

emit-1

By if6-emit if5

ABSTRAK

Gangguan aliran pernapasan selama tidur akan menyebabkan gangguan jantung yang serius jika tidak tangani dengan serius. Salah satu pendekatan yang dilakukan untuk mendeteksi gangguan tidur sejak dini adalah menggunakan rekaman elektrokardiogram (EKG) sebagai representasi aktivitas listrik saat jantung bekerja. Ketidakteraturan dari morfologi dan kompleksitas rekaman memiliki arti klinis yang bisa dijadikan alat bantu untuk mendiagnosa gangguan tidur. Penelitian ini menggunakan rekayasa teknik untuk mendapatkan fitur-fitur dari rekaman EKG yang dilakukan secara otomatis menggunakan mesin Pembelajaran deep learning dengan pendekatan model Convolutional Neural Network (CNN). Rekaman EKG diproses untuk menghilangkan derau sebelum digunakan pada model CNN. Pengujian dilakukan pada model yang paling optimal untuk mendapatkan akurasi yang baik dengan menerapkan dua skenario. Hasil pengujian kedua skenario menunjukkan bahwa skenario satu mendapatkan akurasi 83,03% dibandingkan dengan skenario dua dengan akurasi 76,88%. Sedangkan tingkat presisi, sensitivitas, cohens kappa dan ROC UAC masing-masing sebesar 81,78%, 87,78%, 65,73% dan 82,68% pada pengujian skenario satu pada model CNN dengan pengaturan parameter paling optimal.

Kata kunci: gangguan tidur, diagnosa gangguan tidur, masalah jantung, cnn, ecg

ABSTRACT

Disruption of the flow of breathing during sleeping will result in significant heart problems if not treated seriously. An electrocardiogram (ECG) recordings is one of the most used methods for detecting sleep disorders early on. An ECG is a representation of electrical activity in the heart while it is beating. The irregularities of the morphology and the complexity of the recordings have clinical significance that can be used as a tool for diagnosing sleep disorders. This study uses engineering to obtain features from ECG recordings that are carried out automatically using deep learning machine learning with a Convolutional Neural Network (CNN) model approach. The ECG recordings were processed to remove noise before being used in the CNN model. Tests are carried out on the most optimal model to get good accuracy by applying two scenarios. The test results of the two scenarios show that the first scenario has an accuracy of 83.03% compared to the second scenario with an accuracy of 76.88%. Meanwhile, the precision, sensitivity, cohens kappa, and ROC UAC levels were 81.78%, 87.78%, 65.73% and 82.68% in the first scenario testing on the CNN model with the most optimal parameter settings, respectively.

Keywords: sleep disorder, sleep diagnosing, heart problems, cnn, ecg

1. INTRODUCTION

The common symptoms for people with sleep disorders can usually be seen from the presence of snoring even though they do not feel tired all day, feel tired and have headaches when waking up, not feeling fresh, dry mouth, feeling constantly sleepy during the day, and reduced concentration (AkŞahİN et al., 2018; Wise et al., 2007). Prior study has shown that humans will in fact spend about a third of their lives sleeping (Fekr et al., 2015). A sleep disorder that many people experience is *sleep apnea (SA)*. Grounded in the diagnosis, SA is categorized into three, i.e. *Obstructive Sleep Apnea (OSA)* or commonly known as *Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS)*, *Central Sleep Apnea (CSA)*, and *Mixed Sleep Apnea (MSA)*. These SA categories can be identified by observing the frequency of occurrence of the *Apnea/Hypopnea Index (AHI)* per hour. The criteria for AHI scores based on the standard are AHI classification <5 indicated as normal OSA condition, AHI between 5-14 as mild OSA, AHI between 15-30 as moderate OSA, and AHI > 30 classified as severe OSA (Hudgel, 2016). The prevalence of OSA is estimated at 9%-24% of the total population and 90% is undiagnosed (Singh et al., 2013).

