Signal and Image Processing Letters
Vol. 1., No. 3, November 2019, pp. 32-42
ISSN 2714-6677
https://simple.ascee.org/

Classification of concentration or focus by signal Electroencephalography (EEG) and addiction Watching K-Dramas Using Algoritma K-Nearest Neighbor

Ahmad Azhari [a,1,*], Rizky Ramadan [a,2]

[a] Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[1] ahmad.azhari@tif.uad.ac.id; [2] rizky1800018126@webmail.uad.ac.id
* corresponding author

ARTICLE INFO

Keywords
EEG
Fast Fourier Transform
K-drama
K-Nearest Neighbor
Concentration

ABSTRACT

K-drama or drakor is currently being enjoyed in Indonesia when the Covid-19 pandemic hits, especially by the fair sex. From the sources obtained, the number of k-dramas or dramas also increased during the covid-19 pandemic from the previous 2.7 hours a day to 4.6 hours a day. The issue raised by the authors in this study is whether the impressions of drakor will later affect the concentration of an individual. Data acquisition was carried out using the NeuroSky Mindwave Mobile 2 tool to retrieve EEG data. After the data acquisition is completed, the next process is preprocessing, which is to perform feature extraction using the Fast Furious Transform method to find the average values of the highest and lowest peaks. After the preprocessing is completed go to the classification stage. The classification used is K-Nearest Neighbor with a value of k=9. For evaluation using confusion matrix to determine the accuracy value of the built KNN model. This study used 100 respondents who were37 people who were addicted to drakor. A total of 24 people out of the 37 or about 64.87% turned out to have a lack of concentration level when taking concentration tests. This is enough to prove that drama impressions can reduce the concentration or focus of a person, especially women. For the classification process to have an accuracy of 80% and for variable correlation testing, it turns out that independent variables do have a simultaneous effect on the dependent variables with a calculated f value of 35.642 and a sig value of 0.000b.

1. Introduction

K-drama or drakor is currently being enjoyed in Indonesia, especially by the fair sex. One of the reasons why drama is widely enjoyed in Indonesia today is due to the enactment of PPKM and work from home (WFH), so many people spend time at home and one of the things to do is to watch. From the sources obtained by the author, watching or marathon films ranked second after cooking as the most frequent activities carried out when PPKM and work from home (WFH) were enforced. The duration of watching k-dramas or dramas also increased after the enactment of PPKM due to covid-19 which was previously 2.7 hours a day to 4.6 hours a day [1][2].

One of the factors that is expected to affect watching drakor shows is concentration. Concentration itself is the ability of the individual to focus the attention of each individual on a certain object of activity, if without concentration then the work will be delayed and not good otherwise if we concentrate then the work, we do will be good and correct [3].

In addition to the level of addiction to watching drakor, another factor that will be used as a benchmark is the signal electroencephalography (EEG). Electroencephalography (EEG) is a technique for measuring electrical activity on the scalp produced by brain structures taken with metal electrodes and conductive media. The EEG was recorded on the animal's brain in 1875 by Richard Karton. Then in 1924 for the first time EEG was introduced by Hans Berger to record signals in the human brain (D. Millett, 2001). EEG is the most widely used signal acquisition method because it has a high temporal resolution, a relatively easy and safe technique that communication that occurs in human brain cells occurs through electrical impulses. It is measured by placing electrodes on the subject's scalp. EEG is commonly used to diagnose, monitor, and analyze various abnormalities in the body, especially in the brain.

This electroencephalography (EEG) data collection uses a tool called NeuroSky Mindwave Mobile 2. The tool has 2 sensors, namely on the forehead and on the auricle which will later be attached to the respondent by being given a certain stimulus. The output of the tool is an excel file with a .csv extension. The content of the output is the result of recording electrical activity in the brain which is recorded every few seconds. The columns contained in the results of recording electrical activities are 'Signal Level', 'Blink', 'Attention', 'Meditation', 'Zone', 'Delta', 'Theta', 'Low Alpha', 'High Alpha', 'Low Beta', 'High Beta', 'Low Gamma', 'Mid Gamma'. While the column that will be used is only in the attention column.

