The Effect of Feature Selection on Automatic Sleep Stage Classification Based On Multichannel EEG Signals

Mera Kartika Delimayanti
Department of Computer and Informatics Engineering
Politeknik Negeri Jakarta
Depok, Indonesia
mera.kartika@tik.pnj.ac.id

Mauldy Laya
Department of Computer and Informatics Engineering
Politeknik Negeri Jakarta
Depok, Indonesia
mauldy.laya@tik.pnj.ac.id

Mohammad Reza Faisal
Department of Computer Science
Lambung Mangkurat University
Banjarbaru, Indonesia
reza.faisal@ulm.ac.id

Rizqi Fitri Naryanto
Department of Mechanical Engineering
Universitas Negeri Semarang
Semarang, Indonesia
rizqi_fitri@mail.unnes.ac.id

Kenji Satou
Institute of Science and Engineering
Kanazawa University
Kanazawa, Japan
ken@t.kanazawa-u.ac.jp

***Abstract*— When diagnosing and treating sleep disorders, the manual classification of sleep stages is a time-consuming but crucial step, and the automation of this process has been a focus of recent research. Many kinds of research have been conducted on the automation of sleep stage classification. In this paper, we proposed the effect of feature selection on an automated system based on EEG signals, which was then followed by classification using various supervised classifiers such as Random Forest and SVM. The high dimensional FFT features were used to extract the characteristics of EEG for the classification of sleep stages. The EEG dataset is used from the Sleep-EDF dataset, which is freely available. The accuracy as the performance evaluator on the Random Forest model had gained the best value on 95.93%, 90.41%, 87.91%, 86.92%, and 84.86% and the SVM model reached 96.63%, 91.27%, 88.90%, 87.94%, and 87.94% for 2-6 state classification. Finally, in this proposed research the feature selection phase affects the model's accuracy.**

Keywords—EEG signals, feature selection, Random Forest, Sleep-EDF, Support Vector Machine (SVM).

I. INTRODUCTION

Human beings' most important physiological activity is sleep. Sleep deprivation has been shown to weaken the immune system, putting people's lives and health at risk[1]. Numerous studies have demonstrated that tired driving is causing irreversible implications for an increasing number of accident drivers. Sleep fragmentation and apnea are common in those who have severe sleep-related ailments. When they enter a deeper sleep state, their airways can become obstructed, preventing them from breathing normally. Because of the interference, the body is forced to return to a lighter sleep stage to maintain improved breathing. Sleep apnea patients do not go through the stages of sleep that other people do[2], [3]. As a result, evaluating sleep status can help people better understand their sleeping habits, develop sleep disorder prevention methods, and safeguard their sleep health. Different sleep stages must be categorized to achieve a consistent sleep state throughout the night. Increased accuracy of sleep stage identification is needed in order to better understand sleep-related diseases and disorders [4].

Currently, polysomnography (PSG) is mostly utilized in clinical settings to assess sleep. There are numerous physiological signals that must be recorded by the PSG, such as EOG (electrooculogram), EEG (electroencephalogram), EMG (electromyogram), and ECG (electrocardiogram), as well as pulse oximetry and breathing signals. Experts (electroencephalographers) manually identify sleep stages (scoring) by splitting the full sleep record into 30-second epochs and assigning a specific sleep stage to each epoch[5]. The difference in electric activity between the two electrodes over time is used to determine the electric signal in the brain. The signal steadily decays with distance from the source as it propagates. To sum up, the signal is smaller in value because only one of the parallel combinations of electrodes offers an accurate measurement of the current [6]. The example on the had executed the EEG signal from a single channel on the eye-state characterization[7].

