Journal of Holistic Medical Technologies
Vol.1 No.1 (2024) 31 – 39
Available online at https://journal.innoscientia.org/index.php/jhmt

Classification of normal and relaxed conditions based on brain signal activity with Electroencephalography using k-Nearest Neighbor (kNN)

Thooriq Nur Ali[1] and Fadhillah[2*]
[1] Department of Information Technology, Faculty of Science and Technology, Universitas Islam Negeri Walisongo Semarang, Indonesia
[2] Department of Physics, Faculty of Science and Technology, Universitas Islam Negeri Walisongo Semarang, Indonesia
*Corresponding author's e-mail: Fadhillahf565@gmail.com

ABSTRACT

Research on how to think, patterns that occur when humans act, human conditions observed through brain waves, and various studies related to the human nervous system and coordination are still extensive in scope for development. Electroencephalography (EEG) is an instrument commonly used to develop research related to the mechanism of brain wave activity. Changes in the brain's electrical potential can be exploited by carrying out specific analyses using various signal processing methods, which are grouped into various categories of waves, including delta (0 - 4 Hz), theta (4 - 7 Hz), alpha (8 - 12 Hz) and beta (12 – 30 Hz). This research has succeeded in building a classification model through several stages, namely preprocessing with filtration and data extraction, data processing consisting of clustering using the K-Means algorithm, and classification using the k-Nearest Neighbor (kNN) algorithm to calculate the accuracy value of the model. The classification process produces two categories of conditions: normal and relaxed. The results of testing the classification model using the k-Nearest Neighbor (kNN) produce an accuracy value of 88%.

Keywords:
Elektroencephalography, k-Means, k-Nearest Neighbor, Normal and relaxed conditions.

Introduction

The brain is an organ in humans that stores information that is still difficult to understand directly. The brain contains billions of nerve cells or neurons that regulate all coordination mechanisms of the human body and is one of the most complex systems ever in the universe (Anggara & Rahayu, 2020). Research on how to think, patterns that occur when humans perform an action, human conditions observed through brain waves, and various studies related to the nervous system and human coordination are still extensive scopes to be developed in neuroscience. Instruments commonly used in observing brain activity in the form of functional images resulting from the measurement and observation of brain electrical activity can be done using several instruments, including functional Magnetic Resonance Imaging (fMRI), Magneto-Enchepalography (MEG), and Electroencephalography (EEG) (Aji & Tjandrasa, 2018).

EEG is a tool or instrument used in electrophysiological monitoring to capture and record electrical activity in the brain. EEG is a recording technique that records changes in the brain's electrical activity by placing electrodes on the scalp (Laksono et al., 2019). EEG is the least expensive instrumentation and is widely used as a recording tool for brain activity. It is non-invasive, which means it does not cause damage to human organs (Aji & Tjandrasa, 2018; Laksono et al., 2019). The EEG instrument used in this study is EEG KT-88 with the specification of the number of channels: electrodes 16 channels EEG + 2 channels ECG (optional ECG); sampling rate: 100 dots/s; accuracy: 12 bits; input impedance: ≥10MΩ; Patient leakage current: < 10μA; noise level : ≤ 5μVp-P ; CMRR : ≥ 90dB ; magnification of multiples: 10000 ; Filter Display Speed: 5, 10, 15, 30, 60, 120 mm/s; amplitude : 1. 1.5, 2, 3, 5, 7.5, 10, 12, 15, 20, 30, 50 mm/50μV ; Rotation speed: 1 time, 2 times, 3 times, 10 times, 20 times, 40 times, 60 times; interference suppression of 50 Hz : ≥ 30 dB. Brain signals produced by

EEG represent quantities that are usually produced in signal frequency analysis, namely amplitude and time (Anggara & Rahayu, 2020; Dea & Djamal, 2015).

The placement of electrodes on the scalp generally uses an international rule called montage method 10-20 which is stated by dividing the head into several areas, namely C (Central), F (Forntal), O (Occipital), P (Parietal), and T (Temporal) (Herdiansyah et al., 2017). The electrical activity obtained from the recording results is due to the electrical signals fired from neurons to neurons in the brain (Aji & Tjandrasa, 2018). Changes in the electrical potential of the brain can be utilized by conducting certain analyses using various signal processing methods, which are grouped into various wave categories, including delta waves (0 - 4 Hz), theta (4 - 7 Hz), alpha (8 - 12 Hz) and beta (12 - 30 Hz) (Nacy et al., 2016). The results of this wave analysis can be used to predict a person's condition and help diagnose several health problems, showcasing the potential of EEG in healthcare applications.