The attention data obtained cannot be used immediately because it is still raw data consisting of several data. In this case it must first be distraction. This is done to determine the characteristics of the data to be extracted. Feature extraction using Fast Fourier Transform (FFT). Fast Fourier Transform (FFT) is the source of an algorithm for calculating fast, efficient, and reverse discrete fourier transformations [4]. Then according to Dine Tiara Kusuma [4] Fast Fourier Transform (FFT) is an algorithm to calculate discrete fourier transformations quickly and efficiently by performing mathematical calculations used to transform analog signals into frequency-based digital signals.

The next step is to find a correlation between the independent variable and the dependent variable. Whether independent variables do have a simultaneous effect on dependent variables or not. For correlation testing between variables using Test F (Anova).

After the process of searching for the highest and lowest peaks of attention is complete. The next process is classification. In this study the classification method used is K-Nearest Neighbor (KNN). KNN is a classification algorithm that works by taking a number of K closest data (neighbors) as a benchmark for determining the label of new data. This algorithm classifies data based on their neighbors or proximity to other data. How to find out the closest data using the distance calculation method. In this study the distance calculation method that will be used is Euclidean Distance. The reason for using the K-Nearest Neighbor algorithm is because the method is suitable for managing data with a lot of scalability, according to the number of respondents that will be used in this study, namely as many as 100 respondents. In addition, for researchers themselves the method is easy to implement or easy to use.

After classifying, of course, it is necessary to evaluate the classification model that is built. It aims to find out what is the accuracy level of the classification model being built. In this study, the evaluation of the model used was a confusion matrix.

This research is expected to be able to identify whether k-drama or drakor impressions can affect a person's concentration, especially for people who feel they are addicted to drakor. For the classification process using K-Nearest Neighbor, it is expected to be able to classify with good accuracy values and independent variables do have a simultaneous effect on dependent variables.

2. Literature Review

2.1. Brainwaves

Brainwave signals have different characteristics and individual characteristics, brain waves cannot be imitated or read by people so it is impossible to have something in common. The recognition of identity is necessary to distinguish individual characteristics. Identity recognition aims to recognize a person's identity precisely [5].

Table 1. Wave Oscillation Table

Signal	Frequency (Hz)	Condition
Delta	0.5 - 4	The condition of deep sleep without dreaming (Deep Sleep). The resting phase for the body.
Theta	4 - 8	Conditions of meditation, intuition, and fantasy
Alpha	8 - 13	Conditions of relaxation and awake
Beta	13 - 30	A state of focus, thinking, and concentration. Productivity phases for
Gamma	>30	States of high mental activity such as fear, panic, high concentration

2.2. Electroencephalography (EEG)

Electroencephalography (EEG) is a technique for measuring the electrical activity of the scalp produced by brain structures, taken by attaching metal electrodes and conductive media. The EEG was recorded on the animal's brain in 1875 by Richard Karton. Then in 1924 for the first time EEG was introduced by Hans Berger to record signals in the human brain (D. Millett, 2001). Then according to Vladimir A. et al. EEG is a method for conducting brain research that provides information about brain functions related to human activity [6].

2.3. Feature Extraction

Feature extraction is used to obtain patterns from the tested signals. Feature extraction is the initial process to carry out the process of classification and interpretation of signals. Feature extraction is performed to obtain the characteristics or traits of the signal to be tested [7][8][9].

