EEG-Based Stress Detection Using K-Means Clustering Method

Soumya Samarpita(✉) and Rabinarayan Satpathy

FOS, Sri Sri University, Cuttack, Odisha, India
{soumya.s2020-21ds,rabinarayan.s}@srisriuniversity.edu.in

Abstract. Stress, sadness and panic have all become major issues in our contemporary culture. Stress has become one of the top ten socioeconomic predictors of health inequalities. The electroencephalogram (EEG) signals and machine learning approaches are utilized to predict the mental state of the person. This has become a significant topic of research in recent times in health care system. There are various ways are used to monitor stress. The primary goal of this study is to identify stress in humans. Because of its potential value, stress detection based on EEG signals has emerged as an interesting study topic. This research looks into brain waves to classify a person's mental state. Despite the fact that there is no precise way of defining the optimum feature for a classifier, the features utilized as classifier input have a significant impact on the classification outcomes. An algorithm for stress level detection from EEG is proposed in this paper. The Euclidean distance scale is commonly used in the paper for EEG signal identification. In this study, EEG data is separated into EEG rhythms using a band pass filter method, EEG signals are normalized and a k-mean clustering method is used to classify brain wave signals to detect the mental stress.

Keywords: Brain Waves · Mental Stress · Electro-encephalogram (EEG) · EEG Signals · K-Means clustering

1 Introduction

The majority of individuals encounter stress at some point in their lives. Stress can be caused by a variety of factors. It is a common bodily reaction to being challenged or frightened by the surrounding. The body's response to mental, emotional, or physical discomfort is classified as stress. Stress not only causes unstable behavior, but it may also increase hypertension [17]. Stress is a psychological reaction produced by external stimuli. The two aspects that are followed are mental and physical [15]. Stress can be caused by unfamiliar situations, increasing expectations at work, and emotional reactions to the loss of a loved one or abrupt changes in one's lifestyle. Due to its potential applications, such as in "Human-Computer Interaction" (HCI), the study of mental stress detection using EEG signals has great promise [8]. The utilization of brainwave signals is a step toward the introduction of individual identification through biometric technology based

S. Nandan Mohanty et al. (Eds.): ICISML 2022, LNICST 470, pp. 35–43, 2023.
https://doi.org/10.1007/978-3-031-35078-8_4

on bodily features. Each individual's brainwave signal has distinct properties. Periodic stress tracking is required to control the stress experience. It is critical to alleviate human stress; there are various strategies available for doing so, such as yoga and meditation.

Different patterns of neuronal interaction in the human brain result in different brain states. The signal waves are distinguished by their amplitudes and frequencies based on these patterns. The neuronal contact occurs between a vast numbers of neurons. EEG is a medium of systematically measuring the electrical activity of central nervous system. These recordings are made on the scalp, and numerous electrodes are inserted in several unique areas on the scalp. Scanning of recorded EEGs assists in defining the condition of the brain with deviations from normal, such as sleep problems, epileptic seizures, mental tension, memory loss etc. on [13].

Different methods, including "PET" (Positron Emission Tomography), "MRI" (Magnetic Resonance Imaging), and "fMRI" (Functional Magnetic Resonance Imaging), can be used to capture brain activity. However, because of its asymptomatic, precision in real-time mode of operation, and optimum suitability for the complicated vitality of human brain processes, EEG signals are the ideal choice to capture the neuronal changes [6]. Medical professionals do not need patient involvement when using EEG-based analysis for diagnostic purposes. In order to distinguish between various types of mental stress based on various variables, this study introduces k-nearest neighbor to be employed as classifiers using the frequency domain feature extraction. We also suggest feature choices to drastically minimize the amount of characteristics, hence reducing the computational complexity.

The format of this article is as follows: The comparative analysis is shown in Sect. 2, the proposed methodology for this study is presented in Sect. 3, the experiment's findings are presented in Sect. 4, and the conclusion is presented in Sect. 5.

2 Comparative Analysis

Stress in humans is induced by psychological, intellectual, or bodily resistance to new difficulties. When stress levels surpass a particular threshold, it can have serious consequences for one's health, mood, creativity, relationships, and life experiences. As a result, early recognition of mental stress is critical for avoiding such negative consequences. EEG microstate analysis is also more effective than resting-state fMRI analysis in terms of its capacity to test expensively big groups of people [16]. The brainwaves signal has various typical and characteristic of the individual, and because brainwaves cannot be duplicated or read by individuals, similarity is not conceivable. Identity identification is required to differentiate an individual's traits [4]. EEG waves in different mental states include:

a) Beta waves (13–35 Hz frequency) b) Alpha waves (8–13 Hz frequency) c) Theta waves (4–8 Hz frequency) and d) Delta waves (0.5–4 Hz frequency).