Recording using EEG is not necessarily an easy thing to do because there are challenges that must be faced, namely the signal from EEG recording is susceptible to noise and incoming artifacts, so the filtering process is important before data processing is carried out (Dea & Djamal, 2015). The results of recording using EEG are raw data that is still difficult to observe. Therefore, in addition to the filtering process used to reduce noise and artifacts, the analysis process in the form of clustering and classification is a process that should be carried out. This is done to make it easier to observe a person's activities and condition by using signal activity observation data from EEG recordings.

Previous research on the topic of brain signal classification was conducted by Ekayama et al. (2016) by identifying into two conditions, namely relaxed and non-relaxed; the research was conducted on the FP1 channel, the results of the study in 200 training data and 100 test data resulted in a percentage of recognized training data of 87.5% and the percentage of test data recognized by 47%. Research by classifying brain waves was also carried out by Hilmi et al. (2017), which analyzed alpha and beta signals to classify relaxation conditions in active smokers; the classification methods used in the study were k-Nearest Neighbor (kNN) and Principal Component Analysis with an accuracy of 83.33% on alpha signals and 90% on beta signals. Meanwhile, in this study, classification will be carried out into two conditions, namely normal and relaxed conditions, using the clustering method with k-Means and the classification method with kNN.

Methods

Research Methodology

The research method approach used in this study is a quantitative approach with experiments. Data collection was carried out in the Modern Physics Laboratory Room of Universitas Islam Negeri Walisongo Semarang, a renowned institution for its advanced neuroscience research facilities, on 20 participating students to record brain signal data using Electroencephalography (EEG) instruments directly. The experiment was carried out by collecting brain activity data using EEG in conditions without treatment and treatment conditions with murottal Al-Qur'an so as to provide a stimulus for relaxation to the participants.

Research Procedure

The data analysis method used in this study consists of several stages, namely data acquisition using the Contec KT-88 EEG instrument. The data preprocessing stage uses the Independent Component Analysis (ICA) and Power Spectral Density (PSD) methods to obtain delta, theta, alpha, and beta wave data. After the data preprocessing is successfully carried out, the next stage is the data clustering stage using the k-means algorithm; this data clustering stage aims to provide data labels according to the results of the cluster class obtained using the k-means algorithm. The next stage is the data classification stage using the k-Nearest Neighbor (kNN) algorithm, and then the calculation of the accuracy value of the kNN algorithm is continued. The chart of the data analysis method is shown in Figure 1.