In this study, the extraction of the feature used was Fast Fourier Transform (FFT). Fast Fourier Transform (FFT) is a highly efficient method of calculating the coefficient from a discrete Fourier to a finite sequence of complex data. According to other sources, [10]Fast Fourier Transform (FFT) is a source algorithm for calculating fast, efficient, and reverse discrete Fourier transformations. Then according to Dine Tiara Kusuma, [4] Fast Fourier Transform (FFT) is an algorithm to calculate discrete fourier transformations quickly and efficiently by performing mathematical calculations used to transform analog signals into frequency-based digital signals.

2.4. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a method that uses a supervised algorithm where the results of a new instance query are classified based on the majority of the categories in KNN. When using the K-Nearest Neighbor (KNN) algorithm, we need to specify the value of k to be used, the value of k will later become the closest neighbor of the new data entered and determined its class.

In the KNN method there are many methods used to calculate the distance between data. In this study, the method that will be used to calculate the distance is Euclidean distance.

3. Method

The flow of research can be seen in Fig. 1.

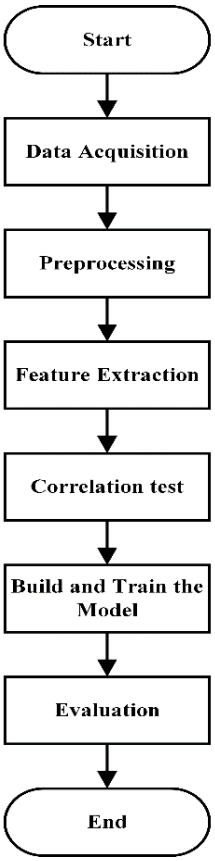

Fig. 1. Research flow

3.1. Data Acquisition

The tool used for data acquisition in this study is NeuroSky Mindwave Mobile 2 to record how signal patterns will be generated by the brain when concentrating. During the sampling of signal patterns, respondents will be given predetermined stimuli.

3.2. Preprocessing

Preprocessing is the initial stage after obtaining raw data before entering the next stage of processing. In this study, the preprocessing used was data filtering, which is choosing what data will be used in the study. The data consists of 'Signal Level', 'Blink', 'Attention', 'Meditation', 'Zone', 'Delta', 'Theta', 'Low Alpha', 'High Alpha', 'Low Beta', 'High Beta', 'Low Gamma', 'Mid Gamma'. The data that will be used later is only attention data because the data in the column refers to the state of the brain that is in a state of focus or concentration.

3.3. Feature Extraction

Once the data is acquired dan in the preprocessing the next stage is feature extraction. This process is carried out to manage the raw data that has been acquired. The data obtained will be normalized by finding the minimum peak value and the maximum peak value in order to get the desired numerical data.

The extraction process uses the Fast Fourier Transform (FFT) method. Data that is still a signal image will be extracted first using Fast Fourier Transform (FFT). Fast Fourier Transform (FFT) itself is widely applied in various fields, ranging from digital signal processing and solving partial differential equations to algorithms for multiplying integers by large quantities. Fast Fourier Transform (FFT) is an algorithm for calculating discrete Fourier transformations (DFT) quickly and efficiently. Since many signals in the communication system are continuous in the time domain, we use Fourier transforms to represent the frequency domain [11].

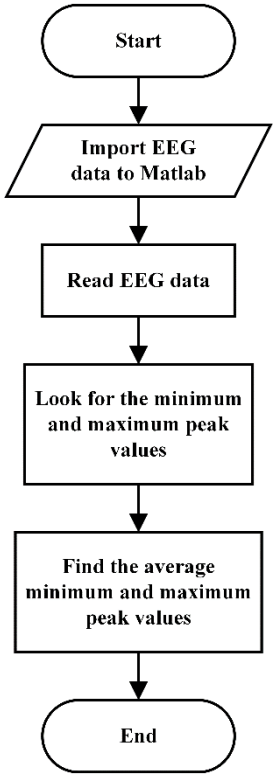

Fig. 2. Feature Extraction Flow

For the flow of the feature extraction process can be seen in the Fig. 2. The steps are as follows:

The extraction process uses Matlab. The raw data acquired is imported into Matlab.

