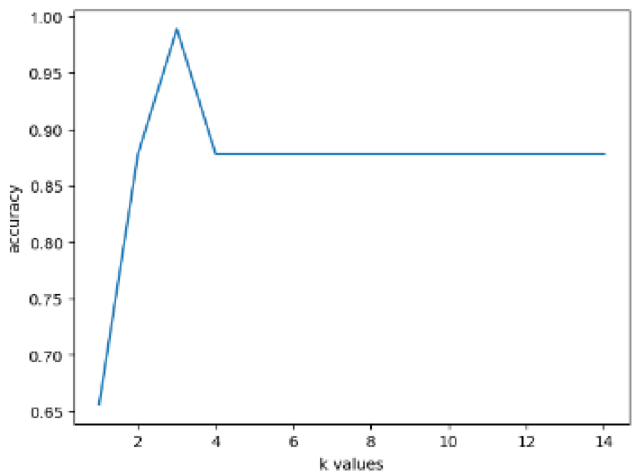

Fig. 15 K-values vs accuracies

From the above Table 3, it is found that, accuracies for logistic regression, SVM, Naïve Bayes method are same as 66.66%. Also, accuracies for decision tree and random forest method are same as 77.77%. So, it is observed that, confusion matrix graph having same accuracies of different classifiers are also same. In the following Fig. 16(a) to 16(f) heatmap of confusion matrix obtained for accuracies of classifiers are shown.