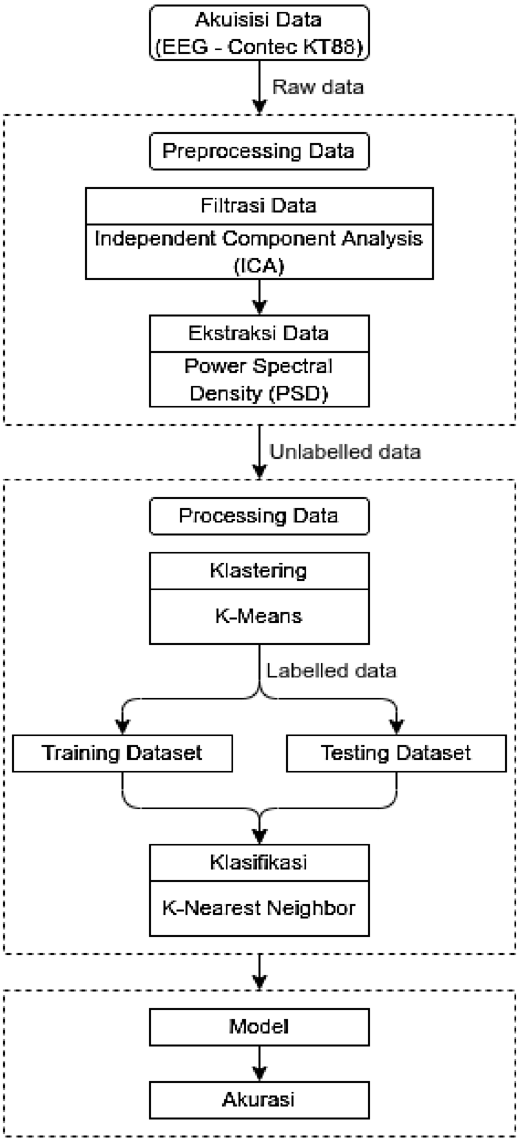


Figure 1. Data analysis flow

a) Data acquisition

Data recording on 20 participating students was carried out in two stages, each producing EEG data for 60 seconds. The first data recording stage was carried out under normal conditions without treatment. Meanwhile, in the second stage, treatment is given by listening to the chanting of the holy verses of the Qur'an with the specifics of surah al-Fatihah verses 1 to 7, which are chanted repeatedly for 60 seconds. The treatment of chanting the holy verses of the Qur'an was carried out with the aim of providing a relaxing stimulus for each participating student. The data obtained amounted to 20 EEG data for normal conditions and 20 EEG data for treatment conditions, with each data lasting 60 seconds. The EEG data obtained is time-based raw data with the European Data Format (EDF) extension, an extension of the EEG data recording standard.

b) Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is a calculation method that utilizes statistics to find hidden factors in random and multivariate variables (Apriyanti, 2020; Riwinoto, 2014). ICA is a signal processing method that can search for components that form mixed signals (Firstanto et al., 2013). ICA separates mixed signals originating from multiple sources with non-Gaussian properties. The purpose of the ICA method is to clean the wave data against noise interference and incoming artifacts so that the obtained wave data will be better for use in the data processing and analysis stage.

c) Power Spectral Density (PSD)

The frequency spectrum is a feature that can be extracted based on raw data taken using EEG instruments (Aji & Tjandrasa, 2018). The EEG recording data is time-based. To facilitate analysis, it is necessary to have a method that can change time-based data into frequency-based data. The Power Spectral Density (PSD) algorithm using the Welch method is one of the methods that can be used to convert time-based data into frequency-based data. The Welch method is applied to estimate the power spectrum of a specified time sequence (Ameera et al., 2019). The Welch method divides the input signal into several short segments, and the periodogram calculation is taken based on the imaginary value of the Fast Fourier Transform (Antonisfia & Wiryadinata, 2008; Husain & Aji, 2019). Segment modification is done with the existing window function before the periodogram calculation (Aji & Tjandrasa, 2018; Husain & Aji, 2019). The periodogram obtained will be averaged so that it produces a good spectrum.

d) k-Means

The k-means algorithm is one of the machine learning algorithms included in the category of clustering algorithms with an unsupervised learning paradigm. Algorithms with an unsupervised learning paradigm approach can be used as a data processing method aimed at data that does not have a label. The k-means algorithm is one of the algorithms that can be used to classify or cluster data (Dhuhita, 2015). K-Means is an algorithm with a non-hierarchical method (partition) that divides or partitions existing data (Nur et al., 2017; Sumadikarta & Abeiza, 2016). The k-Means algorithm can divide cluster class groups into two or more cluster class groups with similar characteristics (Harahap, 2019; Sumadikarta & Abeiza, 2016). K-Means assigns cluster groups based on their proximity to centroid points to data. The proximity of the data to the cluster is measured by the proximity of the cluster object on the centroid using Euclidean distance.

e) The k-Nearest Neighbor (kNN)