Some certain waves and events characterize a recorded signal, especially the EEG. The sleep stages include the night wake (wake) stage, the REM (Rapid Eye Movement) stage, as well as the NREM (Non-Rapid Eye Movement) stage. According to the American Academy of Sleep Medicine (AASM), the NREM stage can be further divided into N1, N2, and N3 stages. There are two distinct stages of N3 development: R&K (R&K) Stage 3 and R&K Stage 4 (S3) (Rechtschaffen, 1968). One of the most frequently acknowledged standards for measuring sleep stages is the Rechtschaffen and Kales criterion (R & K) and American Academy of Sleep Medicine (AASM) criteria. According to the R & K recommendations, sleep is broken down into seven stages: waking, sleepiness, light sleep, deep sleep, REM, and MT/movement time. The AASM criteria, on the other hand, are based on the R&K criteria. Some of the differences between the AASM and R & K criteria are as follows[4]:

1. When referring to the R & K criteria, the stages S1, S2, S3, and S4 are referred to as stages N1, N2, and N3 when referring to the AASM criteria.
2. Deep sleep (N3), according to the AASM criterion, is a combination of the R & K criteria's S3 and S4 stages, respectively.
3. According to the American Academy of Sleep Medicine criteria, movement time (MT) is no longer recognized as a sleep stage.

The sleep stages were manually evaluated and annotated in the previous version. The outcome is a time-consuming and expensive process that relies heavily on human capital. A human expert cannot annotate large EEG datasets for sleep phases because it takes too much time, money, and effort.[8].

As a result, developing a sleep stage classification has become important in order to improve accuracy.

Automatic sleep stage identification research has primarily employed frequency-domain data, such as the power spectral estimate of EEG waveforms' various frequency sub-bands, to extract features from polysomnographic signals. Neural networks and fuzzy logic algorithms were used in the categorizing process, among other methods [9]. Wavelet packets were employed for feature extraction in other methods and alternative classification algorithms, including rule-based reasoning [10], [11]. Some researchers had conducted the research based on sleep stages classification using feature selection, but the dataset used was different[12]. We executed this research from the Sleep EDF dataset[13].

Using PSG signals, specifically EEG signals, we investigated the effect of feature selection on an automated sleep stage classification system. This research was followed by classification using various supervised classifiers, such as Random Forest and SVM. It is fed into an automated system, which evaluates the sleep stages based on the Sleep-EDF dataset, with EEG signals from healthy and sleeps problem patients as inputs. We employed only one or two EEG channels, making practical installation more straightforward than other cutting-edge systems that utilized PSG or multiple EEG channels or other physiological data for automated sleep-stage scoring, which are more complicated to implement [14], [15]. When compared to sleep scoring using multi-modal signals, the subject's comfort level is also improved. Furthermore, we used high-dimensional features derived from single- or multi-channel EEG signals using the Fast Fourier Transform (FFT). FFT is a time-tested and proven technique for extracting features from EEG signals. Our proposed method consists of four basic steps: EEG channel capture of brainwaves saved in the dataset, preprocessing, feature extraction, and classification evaluation using SVM and Random Forest classifiers to measure accuracy.

II. Materials and Methods

A. The Dataset

The EEG dataset used in this investigation was derived from the Sleep-EDF dataset[13], which is freely available. An extended version of the dataset with 61 recordings from 42 Caucasian male and female volunteers was used in this study, and it was one of three versions that were available. The participants ranged in age from 18 to 79 years old. This data was divided into two sections. EEG data from a study conducted between 1987 and 1991 comprised the first subset, which included 39 recordings from 20 participants. These individuals were in good health and were ambulatory. The second collection, which contained 22 recordings from 22 individuals, was derived from EEG data gathered in a 1994 study in which the participants had some difficulty going asleep but were otherwise healthy. The participants had some difficulty falling asleep but were generally healthy. The EEG data was collected over the course of 24 hours while the subjects were going about their normal lives. In a hospital setting, a small telemetry device was used to record the EEG data from four participants over the night. During our suggested research, data was collected from two channels: Fpz-Cz and Pz-Oz, with a sampling frequency of 100 Hz for each channel. Even-numbered electrodes are usually put on the right side of the head, whereas odd-numbered electrodes are placed on the left. The midline electrodes are denoted by the subscript z as zero. Besides that, the electrodes are labeled with letters denoting their locations in reference to anatomical brain divisions: C for central, F for frontal, T for temporal, O for occipital, P for parietal, and Fp for frontal pole, as depicted in fig.1 and fig.2 for the suggested research approach.