1. After importing, read the data so that it can be processed.
2. Looks for minimum and maximum peak values.
3. Looks up the average value of each minimum and maximum peak by 2 times.
4. Looks for the final mean between the minimum and maximum peaks.

3.4. Correlation test

The correlation test aims to determine the presence or absence of the influence of independent variables on dependent variables. In this study, the correlation test used the F test on SPSS. The purpose of the test is to find out the correlation between the EEG signal (attention) and the K-addict score as an independent variable against a person's concentration as a dependent variable simultaneously (together).

The provisions of the F test are as follows:

- If the sig value > 0.05, then the conclusion is that there is no influence of independent variables on the dependent variables simultaneously (together). Vice versa, if the sig value < 0.05, then there is an influence of independent variables on the dependent variables simultaneously (together).
- If the F value is calculated < F of the table, then the conclusion is that there is no influence of the independent variable on the dependent variable simultaneously (together). Vice versa, if the F value is calculated > F table, then there is an influence of the independent variable on the dependent variable simultaneously (together).

3.5. Build and Train the Model

In this study, the model to rebuilt is the K-Nearest Neighbor (KNN) model. Before going into the classification stage should be carried out model development first. This is done so that the data used

is not overfitting, which is a training condition in which the test results against the trained data are very good but tested by other data that are not used in the training very poorly. Model development is carried out by dividing all data into two parts, namely training data and test data. In the development of a model the training data should be larger than the test data. In this study, a comparison between training data and test data of 80:20 will be used.

For the steps of the K-Nearest Neighbor algorithm can be seen in Fig. 3.

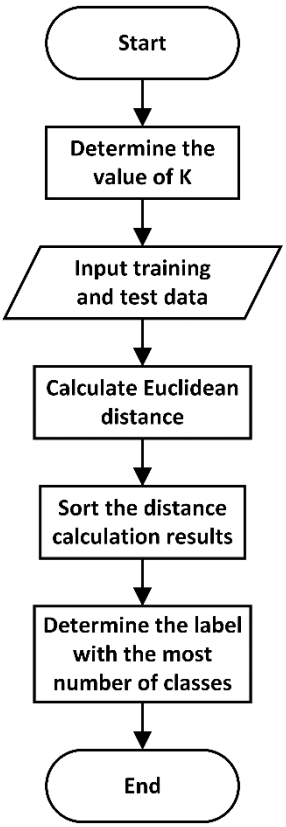

Fig. 3. K-Nearest Neighbor Flow

K-Nearest Neighbor Flow:

1. Determine the value of K. In this study the value of K used is 9.
2. Calculate the distance between the training data and test data using the Euclidean distance method.
3. Ascending sort distance.
4. Determine the test data labels based on the most labels from the K nearest neighbors.

3.6. Evaluation

The last step is to test the accuracy level of the classification method used (KNN) for self-testing using the Confusion Matrix.

Evaluation steps using Confusion Matrix:

1. Know the amount of data used. In this study, 100 data will be used.
2. Determine the True Positive (TP) value
3. Determine the True Negative (TN) value
4. Determine the False Positive (FP) value
5. Determine the False Negative (FN) value
6. Calculating accuracy value: (TP + TN) / (TP + FP + FN + TN)

4. Results and Discussion

4.1. Data Acquisition

Data acquisition scenarios:

1. Respondents filled out a questionnaire to complete their personal data and determine their k-drama addict score.
2. After filling out the questionnaire, respondents were asked to watch Korean drama shows with a minimum duration of 7 minutes until they were completed according to respondents' availability.
3. Installation of the tool Neurosky Mindwave mobile 2. The recording process is carried out while the respondent takes a concentration test given for approximately 2 minutes.
4. The final step of the respondent will be asked to take a concentration test. The test used is an online focus (concentration) test
5. The output on this test is 3, which are less, sufficient, and good. This output will later be used as a label on the dataset. For sufficient output it is still categorized into good concentrations. So that label used are only two i.e. good and less.
6. The recording results during the time the respondent took the test will be saved with an extension (*. log).