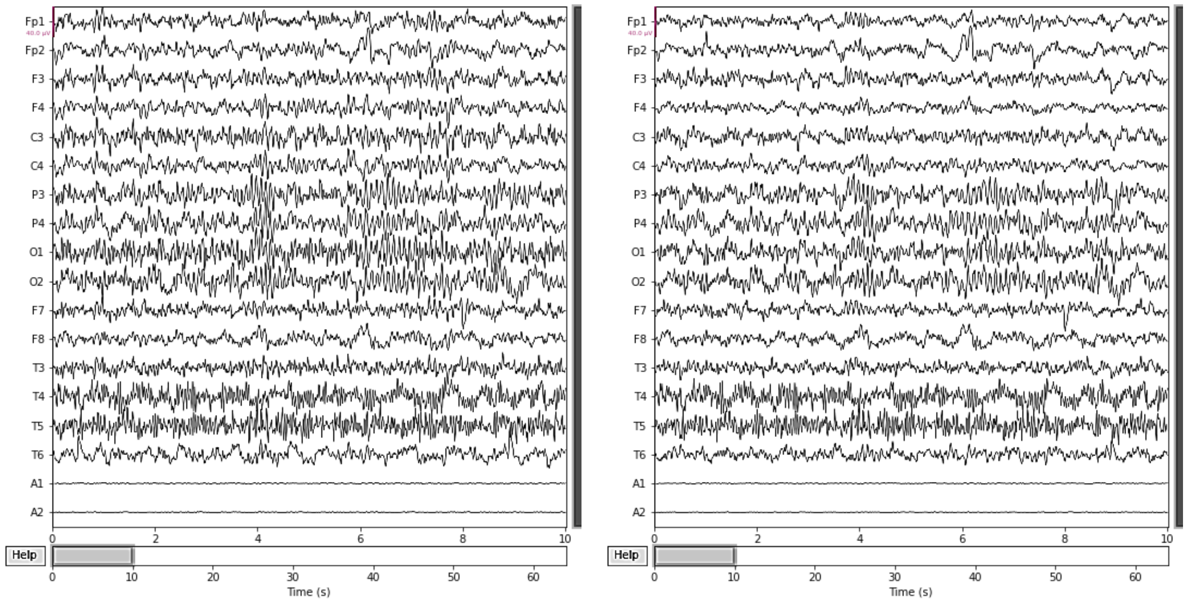
The k-Nearest Neighbor (kNN) is one of the machine learning algorithms with a supervised learning paradigm that performs data processing on data that already has labels. The kNN is a method used to classify data that already has a label. The kNN classifies data based on attributes and samples from training data by using neighborhood classification as the prediction value of the new instance. New instances are classified based on the nearest neighbor (Baharuddin et al., 2019). Meanwhile, to calculate the distance between neighbors, the kNN algorithm calculates the proximity based on the Euclidean distance equation (Baharuddin et al., 2019). The Euclidean distance equation is shown in equation (1) (Yudhana et al., 2020).

$$d(x,y) = \sqrt{\sum_{i=1}^{n} \left(x_{training}^{i} - y_{testing}^{i}\right)^2} \quad (1)$$

where $\mathrm{d(x,y)}$is distance, $\mathrm{x}_{\mathrm{training}}^{\mathrm{i}}$ is data training, $\mathrm{y}_{\mathrm{testing}}^{\mathrm{i}}$ is data testing, i is data variables, and n is data dimensions

Results and Discussions

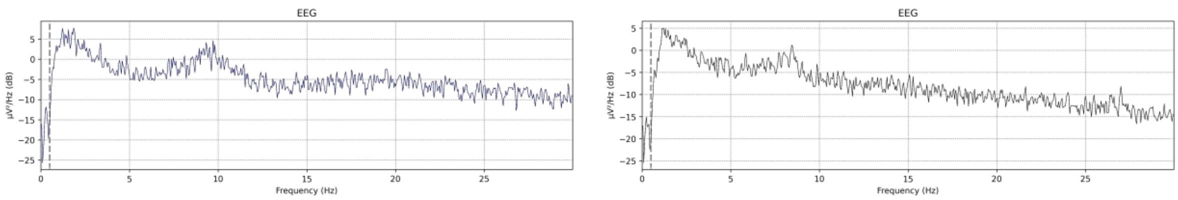
The initial stage in this study is to clean the data (data filtration) using Independent Component Analysis (ICA). Figure 2(a) is the raw data from recording EEG data in participant 1 under normal conditions. Figure 2(a) shows that the raw data obtained during the acquisition process still has noise and artifacts. ICA filters are used on each data to reduce unnecessary interference, such as noise and artifacts. An example of the results of the application of ICA to participant 1's EEG data under normal circumstances can be seen in Figure 2(b), which shows the reduction of noise and artifacts in the data.