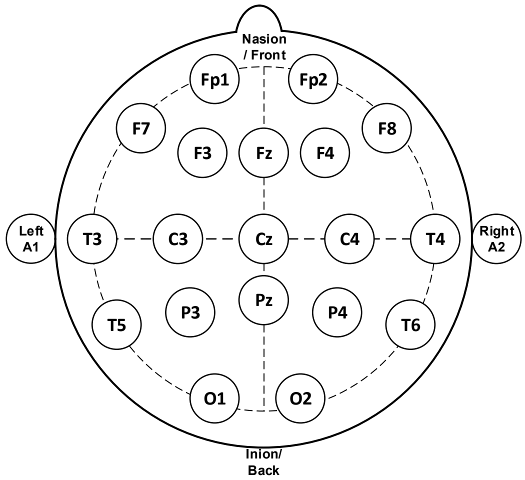

Fig. 1. The Placement of Electrodes on EEG Measurement*[16]*

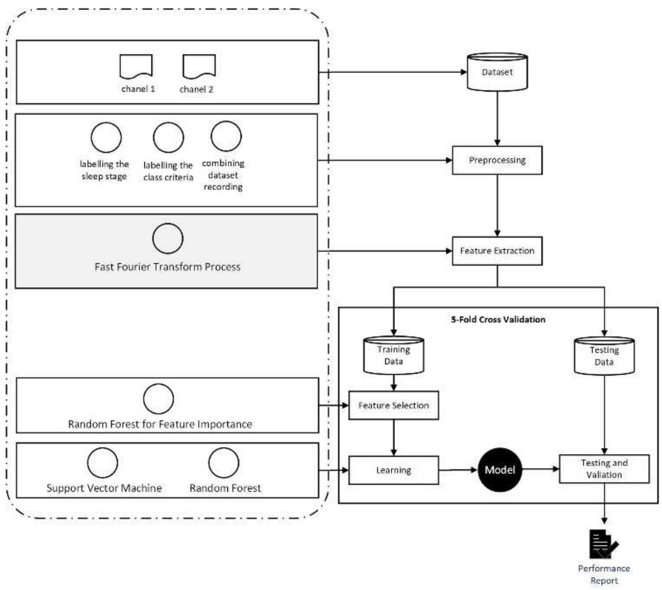

Fig. 2. The Research Procedure

B. Preprocessing

Based on the research procedure that the following processes were performed in this preprocessing, such as sleep labeling of the data set from 2- to 6-state categorization and combining all recorded subjects. EEG recordings of both categories had been interpreted on a 30-second basis by an experienced sleep technologist using R & K criteria. As a result, the period of each epoch was set to 30 seconds, resulting in a total of 3000 samples. "AWA," "REM," S1, S2, S3, S4, "Movement Time," and "Unscored" were the epochs designated by the sleep technologists. Annotations made by the American Academy of Sleep Medicine (AASM) comprised the classifications AWA, REM, N1, N2, N3, and "Unknown sleep state." In TABLE I, the number of samples was calculated according to the R&K criteria. After removing the samples with the labels "Movement Time" and

"Unscored," the total sample size is 127,663. During the period, time stood still for 30 seconds. In terms of the number of features analyzed, we produced the various features generated by feature selection, which began with 250, 500, 1000, 2000, 3000, and 6000 features taken from two channels (Pz-Oz and Fpz-Cz) and progressed to 6000 features taken from two channels (Pz-Oz and Fpz-Cz). The sampling frequency is 100 hertz, and the epoch is 30 seconds. In the studies, the characteristics were used either individually or in combination.