Related studies on sleep disorders are more concerned with dependence on human needs and interventions. One of them is the visual detection of sleep disorders for very large datasets that will hinder the diagnosis process. This takes a very long time which is inefficient and at risk of misdiagnosing. Therefore, the computer-assisted diagnostic process becomes a mandatory requirement to produce a system that is utilized to diagnose sleep disorders. Electrocardiogram (ECG) signals, for instance, are widely used to detect computerized sleep disorders (Fahruzi et al., 2020). The feasibility of software and hardware is a concern to be developed so that it can be used anytime, anywhere, and by anyone without much intervention from other parties. Currently, there are two approaches used to obtain features from ECG recordings, i.e. manual extraction process and automatic extraction process. The difference between the two methods has significance in the final process in the form of simplifying the process and improving the performance of the classification process. Increasing accuracy in the classification process has initially been developed using various extraction methods through modern approaches such as feature engineering based on images with one-minute ECG recordings (Niroshana et al., 2021), prediction of sleep disorders based on ECG recordings *Scalograms* and *Spectrograms* (Nasifoglu and Erogul, 2021), or the current conventional approach which is better known as *handcraft features*, one of which is by increasing the frequency features of Spectrograms (Wu et al., 2021), and diagnosis that grounded in the complexity of recordings based on fractal analysis of a signal (Fahruzi et al., 2021).

In this respect, the present study shows a new approach for automatic identification of sleep disorders using one-lead ECG recordings through deep learning engineering techniques with the Convolutional Neural Network (CNN) model approach. One lead means using the electrodes in only one particular position on the body.

2. METODE PENELITIAN

2.1 Dataset

The present study reports on testing a deep learning model to detect sleep disorders using the ECG-apnea dataset sourced from the *physionet database*. Recording duration ranges from 7 to 10 hours. This study used ECG recordings that were segmented into 10 seconds and

20 seconds. Following Purnomo et al. (2020), the dataset used was then normalized aiming to improve performance by accelerating the training process and not always relying on *dropout* parameters to reduce *overfitting*. The use of *Batch Normalization (BN)* basically produces a distribution of data points for each sample with the best standard normal distribution through a linear transformation process. The implementation of BN on an artificial intelligence network was developed by Ioffe and Szegedy (2015).

2.2 Preprocess of ECG Signals based on Baseline drift or Baseline wander

Preprocessing has a pivotal role to obtain accuracy in detecting the R wave without being affected by various kinds of noise. Mathematically, the ECG signal $y[n]$ is defined by referring to Equation 1. The variable $s[n]$ is a representation of the ECG signal and $w[n]$ is the value obtained from the overall noise that affects the ECG signal. Whilst α is the signal attenuation parameter.

$$y[n] = \alpha s[n] + w[n] \quad (1)$$

Overall, complex QRS is a wave whose shape does not change despite a lot of noise. When recording an ECG signal, in general, the frequency range for obtaining the main information is from 200-500Hz. In this section, no high frequency is required to get the right information. Specifically, the frequency range of QRS waves is between 5 to 25 Hz (Elgendi et al., 2014). While the other waves have a frequency range below the frequency range of the QRS wave. The following are sources of noise in the ECG signal:

- a) *Baseline drift or baseline wander*: This noise effect produces a low-frequency variation from the ECG baseline. This type of noise severely limits the usefulness of the recorded EKG. Hence, it needs to be reduced for better clinical evaluation. Several methods have been developed for noiseless ECG elevation, including the use of the wavelet method and empirical mode decomposition (Blanco-Velasco et al., 2008). This filtering process can reduce high-frequency noise and baseline drift simultaneously without losing important information from the ECG signal.
- b) *Power line interference (PLI)*: is a significant source of noise during bio-potential measurements. The interference is at a frequency of 50 to 60 Hz. This condition reduces signal quality and makes it difficult to obtain small features that may be important and have clinical information for diagnosis.

2.3 Electrocardiogram Signal Morphology

ECG is a very common instrument used to detect the heart's electrical activity over time. The heart is enclosed by a membrane of cells, each of which contains a charge that gets disconnected during each heartbeat. Figure 1 shows the anatomy and electrical system of human heart. The cardiac cycle refers to a complete heartbeat from the beginning to the start of the next beat and has a series of waves labeled P, QRS, and T as shown in Figure 2. Such definition of each recording in the ECG is as follows:

- P wave is a process of atrial depolarization. When the valves between the atria and ventricles open, 70% of the blood in the atria is extracted by the ventricles as they expand. The contractions applied by the atria are required for the next 30%. Accordingly, the work carried out by the muscles requires only low tension;
- PQ segment represents the stage before the initial contraction where blood moves into the ventricles;
- QRS wave represents ventricular depolarization by ventricular muscle contraction;
- The ST segment connects the QRS wave and T wave. This segment is the interval between ventricular depolarization and repolarization whereby irregularity will lead to infarction or myocardial ischemia;