EEG data was obtained from the results of recording brainwave signals in 100 respondents using the EEG tool and MyndPlayer Pro software. The process of recording brainwave signals to respondents was carried out within ± 2 minutes. The successfully recorded EEG data will be saved in a file with the format (*.log).

Table 2. Sample EEG Data from Respondent 1

Time	Vid Time	Signal Level	Blink	Attention	Meditation	Zone	Delta	Theta	Low Alpha	High Alpha	Low Beta	High Beta	Low Gamma	Mid Gamma
22:37:53.000	07:00:00.000	100	2	13	44	28	127646	112811	1734	2225	1782	609	312	578
22:37:54.000	07:00:00.000	100	2	37	67	52	682452	333635	2586	7088	1605	4255	1361	3728
22:37:55.000	07:00:00.000	100	2	54	70	62	23923	180116	11261	6045	8801	7306	3786	8308
22:37:56.000	07:00:00.000	100	2	50	61	55	618498	341240	56414	8175	9894	7297	3430	9055
22:37:57.000	07:00:00.000	100	2	35	53	44	3042995	789520	33542	65521	11301	7922	5977	12487
22:37:58.000	07:00:00.000	100	2	11	41	26	1497698	823280	61016	12718	26551	8759	6291	5446
22:37:59.000	07:00:00.000	100	2	1	38	19	72607	955545	26534	5745	1074	2210	2558	2405
22:38:00.000	07:00:00.000	100	2	1	34	17	788904	199313	25069	5182	25832	10988	5306	13626
22:38:01.000	07:00:00.000	100	2	14	51	32	614445	476978	23069	16135	1826	5655	2453	1080
22:38:02.000	07:00:00.000	100	2	13	40	26	173509	234366	2381	1191	6189	2208	2147	3412

4.2. Prepocessing

At this stage, the EEG recording data will be selected first to choose what attributes to look for maximum and minimum peak values can be seen in the Fig. 4.

Preprocessing Flow using Matlab:

1. Add a raw signal as an input to be converted into spectrum s.
2. Selection of attributes from the EEG data to be searched for maximum and minimum values. In this study, the selected attribute was the attention attribute.

3. Get the maximum and minimum values of the peak of the signal.

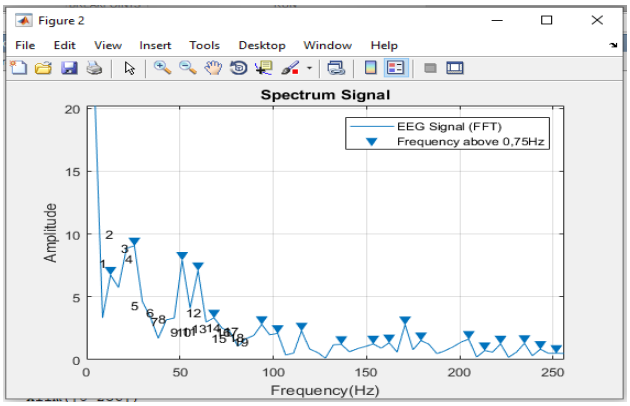

Fig. 4. Sinyal Spectrum

4.3. Feature Extraction

For feature extraction using Fast Fourier Transform with Microsoft Excel software. The extraction of features using Fast Fourier Transform aims to find the average of the minimum and maximum peaks.