After analyzing various research papers [8, 13], we found that theta waves are mostly produced in the drowsy state, while theta waves are primarily connected with deep sleep. The alpha wave represents a relaxed level of awareness, whereas the beta wave mostly represents a focused state. Gamma waves seldom ever occur.

In this section, the EEG signals from different articles are analyzed with different techniques which are described in Table 1.

Table 1. Outline of papers examining how EEG waves affect the human brain.

Reference, Years	Aims	Samples	Methods	Conclusions
[14], 2022	"To examine features based on time-domain and frequency-domain for categorising the EEG responses to auditory evoked potentials (AEPs)"	8 healthy participants	As a method for extracting EEG characteristics, the Fast Fourier Transform (FFT), power spectral density (PSD), spectral centroids have been used Classification:- SVM, LDA, K-NN	Accuracy:- 82.86%
[18], 2022	"To calculate the theta-to-alpha transition frequency from electroencephalographic data in the resting state"	Dataset:- OpenNeuro 25 Parkinson's patients and 25 matched controls	Klimesch's method K-mean Clustering	Accuracy:- 88% for 2D adjusted k-mean
[7], 2022	"Utilizing audio, video, EEG, and EMG to identify emotions"	Emotion Dataset:- CK, CK+ (11 Human Subjects) EEG Dataset:- DEAP	Feature Extraction:- PCA Classification:- SVM, K-NN,CNN,MLP	Emotion accuracy ranging from 43.4% to 86.1%. Using only eight channels of EEG data and the conventional KNN model, they attained 39.70% accuracy
[3], 2022	"To provide a quick-processing strategy for efficiently extracting and choosing spatial features for emotion recognition"	Dataset:- SEED (15 Chinese participants), DEAP (32 participants)	Features are extracted from DNN Classification:- SVM, K-NN	Accuracy for SEED:- 96.3% (SVM), 86.4% (K-NN) Accuracy for DEAP:- 81.1% (SVM), 79.3% (K-NN)

(*continued*)

Table 1. (*continued*)

Reference, Years	Aims	Samples	Methods	Conclusions
[16], 2022	"To evaluate EEG microstates in MwoA patients during the interictal period"	50 Female participants. "NeuroScan" was used to record the EEG data	Feature Extraction:- ICA	Microstate classes B, C, and D were the main areas where microstate syntactic analysis revealed significant differences in transition probabilities between the two groups
[10], 2022	"To suggest a cutting-edge signal processing method for EEG signal emotion identification utilising continuous wavelet transform"	Dataset:- SEED (15 participants)	The BoDF reduces the features Classifiers:- SVM, K-NN	Accuracy for SVM:- 96.7% Accuracy for K-NN:- 95.3%
[1], 2022	"To choose the optimum classification method for the death and surviving cases using COVID19MPD in Mexico"	Dataset:- COVID19MPD Samples:- 200,000patients (30 to 50ages)	Classifiers:- KNN, Naïve Bayes, Random Forest	Accuracy:- 94.88%
[15], 2021	"To detect stress management using BCI technique"	Raw EEG set	Feature Extraction:- Hjorth, KDE, RER, ELC Classification:- KNN, SVM, NN	Accuracy:- 89.03% for ELC + KNN
[13], 2021	"To examine a discrete wavelet-based feature extraction model for the categorization of EEG signals during a mental arithmetic exercise"	32 participants	Feature Selection:- Neighborhood Component Analysis (NCA) Classification:-KNN	Accuracy:- 91%
[2], 2021	"To investigate the electroencephalogram (EEG) data generated when typing"	5 engineering students	Classification:- kNN, Random Forest	Accuracy:- 98.91% (kNN) 99.89% (RF)
[19], 2021	"To recognize emotions from brain signals for efficient human-robot interaction"	Dataset:- SEED 15 Subjects	Classification:-kNN Optimization:- Genetic Algorithm	Accuracy:- 80.59%

(*continued*)

Table 1. (*continued*)

Reference, Years	Aims	Samples	Methods	Conclusions
[9], 2018	"To Correlate RAW EEG Signals and Attention Levels of BCI using KNN Technique"	Dataset of 300 mind waves samples	Classification:- kNN	Researchers discovered an excellent relationship between "attention levels" and "RAW EEG" data
[12], 2017	"To evaluate the effectiveness of the K-Nearest Neighbors (K-NN) algorithm for classifying motor imagery using an EEG signal"	4 healthy human subjects	Classification:- SVM, kNN Distance metrics:- Manhattan, Euclidean, Minkowski, Chebychev and Hamming	The Minkowski distance has the best classification accuracy (70.08%)