TABLE I. The total samples in Sleep-EDF dataset based on R&K criteria

∑ classes	AWA	REM	S1	S2	S3	S4
6	74,827	11,848	4,848	27,292	5,075	3,773
5	74,827	11,848	4,848	27,292	8,848	
4	74,827	11,848	32,140		8,848	
3	74,827	11,848	40,988			
2	74,827	52,836				

C. Feature Extraction

The feature represents a distinctive quality or an operative component that has been recognized in a pattern section, as well as an identifiable measurement that may be identified. The extraction of features from EEG signals is a critical stage in the processing of EEG signals. As a result, one of the objectives of feature extraction is to reduce the amount of meaningful information associated with the signal that is lost during processing. Furthermore, feature extraction minimizes the amount of time and resources required to accurately characterize a huge amount of information. It is possible to reduce the cost of data processing, simplify data implementation, and reduce the need to compress data if feature extraction is done appropriately[17]. Because EEG signals are dynamic and typically nonstationary, their frequency components must be known to determine when they occur. Time-frequency analysis is particularly well suited to dealing with such problems. In high frequency locations with transient waves, we usually need higher time accuracy, and in steady waves, we usually need more frequency resolution[18].

In this study, the Fast Fourier Transform (FFT) was utilized to extract the properties of EEG data in order to categorize sleep stages using a classification system. Because of this, the values of a specific time series are translated into a specific frequency domain and stored as numeric sequence data. Then we separated them into equal time periods called times in order to deconstruct the data into segmented EEG signal sequences, which we named segments. The length of each epoch dictated the length of the entire EEG signal, which was 30 seconds. The frequency spectra were produced using FFT when the frequency analysis was completed, and the epochs were processable. We used Fourier Transform (FFT) to transform a signal from its original time domain signal into a frequency domain signal[19]. Fig . 3 depicts a dataset from the Sleep EDF in the form of a time-domain signal and a frequency-domain signal on one of the datasets[13].

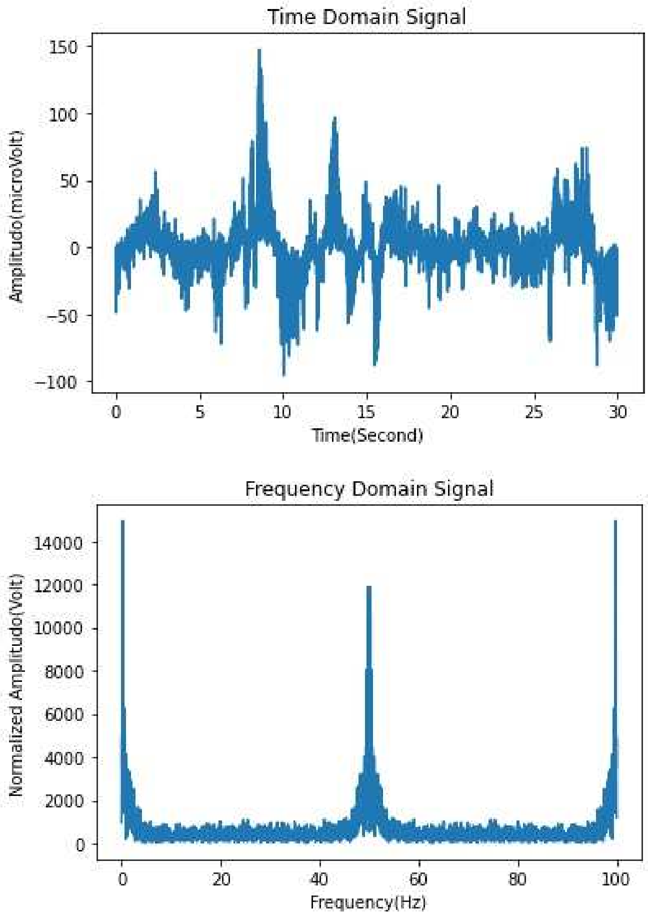

Fig. 3. The Form of Time Domain and Frequency Domain Signal

D. Feature Selection

Feature selection is the third step in the machine learning process, and it is crucial to the classifier's effectiveness. The purpose of this stage is to identify a collection of N features that are most useful in distinguishing between sleep categories based on the total number of classes that we have chosen. When it came to the Sleep-EDF dataset, we looked at the number of relevant features that could be selected from among 250, 500, 1000, 2000, 3000, and 6000 for each number of classes. The most acceptable feature subset for each number of classes was then chosen. More detail on the feature selection technique utilized in this work can be found in, which addresses essentially the same procedure as the one described in this study[20].