Flow extraction features:

1. Sort the minimum and maximum peak data obtained starting from the largest (descending).

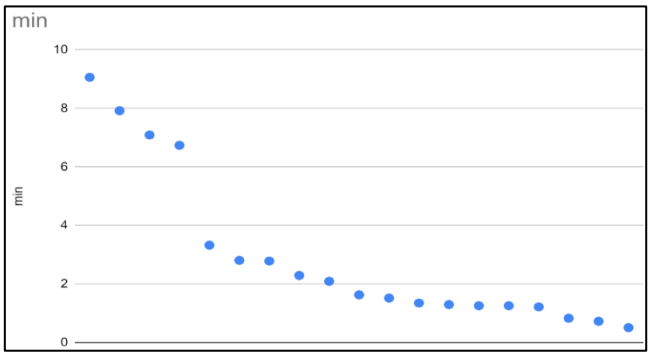

Fig. 5. Minimum Data Order by Descending

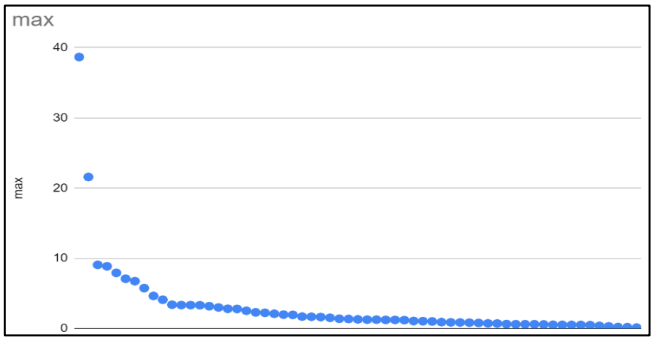

Fig. 6. Descending Maximum Data Order

2. Looks up the average value of each minimum and maximum peak by 2 times.
3. Looks for the final mean between the minimum and maximum peaks.

The results of the final mean between the minimum and maximum peaks will later be used as one of the attributes (attention) in the research dataset.

4.4. Correlation Test

The correlation test was performed using the f test on SPSS. The result of the test is F count > F table, which is 35.642 > 3.0872959 then there is an influence of independent variables on the dependent variables simultaneously (together). For sig values of 0.000b < 0.05, there is an influence of independent variables on the dependent variables simultaneously (together).

4.5. Build and Train the Model

Based on the K-Nearest Neighbor model that was built, the results of determining classes or labels for 20 test data were obtained by calculating the Euclidean distance with a value of k = 9, the results are obtained in Table 3 and Fig. 7.

Table 3. Concentration Test Results

Respondent	Concentration
Subject 1	Good
Subject 2	Good
Subject 3	Not Good
Subject 4	Good
Subject 5	Good
Subject 6	Good
Subject 7	Not Good
Subject 8	Good
Subject 9	Not Good
Subject 10	Not Good
Subject 11	Good
Subject 12	Good
Subject 13	Good
Subject 14	Not Good
Subject 15	Not Good
Subject 16	Not Good
Subject 17	Not Good
Subject 18	Not Good
Subject 19	Not Good
Subject 20	Good

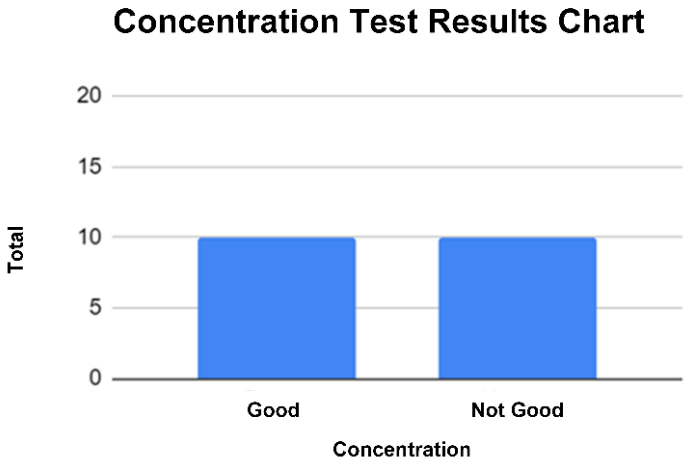

Fig. 7. Concentration Test Results

4.6. Evaluation

The evaluation of the model used is the Confusion Matrix. For evaluation data using 20 random or random data. Here's the matrix being revamped in Table 4.