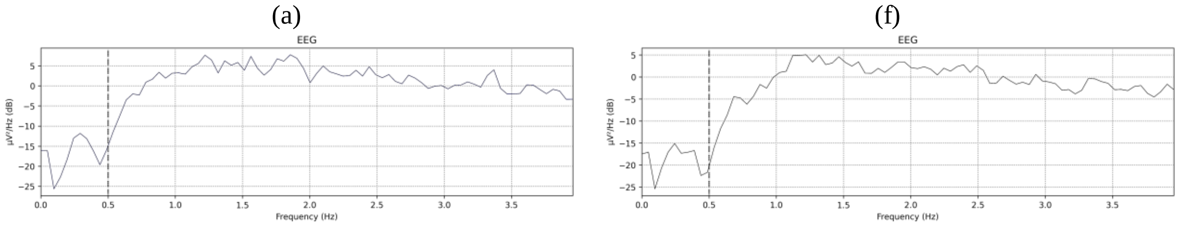
(a) (b)

Figure 2. EEG data on participant 1 (a) Before data filtration with ICA (b) After data filtration with ICA

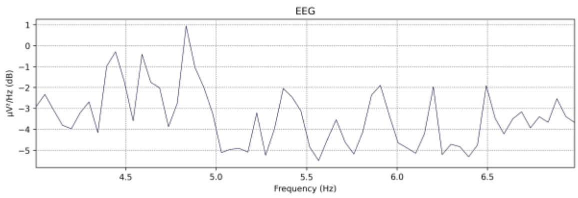
The EEG data obtained from the recording results is time-based raw data. After going through the filtration process, this study's following analysis stage is to change the time-based raw data into frequency-based data. The PSD welch method is one of the methods that can be used to convert time-based data into frequency-based data. This data change aims to facilitate data analysis and determine the characteristics of the human condition based on wave frequency band data, which is divided into several wave frequency ranges, including delta waves (less than 4 Hz), theta (4 – 7 Hz), alpha (8 – 12 Hz) and beta (12 – 30 Hz). The concentration of this wave frequency range can indicate a person's state, whether in a state of sleep, relaxation, concentration, or anxiety. Figures 3(a) and 3(f) show PSD graphs on participant 1's EEG data in a normal and relaxed state. Meanwhile, the division according to the frequency band range is specifically shown in Figures 3(b) and 3(g), which show delta waves in a normal and relaxed state, Figures 3(c) and 3(h), which show theta waves in a normal and relaxed state, Figures 3(d) and 3(i) which show alpha waves in a normal and relaxed state, and Figures 3(e) and 3(j) showing the beta wave in a normal and relaxed state.



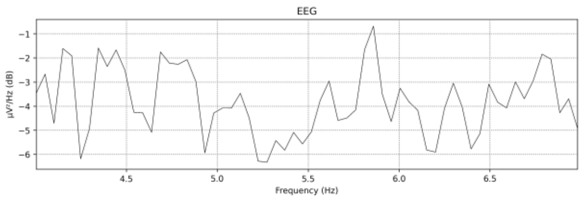
(a) (f)



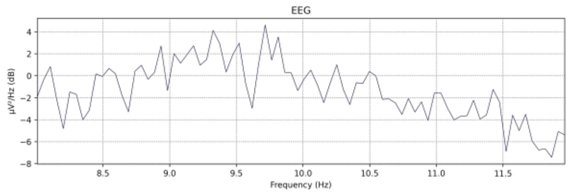
(b) (g)



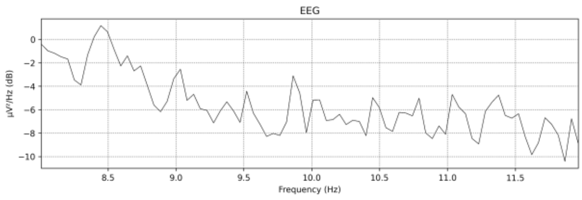
(c)



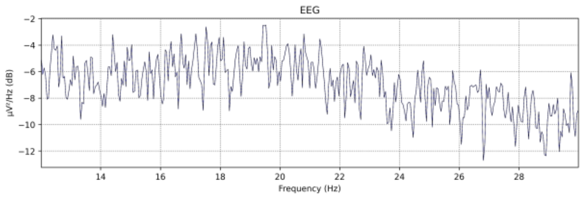
(h)