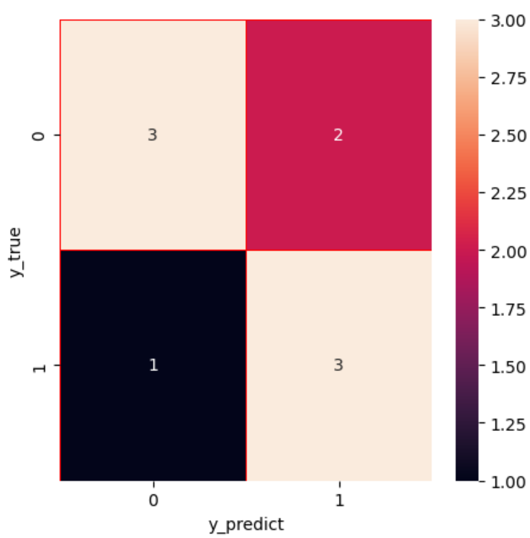

Fig. 16(a) confusion matrix for logistic regression

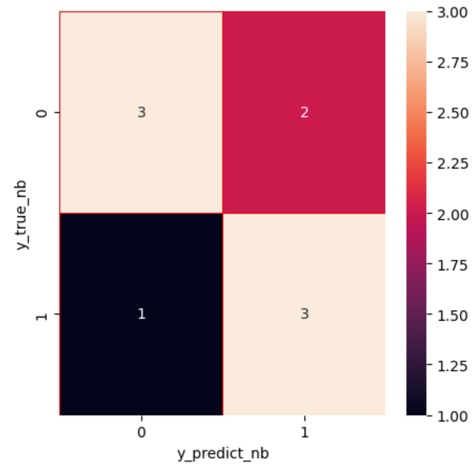

Fig. 16(b) confusion matrix for Naïve Bayes method

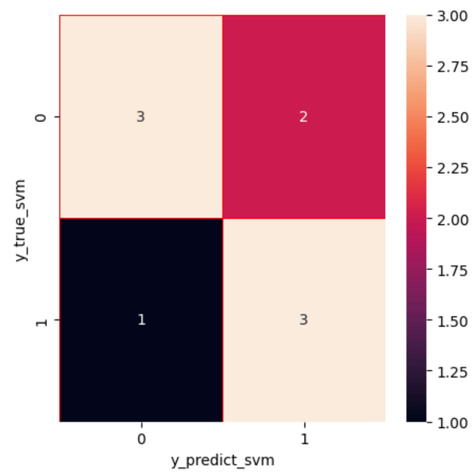

Fig. 16(c) confusion matrix for SVM

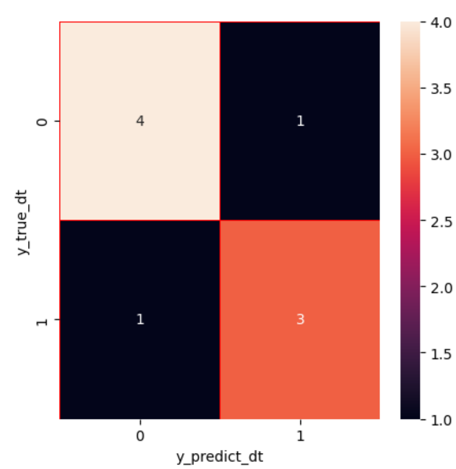

Fig. 16(d) confusion matrix for Decision Tree

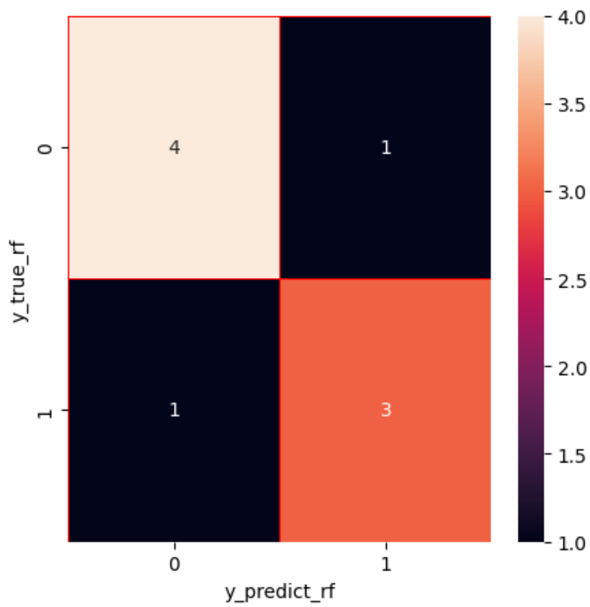

Fig. 16(e) confusion matrix for Random Forest method

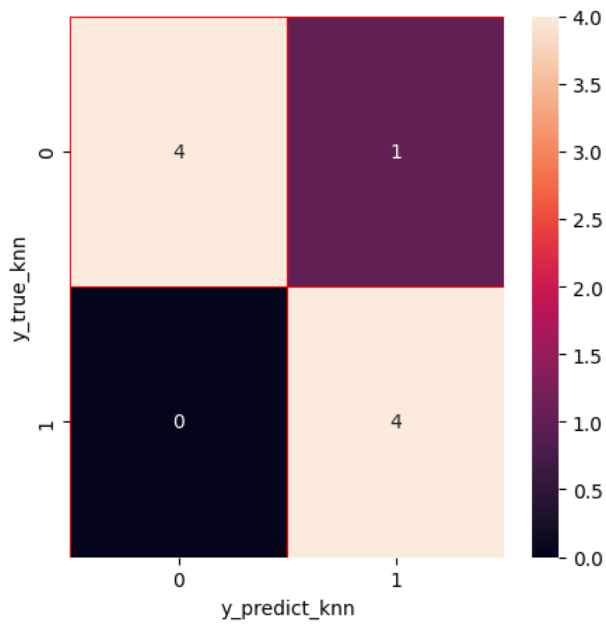

Fig. 16(f) confusion matrix for KNN classifier (K=3)

4. Results and Discussion

All accuracies achieved in various classifiers are shown in following Fig. 17. Accuracy for KNN classifier (K=3) is found highest 98.88% among others.

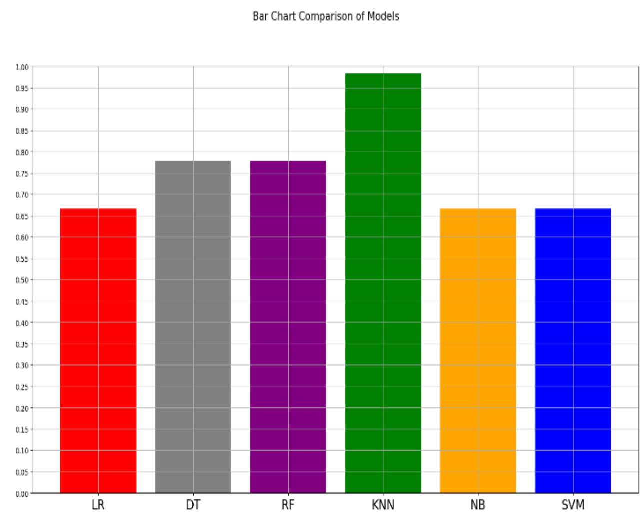

Fig. 17 Classifier vs. accuracy

The result of the present study in comparison to the result of reference [2] is shown in Table 4.