To avoid selecting features that are dominant in only a few patterns, the best N features were chosen using a 5-fold cross validation technique, with each fold including roughly the same number of segments for each sleep category. The 5-fold cross validation is an iterative procedure in which four folds are employed for feature selection in each iteration. The N features with the highest number of repeats (probability of appearance) were chosen as the final set of selected features after all iterations were completed. It is preferable to choose N as small as possible to avoid over-fitting. We have tried with the total number of different features in feature selection because it has been conveyed by Delimayanti et al. that more features are used to improve the performance of classification[4]. We employed a random forest classifier and a support vector machine in this study to categorize five examples into two to six classes using R& K criteria[14], [21].

III. RESULT AND DISCUSSION

As indicated in the pre-processing stage, the EEG data retrieved from the polysomnographic recording was processed, filtered, and segmented into 30 s epoch. The high-dimensional FFT features were then used to perform time–frequency analysis. The technique was performed for all 61 records, and the features data from all records was combined into a single features data set comprising 127,663 sleep epochs in total. In a random way, the entire data set was divided into two sets: a training set containing 4/5 of the features data and a testing set containing 1/5 of the remaining features data. A variety of classifiers, including Random Forest and SVM, were developed and tested using the training data set and the test data set, respectively. The performance of the proposed system was evaluated by comparing the output score of the classifiers to the score of the experts.

In addition, we had picked the multiclass support vector machine (SVM) technique, which is a supervised machine learning approach, from among the numerous classification algorithms that were available for consideration as a supervised machine learning approach. The SVM classifier is a well-known machine learning algorithm used to address a wide range of problems in various fields. The SVM algorithm is used to calculate the most significant margin around the separation hyperplane between the classes. In this inquiry, we used a Radial Basis Function (RBF) Kernel to solve the problem. One of the advantages of the SVM approach is that it may be employed when the number of features exceeds the number of samples, which is one of the most common situations. Furthermore, the model is suitable for use as a categorization model for EEG signals. After that, the performance of the SVM approach is compared to that of the widely used random forest classification method.

For two -six classes specified by R & K criteria were chosen as class labels in this experiment. Two channels' features were evaluated separately or in combination. 3000 features (100 Hz 30 s) were retrieved from a single channel. In addition, five-fold cross-validation was used to assess performance. The findings of Random Forest and SVM are compared to the outcomes of the acquisition. In the categorization of 6 to 2 classes, our technique with 6000 characteristics from two channels (Pz_Oz and Fpz_Cz) outperformed all other methods, as shown in tables 2 and 3.

TABLE II. The Performance of Reduced Number of Features with Random Forest Classifier (R&K Criteria, 2-6 Classes)

#of Classes	6000	3000	2000	1000	500	250
6	84.45%	84.81%	**84.86%**	84.84%	84.35%	83.89%
5	86.67%	**86.92%**	86.91%	86.78%	86.20%	85.71%
4	87.64%	87.78%	**87.91%**	87.75%	87.23%	86.63%
3	90.02%	90.26%	**90.41%**	90.19%	89.57%	89.01%
2	95.60%	95.81%	**95.93%**	95.77%	95.21%	94.29%

TABLE III. The Performance of Reduced Number of Features with SVM Classifier (R&K Criteria, 2-6 Classes)

#of Classes	6000	3000	2000	1000	500	250
6	85.92%	**87.94%**	86.31%	84.92%	83.09%	83.31%
5	86.87%	**87.94%**	87.75%	86.41%	84.60%	84.18%
4	87.52%	**88.90%**	88.66%	86.74%	84.30%	83.85%
3	90.38%	**91.27%**	90.99%	88.14%	86.86%	86.47%
2	96.53%	**96.63%**	96.44%	95.65%	93.63%	91.55%