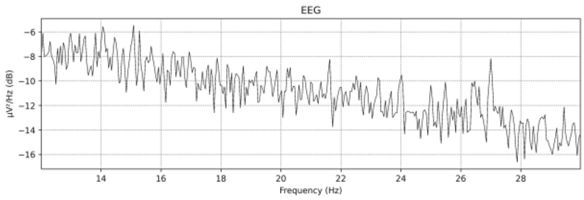
(d)



(i)



(e)



(j)

Figure 3. PSD graphics on participant 1 (normal) (a) PSD 0Hz – 30Hz (b) delta 0Hz – 4Hz (c) tetha 4Hz – 7Hz (d) alpha 8Hz – 12Hz (e) beta 12Hz – 30Hz PSD graph on participant 1 (relax) (f) PSD 0Hz – 30Hz (g) delta 0Hz – 4Hz (h) tetha 4Hz – 7Hz (i) alpha 8Hz – 12Hz (j) beta 12Hz – 30Hz

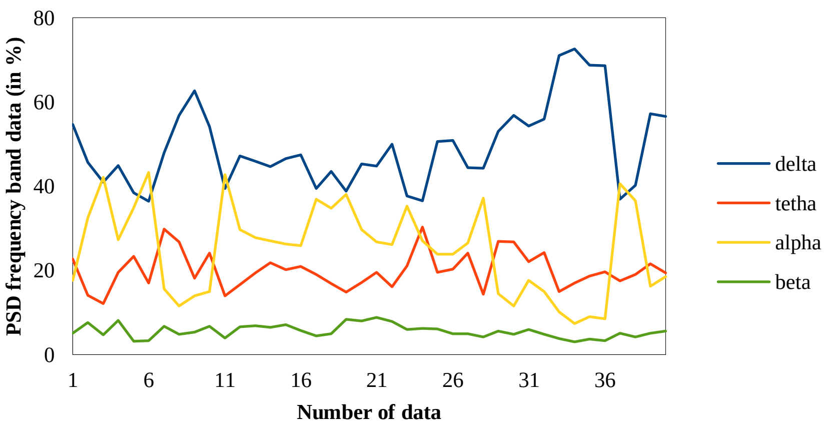
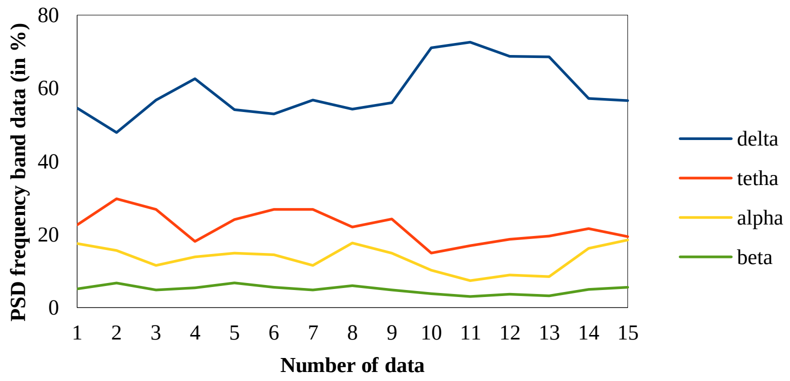


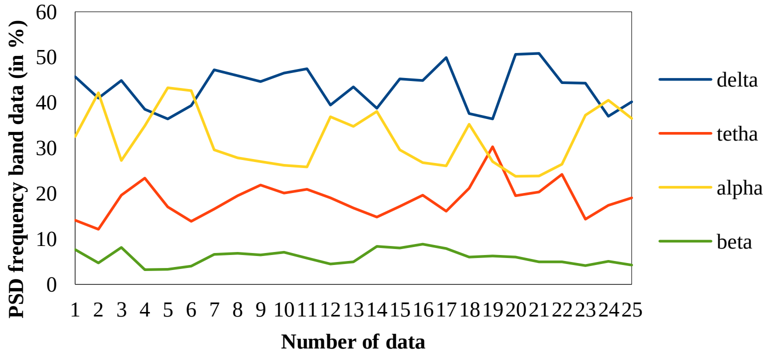
Figure 4. PSD frequency band data (in %)

The change from time-based data to frequency-based data using the PSD Welch method was performed on each participant's EEG data, and the data is produced as shown in Figure 4. In Figure 4, the x-axis represents the number of as many as 40 data, and the y-axis represents the number of percentages in each wave, divided into 4 waves, namely delta, theta, alpha, and beta.