Table 4. Confusion Matrix

		ACTL
PREDICTED	10 (TP)	2 (FP)
	2 (FN)	6 (TN)

- True Positive (TP) = 10
- False Positive (FP) = 2
- False Negative (FN) = 2
- True Negative (TN) = 6

Based on the matrix above, the accuracy rate of the K-Nearest Neighbor l mode built is 80%.

5. Conclusion

Based on the research conducted, it consisted of 100 respondents with 63 people who assumed that they were not addicted to drakor impressions and 37 people who assumed they were addicted to drakor impressions. Of the 37 people, there were 34 women and 3 men. A total of 24 people out of 37 people who assumed they were addicted to drakor impressions or about 64.87% turned out to have a lack of concentration level when taking concentration tests. Of the 24 people, there were 23 women and 1 man. This is enough to prove that drama impressions can reduce the concentration or focus of a person, especially women.

For classification using the K-Nearest Neighbor algorithm, it can be said to be successful because it is able to do classification well. The accuracy rate of the built K-Nearest Neighbor model is 80%.

Correlation testing using the f (anova) test between independent variables and dependent variables has also been proven by the results of calculated F values (35.642) > F tables (3.0872959) and sig values (0.000b) < 0.05. It can therefore be concluded that independent variables have a simultaneous effect on dependent variables.

K-drama or drama lovers are expected to be able to control the feeling of wanting to watch. Don't every free time be spent watching drakor. It is better to fill it with other productive activities such as sports or studying.

References

[1] I. F. Wati, “Digital game-based learning as a solution to fun learning challenges during the Covid-19 pandemic,” In *1st International Conference on Information Technology and Education (ICITE 2020)*, pp. 202-210, 2020.

[2] V. K. R. Chimmula and L. Zhang, “Time series forecasting of COVID-19 transmission in Canada using LSTM networks,” *Chaos, Solitons & Fractals*, vol. 135, p. 109864, 2020.

[3] P. D. Purnamasari, P. Yustiana, A. A. Putri Ratna and D. Sudiana, "Mobile EEG Based Drowsiness Detection using K-Nearest Neighbor," *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, pp. 1-5, 2019,

[4] W. K. Lee, M. M. Ratnam, and Z. A. Ahmad, ”Detection of chipping in ceramic cutting inserts from workpiece profile during turning using fast Fourier transform (FFT) and continuous wavelet transform (CWT),” *Precision Engineering*, vol. 47, pp. 406-423, 2017.

[5] A. Azhari, and L. Hernandez, “Brainwaves feature classification by applying K-Means clustering using single-sensor EEG,” *International Journal of Advances in Intelligent Informatics*, vol. 2, no. 3, pp. 167-173, 2016.

[6] V. A. Maksimenko, *et al.,* "Artificial neural network classification of motor-related eeg: An increase in classification accuracy by reducing signal complexity," *Complexity*, *2018.*

[7] A. K. Jaiswal and H. Banka, "Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals," *Biomedical Signal Processing and Control*, 34, 81-92, 2018.

[8] P. Dawar and S. Dhage, "Feature extraction methods for electroencephalography based brain-computer interface: A review," *IAENG International Journal of Computer Science*, vol. 47, no. 3, 2020.

[9] P. Ozel, A. Akan, and B. Yilmaz, "Synchrosqueezing transform based feature extraction from EEG signals for emotional state prediction," *Biomedical Signal Processing and Control*, vol. 52, 2019.

[10] N. Thammasan, K. -i. Fukui and M. Numao, "Application of deep belief networks in eeg-based dynamic music-emotion recognition," *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 881-888, 2016.

[11] A. R. Hassan, and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowledge-Based Systems*, *128*, 115-124, 2017.