3 Proposed Methodology

The technique, suggested model, and how it operates are all included in this section. Additionally, it provides a thorough summary of the feature extraction, classification, and EEG pre-processing. The brain interface method is far superior to other methods because it can analyze stress with a high degree of accuracy using a variety of feature extraction and classifier algorithm combinations. Based on different literature review, our proposed model for the mental stress detection approach utilizing the brain interface technology is illustrated in Fig. 1.

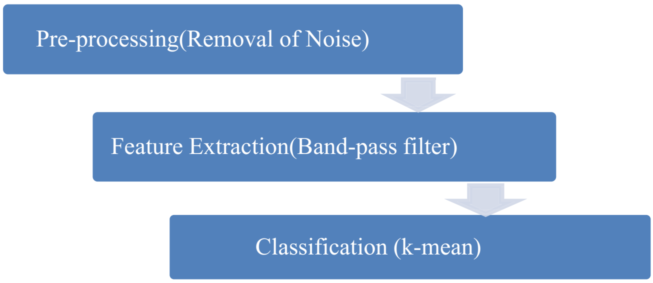

Fig. 1. Stress detection with brain interface techniques

The power per frequency band is the EEG signal characteristic that is most frequently employed for mental stress detection. This suggested method leverages EEG generated by sensors for analysis and is useful for real-time mode stress detection. In this study, multiple temporal and frequency-domain characteristics are used to analyze the EEG signals, and the recovered features are then categorized using the K-means clustering method. In this regard, a bandpass filter with a range from 0.5 to 60 Hz is applied for EEG classification. The experiment in this paper uses four bands (BETA, ALPHA, THETA, and DELTA). The frequency domain properties are used to compute band power for these four bands.

The K-means [5], simplest unsupervised clustering technique, which divides each item into a cluster based on the observation's mean. K-means typically moves from one cluster to another in finding of the k-partition calculating within cluster sum of square. K-means clustering may divide an input set of n parameters into an equal number of clusters, with each observation belonging to the cluster that is closest to the centroids, which act as prototypes for the clusters. By reducing the Euclidean distance between the "data" and the matching "cluster centroid", data are clustered. K-mean clustering divides the data into 'k' clusters in this manner [11].

K-Means is a clustering technique that categorizes data into k groups. Because we may choose the number of clusters, it is easily applicable in classification, where we split data into clusters that are equal to or more than the number of classes.
K-Means Algorithm:-

Step 1: First, determine the number of clusters 'K.'
Step 2: As centroids for each cluster, randomly select K data points. The value of 'K' will be 2 if there are two clusters.
Step 3: Iterate numerous times until the data points allocated to clusters do not change.
Step 4: Sum the squared distances between the data points and the centroids.
Step 5: To minimise the distance, assign each data point to the nearest cluster (centroid).
Step 6: Take an average of the centroids of the clusters that are related.

This is a one-time operation that computes the centroid and assigns points to the cluster depending on their distance from the centroid. The procedure is terminated once all centroids have been established.

In this paper, Euclidean distance is calculated to find the mean. The k-means clustering technique measures the similarity of items using the Euclidean distance. For typical k-means clustering, both iterative and adaptive algorithms are available. K-means clustering techniques must make the assumption that the number of groups (clusters) is known in advance.

The Euclidean distance is as follows:-

$$dis(m, n) = \sqrt{\sum_{i=1}^{N} (m_i - n_i)^2} \tag{1}$$

where, dis = distance from point m to point n, N = N- sample space and m, n = two points, m_i, n_i = Euclidean vectors are drawn from the origin of space (initial point)

4 Results

The data extracted from the characteristic findings are normalized and grouped into four groups using K-Means Clustering. These four groups are alpha, beta, theta and delta waves respectively. Then these waves are converted to corresponding EEG signals.

These EEG signals are normalized using band-pass filter method to get the features like frequency and phase. After normalization, frequencies vs. phase points are coordinated to form the cluster. Figures 2 shows the frequency of the alpha wave, beta wave, theta wave and delta wave plotted with phase in degree.