The k-means algorithm in this study was used to cluster the data into two classes indicated by binary numbers 0 and 1. The results of data clustering by the K-Means method on frequency band data are shown in Figure 5. The label value of 0 in Figure 5(a) indicates a person's condition in a normal state with the number of data shown on the x-axis as much as 25 data. In comparison, the label value 1 in Figure 5(b) indicates a more relaxed condition of 15 data. Meanwhile, the y-axis in the graph represents the sum of the percentage magnitudes in each wave.



(a)



(b)

Figure 5. EEG data (a) labeled 0 (normal) (b) labeled 1 (relaxed)

The data division in this study was divided into 80% training data and 20% testing data; 32 data were used as training data, and 8 were used as test data. The classification model generated using the-Nearest Neighbor (kNN) Algorithm produces a precision value of 1.00 or 100% under normal conditions, a precision value of 0.80 or 80% for relaxation conditions, a recall value of 0.75 or 75% for normal conditions, a recall value for relaxation conditions of 1.00 or 100%, and an F1-score value Normal conditions are 0.86 or 86%, F1-Score values in relaxation conditions are 0.89 or 89% and the total accuracy obtained is quite good with a value of 0.88 or 88%.

Table 1. Model accuracy value data

	Precision	Recall	F1-Score
Normal (0)	1	0.75	0.86
Relaxe (1)	0.8	1	0.89
Accuracy			0.88

Conclusion

This research has successfully built a classification model through several stages: data preprocessing, which consists of data filtration using ICA, and data extraction with PSD. Data processing consists of clustering using the k-means algorithm and classification using the k-Nearest Neighbor (kNN) algorithm. The classification process produces two categories of conditions, namely normal and relaxed. The results of the classification model test using the kNN produced an accuracy value of 88%.

Acknowledgments

We want to thank Universitas Islam Negeri Walisongo Semarang for supplying this research's essential facilities and resources.

Conflicts of interest

The authors affirm that they have no conflicts of interest.

References

Aji, N. B., & Tjandrasa, H. (2018). Klasifikasi EEG Epilepsi Menggunakan Singular Spectrum Analysis, Power Spectral Density dan Convolutional Neural Network. *Jurnal Ilmiah Teknologi Informasi*, 185–194. https://doi.org/dx.doi.org/10.12962/j24068535.v15i2.a662

Ameera, A., Saidatul, A., & Ibrahim, Z. (2019). Analysis of EEG Spectrum Bands Using Power Spectral Density for Pleasure and Displeasure State. *IOP Conference Series: Materials Science and Engineering*, *557*(1). https://doi.org/10.1088/1757-899X/557/1/012030

Anggara, R., & Rahayu, Y. (2020). Sistem Electroencephalogram (EEG) Untuk Analisis Sinyal Gelombang Otak Pada Pasien Depresi. *Jom FTEKNIK*, *7*(1), 1–6.

Antonisfia, Y., & Wiryadinata, R. (2008). Ekstraksi Ciri pada Isyarat Suara Jantung Menggunakan Power Spectral Density Berbasis Metode Welch. *Media Informatika*, *6*(1), 71–84.

Apriyanti M., V. (2020). *Klasifikasi Tingkat Stres Manusia Berdasarkan Pengamatan Gelombang Otak Menggunakan EEG dengan Algoritma K-Nearest Neighbor*. Universitas Sumatera Utara.

Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, *11*(3), 269–274. https://doi.org/10.33096/ilkom.v11i3.489.269-274

Berbano, A. E. U., Pengson, H. N. V., Razon, C. G. V., Tungcul, K. C. G., & Prado, S. V.(2017). Classification of Stress Into Emotional, Mental, Physical and No Stress Using Electroencephalogram Signal Analysis. *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*.11–14.