Table. 4 Validation of present study result with result of the Reference

Comparative parameter	Result of the research	Result of the Reference [2]
Accuracy	98.88%	97.77%

5. Conclusions

This research is a novel, experimental study in which bottom-up, inductive method is applied. The present study is a unique attempt of statistical analysis and classification on EEGBCI dataset. The present research has taken online EEG data which consists of 109 subjects or volunteers who had performed 14 different tasks using their hands and feet. EEG signals have been recorded using 64 channel electrodes at a sampling rate of 160 samples per second. XDAWN denoising method is used at pre-processing which improved signal to noise ratio (SNR). Then, various types of brain waves like delta, theta, alpha, beta, gamma waves are filtered using FIR filtering. Statistical tests i.e. mean, median, mode, standard deviation, variance, skewness, kurtosis are performed on epochs and events. Cross-correlation between epochs and runs also, between events and runs is shown in the methodology section. Hypothesis tests are performed to get statistically significant features. Using T-test, most prominent features are extracted. These features are used for classification. Different types of machine learning algorithms are used like logistic regression, KNN, SVM, Naïve Bayes, decision tree, random forest technique. Among all these techniques, KNN (K=3) classification accuracy is found highest as 98.88% which is slightly greater than the accuracy of the reference [2] which is 97.77%.

Future Scope: Authors are working on other methods to improve classification accuracy than the present study. Authors are also working to explore other aspects of EEG signals.

REFERENCES

1. EEG Motor Movement/Imagery dataset (2009). Version: 1.0.0, https://physionet.org/content/eegmmidb/1.0.0.
2. Shweta Suresh Pathak, Sanjiv V. Bonde (2023). Classification of Electroencephalogram (EEG) Signals Using Linear Discriminant Analysis, International Journal of Microsystems and IoT (Vol. 1-5), 2584-0495.
3. Lisha Sun, Ying Liu, P. J. Beadle (2005). Independent component analysis of EEG signals, IEEE International Workshop on VLSI Design and Video Technology, 8642877.
4. Hasminda Hassan, Zunairah Murat and Valerie Ross (2012). A preliminary study on the effects of music on human brainwaves, www.researchgate.net/publication/261206700.
5. Xu Huang, Salahiddin Altahat, Dat Tran, Sharma (2012). Human identification with electroencephalogram (EEG) signal processing, International Symposium on Communications and Information Technologies (ISCIT), 13187452.
6. J. Satheesh Kumar, P. Bhuvaneswari (2012). Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study, Procidia Engineering, Volume 38, Pages 2525-2536.
7. Amjed S. Al-Fahoum and Ausilah A. Al-Fraihat (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains, https://www.hindawi.com / journals/isrn/2014/730218.
8. Jasjeet Kaur, Amanpreet Kaur (2015). A review on analysis of EEG signals, IEEE International Conference on Advances in Computer Engineering and Applications, 15309893.
9. Nilay Yildirim, Asaf Varol (2017). A research on estimation of emotion using EEG signals and brain computer interfaces, IEEE International Conference on Computer Science and Engineering, 17318593.
10. Rab Nawaz, Humaira Nissar, Yap Voi Voon (2018). The Effect of Music on Human Brain; Frequency Domain and Time Series Analysis Using Electroencephalogram, IEEE International Conference on Computer Science and Engineering, 18068899.
11. S. Dhivya and A. Nithya (2018). A Review on Machine Learning Algorithm for EEG Signal Analysis, https://www. researchgate.net /publication/ 328768242.
12. Jaehoon Cha, Kyeong Soo Kim, Hao Lan Zhang, Sanghyuk Lee (2019). Analysis on EEG signal with machine learning, researchgate.net, DOI:10.11.17/ 12.2548313.
13. Tom Di Fulvio (2019). Using Machine Learning to Categorize EEG Signals from the Brain, https://towardsdatascience.com/using-machine-learning-to categorise-eeg-signals-from-the-brain-to-words-728aba93b2b3.
14. Sho Nakagome, Trieu Phat Luu, Yongtian He, Akshay Sujatha Ravindran and Jose L. Contreras-Vidal (2019). An empirical comparison of neural networks and machine learning algorithms for EEG decoding, https://www.nature.com / articles/s41598-020-60932-4.
15. Sheng Li, Hanxin Feng (2019). EEG Signal Classification Method Based on Feature Priority Analysis and CNN, IEEE International Conference on Communications, Information System and Computer Engineering, 19007751.
16. Jianhua Wang, Gaojie Yu, Liu Zhong, Weihai Chen, Yu Sun (2019). Classification of EEG signal using convolutional neural networks, IEEE Conference on Industrial Electronics and Applications (ICIEA), 18993324.
17. Raju Vishwakarma, Hazim Khwaja, Varad Samant, Prajyot Gaude, Mayur Gambhir, Shailendra Aswale (2020). EEG Signals Analysis And Classification For BCI Systems: A Review" IEEE International Conference on Emerging Trends in Information Technology and Engineering, 19608556.
18. Ayman M. Anwar, Ayman M. Eldeib (2020). EEG Signal Classification Using Convolutional Neural Networks on Combined Spatial and Temporal Dimensions for BCI Systems, 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 19916239.
19. Pengcheng Ma, Qian Gao (2020). EEG Signal and Feature Interaction Modeling-Based Eye Behavior Prediction Research, Computational Intelligence Methods for Brain-Machine Interfacing or Brain-Computer Interfacing, Volume 2020, Article ID : 2801015.
20. Zheng Li (2020). Electroencephalography Signal Analysis and Classification Based on Deep Learning, International Conference on Information Science, Computer Technology and Transportation (ISCTT), 20411484.
21. Vikrant Doma and Martin Pirouz (2020). A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals, https://journalofbigdata. springeropen.com/articles/10.1186/s40537-020-00289-7.
22. Xiongliang Xiao, Yuee Fang (2021). Motor Imagery EEG Signal Recognition Using Deep Convolution Neural Network, researchgate.net, License: CC BY 4.0
23. Ghita Amrani, Amina Adadi, Mohammed Berrada, Zouhayr Souirti, Saïd Boujraf (2021). EEG signal analysis using deep learning: A systematic literature review, Fifth IEEE International Conference On Intelligent Computing in Data Sciences, 21488992.
24. Indurani P., Firdaus Begam B. (2021). A Detailed Analysis of EEG Signal Processing in E-healthcare Applications and Challenges, International Journal of Innovative Research in Science Engineering and Technology.
25. Swarali Belsare, Maitreyi Kale, Priya Ghayal, Aishwarya Gogate, Suhasini Itkar (2021). Performance Comparison of Different EEG Analysis Techniques Based on Deep Learning Approaches, IEEE International Conference on Emerging Smart Computing and Informatics, 20553606 .