In the performance report in table II and III, we can see that the accuracy in models with Random Forest and SVM increases as the number of features is increment value from feature selection. The accuracy as the performance evaluator on the Random Forest model had gained the best value on 95.93%, 90.41%, 87.91%, 86.92%, and 84.86% for 2-6 state classification. On the other hand, the SVM model reached 96.63%, 91.27%, 88.90%, 87.94%, and 87.94% for 2-6 state classification. The model with the SVM algorithm has given the best accuracy value compared to the Random Forest algorithm. Sleep stage classification in the form of time-series signals, as well as results obtained in text data classification, are made possible by the SVM algorithm, which produces classification results with the highest accuracy values. The SVM algorithm is also used in text data classification[22].

Moreover, it was found that for a certain number of features, the accuracy value reached the best. In Table II, with a feature selection of 2000 features, the best accuracy value was reached compared to some of 3000 and 6000 features. Moreover, in Table III, the research had conducted the model on an SVM classifier. The optimum number of features had been gained in the number of 3000 features from feature selection. The accuracy as the performance. It can be stated that feature selection improving the accuracy of the proposed model.

IV. CONCLUSION

It was successfully investigated on the effect of feature selection on automated sleep stage identification system based on EEG signal using R&K standard. This EEG signal was captured from the multichannel. The extraction of features from EEG signals has proven to be successful in utilizing time-frequency analysis using high dimensional FFT features. Although the dataset used records individual activities for almost one day (20- 24 hours), the accuracy obtained has provided more than 88% to overcome the imbalanced data. At the same time, the performance evaluation with the 2-6 state classification improved as the number of features executed and the feature selection attribute. In addition, feature selection affects the accuracy value of the model used. The model with SVM achieves the highest accuracy value with 3000 features, while the model with Random Forest has 2000 features. It can be concluded that SVM classifiers had gained the highest accuracy compared with the Random Forest on sleep stage classification on the sleep-EDF dataset.

ACKNOWLEDGMENT

We would like to express our highest gratitude to Politeknik Negeri Jakarta for supporting this research through Penelitian Kerjasama Luar Negeri (PKLN) Research Schema 2021.

REFERENCES

[1] K. Aboalayon, M. Faezipour, W. Almuhammadi, and S. Moslehpour, “Sleep Stage Classification Using EEG Signal Analysis: A

Comprehensive Survey and New Investigation," *Entropy*, vol. 18, no. 9, p. 272, Aug. 2016, doi: 10.3390/e18090272.

[2] V. Gurrala, P. Yarlagadda, and P. Koppireddi, "Detection of Sleep Apnea Based on the Analysis of Sleep Stages Data Using Single Channel EEG," *Trait. Signal*, vol. 38, no. 2, pp. 431–436, Apr. 2021, doi: 10.18280/ts.380221.

[3] S. Chokroverty, Ed., *Sleep Disorders Medicine: Basic Science, Technical Considerations and Clinical Aspects*. New York, NY: Springer New York, 2017. doi: 10.1007/978-1-4939-6578-6.

[4] M. K. Delimayanti *et al.*, "Classification of Brainwaves for Sleep Stages by High-Dimensional FFT Features from EEG Signals," *Appl. Sci.*, vol. 10, no. 5, p. 1797, Mar. 2020, doi: 10.3390/app10051797.

[5] L.-X. Feng *et al.*, "Automatic Sleep Staging Algorithm Based on Time Attention Mechanism," *Front. Hum. Neurosci.*, vol. 15, p. 692054, Aug. 2021, doi: 10.3389/fnhum.2021.692054.

[6] E. Kaniusas, "Fundamentals of Biosignals," in *Biomedical Signals and Sensors I: Linking Physiological Phenomena and Biosignals*, E. Kaniusas, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1–26. doi: 10.1007/978-3-642-24843-6_1.

[7] S. K. Ramakuri, C. Chakraborty, S. Ghosh, and B. Gupta, "Performance analysis of eye-state charecterization through single electrode EEG device for medical application," in *2017 Global Wireless Summit (GWS)*, Cape Town, Oct. 2017, pp. 1–6. doi: 10.1109/GWS.2017.8300494.