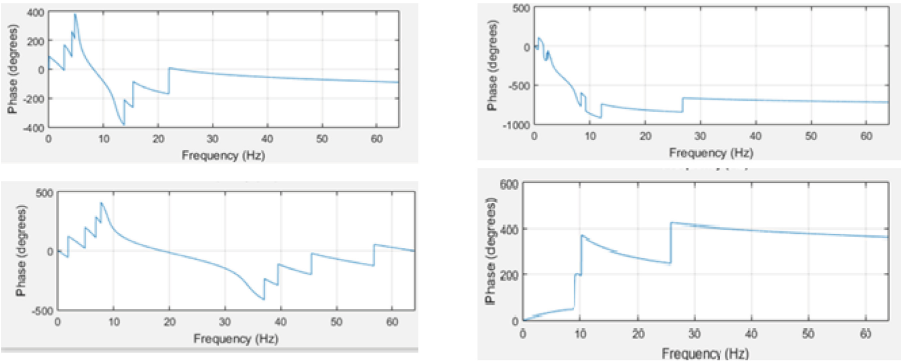

Fig. 2. Frequency of alpha, beta, theta and delta Wave

The findings are achieved using Matlab to execute the K-Mean approach. In this study, 32 samples are taken to classify the EEG signals. Here, only four clusters C1, C2, C3 and C4 are taken into consideration and the data matrix is grouped using the k-means clustering method. These four clusters represent the mental state of the person. The C1 cluster indicates that the subject is sleeping deeply (delta wave). The C2 cluster indicates that a person is extremely calm (theta wave). The C3 cluster denotes when the individual is in a highly calm condition (alpha wave), whereas the C4 cluster reflects the beta wave, which means the person is in an active or relaxed state.

Table 2. Output result of Stress Analysis.

Cluster #	Output sample points for each cluster							
	V1	V2	V3	V4	V5	V6	V7	V8
C1	0.07	1.04	1.48	1.98	1.52	2.84	3.11	3.97
C2	4.75	5	5.75	6.11	6.74	8.13	8.44	9.6
C3	10.52	11.06	12.07	12.98	13.91	14.33	15.07	16.67
C4	18.13	19.2	21.21	22.48	23.89	24.96	25.63	27.94

Here Table 2 depicts a person's stress analysis.. Four clusters are classified as being in various states of stress because we are considering 32 samples and each cluster has 8. From these points, C1 shows the delta wave (0 Hz–4 Hz), C2 represents the theta wave (4 Hz–8 Hz), C3 represents the alpha wave (8 Hz–16 Hz), and C4 shows the beta wave (16 Hz–35 Hz).

The summaries of different clusters that represent stressed or relaxed condition of the person are depicted in the Fig. 3.

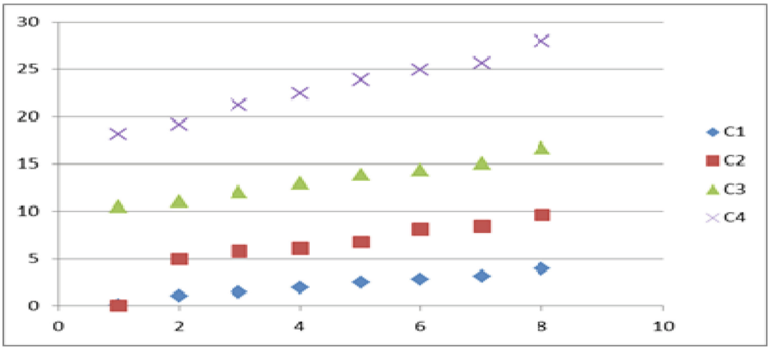

Fig. 3. Clustering with K-means

5 Conclusion

The EEG signal is an electrical signal produced by coordinated brain activity. EEG may be used to detect anomalies in brain waves and classify distinct mental states. Classifiers are used to classify EEG signals in order to detect anomalies in brain waves. This suggested classification approach is distinct and makes it extremely simple to identify EEG data. In this paper, a real-time EEG-based stress detection algorithm is used. The K-Mean clustering method is used to produce four stages of stress and EEG data is used to check the suggested stress detection system. The K-mean clustering method is also used to investigate frequency characteristics. In the future, it is intended to collect EEG data from the participants and a database can be created which will be used for EEG-based stress identification. And, different feature extraction and classification techniques could be analyzed for further study.