Dea, A. P. D., & Djamal, E. C. (2015). Klasifikasi Sinyal Elektroensephalogram terhadap Kewaspadaan Menggunakan Support Vector Machine. *Prosiding SNIJA*, 242–246.

Dewi, D., A., P., & Djamal, E., C., (2015). Klasifikasi Sinyal Elektroensephalogram terhadap Kewaspadaan Menggunakan Suppport Vector Machine. *Prosiding SNIJA*, 242 -246.

Dhuhita, W. M. P. (2015). Clustering Menggunakan Metode K-Means untuk Menentukan Status Gizi Balita. *Jurnal Informatika*, *15*(2), 160–174.

Ekayama, R., Djamal E., C., & Komarudin, A., (2016). Identifikasi Kondisi Rileks dari Sinyal EEG Menggunakan Wavelet dan Learning Vector Quantizartion. *Prosiding SNST*, 150-155.

Firstanto, M. A., Wirawan, & Widjiati, E. (2013). Kombinasi Metode Independent Component Analysis (ICA) dengan Beamforming untuk Pemisahan Sinyal Akustik Bawah Air. *JURNAL TEKNIK POMITS*, *2*(2), 300–305. https://doi.org/doaj.org/toc/2337-3539

Harahap, B. (2019). Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung). *Regional Development Industry & Healt Science, Technology and Art of Life*, *2*(1), 394–403.

Herdiansyah, I., Djamal, E. C., & Komarudin, A. (2017). Klasifikasi Sinyal EEG Terhadap Tiga Kondisi Pikiran Menggunakan Autoregressive dan Adaptive Backpropagation. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 47–51.

Hilmi, A., Wijayanto, I., & Hadiyoso, S. (2017). Analisis Perbandingan Pola Sinyal Alfa dan Beta Untuk Klasifikasi Kondisi Rileks pada Perokok Aktif dengan Menggunakan K-Nearest Neighbor. Universitas Telkom, Bandung.

Husain, N. P., & Aji, N. B. (2019). Klasifikasi Sinyal EEG Dengan Power Spectra Density Berbasis Metode Welch Dan MLP Backpropagation. *Jurnal ELTIKOM*, *3*(1), 17–25. https://doi.org/10.31961/eltikom.v3i1.99

Laksono, I. K., Matul, E., & Kom, M. (2019). Klasifikasi Schizophrenia Berdasarkan Sinyal EEG Menggunakan Algoritma Support Vector Machine. *Jurnal Ilmiah Matematika*, *7*(2).

Nacy, S. M., Kbah, S. N., Jafer, H. A., & Al-Shaalan, I. (2016). Controlling a Servo Motor Using EEG Signals from the Primary Motor Cortex. *American Journal of Biomedical Engineering*, *6*(5), 139–146. https://doi.org/10.5923/j.ajbe.20160605.02

Niedermeyer E, L. da S. F. (1993). Electroencephalography: Basic Principles, Clinical Applications and Related Fields. In Journal of Neurology, Neurosurgery & Psychiatry (Vol. 47, Issue 6).

Nur, F., Zarlis, M., & Nasution, B. B. (2017). Penerapan Algoritma K-Means pada Siswa Baru Sekolah Menengah Kejuruan untuk Clustering Jurusan. *Jurnal Nasional Informatika dan Teknologi Informasi*, *1*(2), 100–105. https://doi.org/doi.org/10.30743/infotekjar.v1i2.70

Riwinoto. (2014). Penemuan Sinyal Asli dengan Metode Independent Component Analysis pada Sinyal Tercampur Tunggal. *Jurnal Integrasi*, *6*(1), 77–83.

Sumadikarta, I., & Abeiza, E. (2016). Penerapan Algoritma K-Means pada Data Mining untuk Memilih Produk dan Pelanggan Potensial (Studi Kasus: PT Mega Arvia Utama). *Jurnal Satya Informatika*, *1*(1), 12–22.

Yudhana, A., Sunardi, & Hartanta, A. J. (2020). Algoritma K-NN dengan Euclidean Distance untuk Prediksi Hasil Penggergajian Kayu Sengon. *TRANSMISI*, *22*(4), 123–129. https://doi.org/10.14710/transmisi.22.4.107-141