[8] R. G. Norman, I. Pal, C. Stewart, J. A. Walsleben, and D. M. Rapoport, "Interobserver Agreement Among Sleep Scorers From Different Centers in a Large Dataset," *Sleep*, vol. 23, no. 7, pp. 1–8, Oct. 2000, doi: 10.1093/sleep/23.7.1e.

[9] Z. Ge and Y. Sun, "Sleep Stages Classification Using Neural Networks with Multi-channel Neural Data," in *Brain Informatics and Health*, Cham, 2015, pp. 306–316.

[10] R. K. Tripathy, S. K. Ghosh, P. Gajbhiye, and U. R. Acharya, "Development of Automated Sleep Stage Classification System Using Multivariate Projection-Based Fixed Boundary Empirical Wavelet Transform and Entropy Features Extracted from Multichannel EEG Signals," *Entropy*, vol. 22, no. 10, 2020, doi: 10.3390/e22101141.

[11] S. Taran, P. C. Sharma, and V. Bajaj, "Automatic sleep stages classification using optimize flexible analytic wavelet transform," *Knowl.-Based Syst.*, vol. 192, p. 105367, Mar. 2020, doi: 10.1016/j.knosys.2019.105367.

[12] B. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms," *J. Med. Syst.*, vol. 38, no. 3, p. 18, Mar. 2014, doi: 10.1007/s10916-014-0018-0.

[13] S. A. Imtiaz and E. Rodriguez-Villegas, "An open-source toolbox for standardized use of PhysioNet Sleep EDF Expanded Database," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, pp. 6014–6017. doi: 10.1109/EMBC.2015.7319762.

[14] O. Yildirim, U. Baloglu, and U. Acharya, "A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals," *Int. J. Environ. Res. Public. Health*, vol. 16, no. 4, p. 599, Feb. 2019, doi: 10.3390/ijerph16040599.

[15] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowl.-Based Syst.*, vol. 128, pp. 115–124, Jul. 2017, doi: 10.1016/j.knosys.2017.05.005.

[16] S. A. Keenan, "Handbook of Clinical Neurophysiology, An Overview of Polysomnography," *Elsevier BV Amst. Neth.*, vol. 6, no. 3, p. 18, 2005.

[17] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains," *ISRN Neurosci.*, vol. 2014, p. 730218, Feb. 2014, doi: 10.1155/2014/730218.

[18] M. Ravan and Department of Electrical & Computer Engineering, New York Institute of Technology (NYIT), 1855 Broadway, New York, NY 10023, USA, "A machine learning approach using EEG signals to measure sleep quality," *AIMS Electron. Electr. Eng.*, vol. 3, no. 4, pp. 347–358, 2019, doi: 10.3934/ElectrEng.2019.4.347.

[19] H. J. Nussbaumer, "The Fast Fourier Transform," in *Fast Fourier Transform and Convolution Algorithms*, H. J. Nussbaumer, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 80–111. doi: 10.1007/978-3-662-00551-4_4.

[20] C.-S. Huang, C.-L. Lin, L.-W. Ko, S.-Y. Liu, T.-P. Su, and C.-T. Lin, "Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels," *Front. Neurosci.*, vol. 8, Sep. 2014, doi: 10.3389/fnins.2014.00263.

[21] T. Nakamura, V. Goverdovsky, M. J. Morrell, and D. P. Mandic, "Automatic Sleep Monitoring Using Ear-EEG," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–8, 2017, doi: 10.1109/JTEHM.2017.2702558.

[22] M. K. Delimayanti, Sari, Risna, Laya, Mauldy, Faisal, M. Reza, Pahrul, and Naryanto, R. Fitri, "The Effect of Pre-Processing on the Classification of Twitter's Flood Disaster Messages Using Support Vector Machine Algorithm," presented at the International Conference on Applied Engineering (ICAE), Batam, Indonesia, Oct. 2020.