26. Tao Wu, Xiangzeng Kong, Yiwen Wang, Xue Yang, Jingxuan Liu, Jun Qi (2021). Automatic classification of EEG signals via deep learning, IEEE International Conference on Industrial Informatics, 21229539.
27. Min Wu, Bo Nan, Changqing Li (2022). EEG Classification Based On Deep Learning, IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 21931788
28. German Guyo (2022). Analysis of EEG signals using multiresolution wavelet analysis and its extensions, IEEE International Conference on 6th Scientific School Dynamics of Complex Networks and their Applications (DCNA), 22159980
29. Prakash Chandra Sharma, Rohit Raja (2022). Analysis of brain signal processing and real-time EEG signal enhancement, https://www.researchgate.net/publication/360628946_Analysis_of_brain_signal_processing_and_real-time_EEG_signal_ enhancement
30. Anupreet Kaur Singh, Sridhar Krishnan (2022). Trends in EEG signal feature extraction applications, https://www.frontiersin.org/articles/10.3389/frai.2022.1072801/full, Volume 5
31. Motor imagery decoding from EEG data using the Common Spatial Pattern (CSP), https://mne.tools/dev/auto_examples/decoding/decoding_ csp_eeg.html #ex-decoding-csp-eeg

AUTHORS:

Shweta S. Pathak received her bachelor's degree in Electronics Engineering from Shri Ramdeobaba College of Engineering and Management, Nagpur in 2020 and Master's degree in Artificial Intelligence from Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded in 2023. Her areas of interest are signal processing and machine learning.
E-mail: shwetapathak2798@gmail.com

Nilkanth H. Kulkarni received his bachelor's degree in civil-W.M. Engineering from Shri Guru Gobind Singhji College of Engineering and Technology, Nanded in 1988 and Master's degree in Civil –W.M. Engineering from Shri Guru Gobind Singhji College of Engineering and Technology, Nanded in 1996 and Ph.D. in Civil Engineering from Indian Institute of Technology, Bombay in 2008. His areas of interest are soft computing tools in hydraulic engineering and water resources engineering.

Corresponding author E-mail: nhkulkarni@sggs.ac.in