References

1. Almustafa, K.M.: Covid19-Mexican-patients' dataset (Covid19MPD) classification and prediction using feature importance. Concur. Comput. Pract. Exp. **34**(4), e6675 (2022)
2. Amalina, I., Saidatul, A., Fook, C.Y.: Frequency bands based on EEG typing for biometric authentication. In: AIP Conference Proceedings, vol. 2339, no. 1, p. 020170. AIP Publishing LLC, May 2021
3. Asghar, M.A., Khan, M.J., Rizwan, M., Shorfuzzaman, M., Mehmood, R.M.: AI inspired EEG-based spatial feature selection method using multivariate empirical mode decomposition for emotion classification. Multimedia Syst. **28**(4), 1275–1288 (2022)
4. Azhari, A., Hernandez, L.: Brainwaves feature classification by applying K-means clustering using single-sensor EEG. Int. J. Adv. Intell. Inform. **2**(3), 167–173 (2016)

5. Bablani, A., Edla, D.R., Kuppili, V., Ramesh, D.: A multi stage EEG data classification using k-means and feed forward neural network. Clin. Epidemiol. Glob. Health **8**(3), 718–724 (2020)
6. Bavkar, S., Iyer, B., Deosarkar, S.: Detection of alcoholism: an EEG hybrid features and ensemble subspace K-NN based approach. In: Fahrnberger, G., Gopinathan, S., Parida, L. (eds.) ICDCIT 2019. LNCS, vol. 11319, pp. 161–168. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05366-6_13
7. Chen, J., Ro, T., Zhu, Z.: Emotion recognition with audio, video, EEG, and EMG: a dataset and baseline approaches. IEEE Access **10**, 13229–13242 (2022)
8. Diah, K.T., Faqih, A., Kusumoputro, B.: Exploring the feature selection of the EEG signal time and frequency domain features for k-NN and weighted k-NN. In: 2019 IEEE R10 Humanitarian Technology Conference (R10-HTC) (47129), pp. 196–199. IEEE, November 2019
9. Garg, P., Singh, R.P., Mehra, M.: To correlate RAW EEG signals and attention levels of BCI using KNN technique (2018)
10. Haq, Q.M.U., Yao, L., Rahmaniar, W., Islam, F.: A hybrid hand-crafted and deep neural spatio-temporal EEG features clustering framework for precise emotional status recognition. Sensors **22**(14), 5158 (2022)
11. Hegde, N.N., Nagananda, M.S., Harsha, M.: EEG signal classification using k-means and fuzzy c means clustering methods. Int. J. Sci. Technol. Eng **2**(1), 1–5 (2015)
12. Isa, N.E.Z.M., Amir, A., Ilyas, M.Z., Razalli, M.S.: The performance analysis of K-nearest neighbors (K-NN) algorithm for motor imagery classification based on EEG signal. In: MATEC Web of Conferences, vol. 140, p. 01024. EDP Sciences (2017)
13. Islam, A., Sarkar, A.K., Ghosh, T.: EEG signal classification for mental stress during arithmetic task using wavelet transformation and statistical features. In: 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), pp. 1–6. IEEE, July 2021
14. Islam, M.N., Sulaiman, N., Rashid, M., Mustafa, M., Jadin, M.: Investigation of time-domain and frequency-domain based features to classify the EEG auditory evoked potentials (AEPs) responses. In: Ab. Nasir, A.F., Ibrahim, A.N., Ishak, I., Mat Yahya, N., Zakaria, M.A., P. P. Abdul Majeed, A. (eds.) Recent Trends in Mechatronics Towards Industry 4.0. LNEE, vol. 730, pp. 497–508. Springer, Singapore (2022). https://doi.org/10.1007/978-981-33-4597-3_45
15. Lahane, P.: Brain computer interfaces techniques for stress management. Inf. Technol. Ind. **9**(3), 767–775 (2021)
16. Li, Y., et al.: Abnormalities in resting-state EEG microstates are a vulnerability marker of migraine. J. Headache Pain **23**(1), 1–12 (2022). https://doi.org/10.1186/s10194-022-01414-y
17. Shon, D., Im, K., Park, J.H., Lim, D.S., Jang, B., Kim, J.M.: Emotional stress state detection using genetic algorithm-based feature selection on EEG signals. Int. J. Environ. Res. Public Health **15**(11), 2461 (2018)
18. Vallarino, E., Sommariva, S., Famà, F., Piana, M., Nobili, F., Arnaldi, D.: Transfreq: A Python Package for Computing the Theta-to-Alpha Transition Frequency From Resting State Electroencephalographic Data. Wiley, Hoboken (2022)
19. Vrochidou, E., Lytridis, C., Bazinas, C., Papakostas, G.A., Wagatsuma, H., Kaburlasos, V.G.: Brain signals classification based on fuzzy lattice reasoning. Mathematics **9**(9), 1063 (2021)