- The T wave corresponds to the final phase of the ventricular muscle cell potential action. The same potential handles contraction and repolarization, but one is an upstroke and the other is a downstroke. For this reason, T waves may be associated with unspecified events. To get the value of the RR interval as a basis for calculating other parameters as a representation of HRV, each beat on the ECG recording displays the electrical activity of the heart. In the complex R waveform, QRS is always pointing upwards and therefore is usually used to identify each beat and its duration which is called the RR interval (from one R to successive R waves). Figure 2 shows that two R beats produce one RR interval, and so on to get the average value of the RR interval.

2.3 Characteristics of Heart rate Variability (HRV)

Normal biological systems exhibit heart rate patterns with complex variability and can be explained mathematically. This variability describes the change in the time interval between successive heartbeats called interbeat intervals (IBIs). In this section, an overview of changes in heart rate oscillations will be described based on the HRV time domain, namely the frequency domain and the nonlinear metric domain. A time domain will measure the number of HRV observed over a certain period. Changes in the frequency domain to determine the absolute or relative value of signal energy in the component spectrum band range. Meanwhile, nonlinear measurements are used to measure the uncertainty and complexity of the observed series of IBIs.

Heart rate is the number of heartbeats counted per minute. An HRV is basically a fluctuation in the time interval between adjacent heartbeats. High HRV values are not always better because pathological conditions can produce HRV. When there is an abnormality in heart function, it increases the HRV measurement. It is strongly associated with an increased risk of death. Therefore, an in-depth examination of ECG morphology and periodic patterns of ECG signals can reveal whether the increase in HRV values is caused by problems such as breathing problems due to obstruction during sleep.

2.5 Modern Deep Learning Method for Classification of Sleep Disorders

In this study, one-lead ECG recordings were also tested on modern machine learning such as deep learning, specifically using the Convolutional Neural Network (CNN) architecture with a framework as shown in Figure 3. The main objective was to test the different stages and classification results based on the features that previously have been extracted. In the era of artificial intelligence in recent years, it has developed rapidly and has become a solution in solving various cases, including those related to image processing, signal processing, text processing, voice recognition, video processing, and case resolution in the biomedical field. The growing use of deep learning increases the capabilities of computing devices, as well as the development of algorithmic capabilities, such as the implementation of activation functions, weight initialization schemes, optimizations schemes, batch normalization, and depth-wise separable convolutions.

2

3. RESULTS AND DISCUSSION

In this section, the use of deep learning is to get the best accuracy results based on a one-dimensional CNN architecture (1D-CNN) on a dataset that has been segmented by 10 seconds and 20 seconds. The purpose of this division is to find out which dataset has the best results based on settings on several parameters to improve the performance of the selected CNN model. The number of records used is adjusted to the scenario to be applied. Testing

according to this scenario is based on the number of samples, optimization of parameters, and implementation of different architectures.

3.1 Scenario 1: This Schema Uses a 10 sec Dataset

Prior to use, the dataset is preprocessed to reduce the effect of noise. Thereafter, the preprocessed dataset is normalized using *zscore* to obtain a feature value scaling between 0 and 1, but without destroying or losing information because it is still within its limits. The proposed model is shown in Figure 4. The main objective of the test is to find out whether CNN is able to classify apnea and non-apnea with better accuracy results. Parameter settings are carried out in each scenario starting from testing using standard settings to adjusting parameters. Parameter settings are repeated until the best accuracy results are obtained during the training, validation, and testing processes. To see performance measures during the training process, it can be a graph in the form of accuracy and errors during the process. The goal is to analyze and evaluate a model, if the results are less promising, the model can be improved and parameter settings can be made until promising results are obtained with the appropriate model.

Furthermore, if CNN model has been selected, then, the next step is to test it with new data as testing data. The composition of the dataset is divided into three parts, i.e. training data set (90%) with 38,344, validation data for 3634 samples and test data set (20%) with 8925 samples. Test data contains a completely separate data set and is not part of the training data. This is to avoid biased results and do not reflect the superiority of a model over a new data set. However, before carrying out testing process, the selected model is validated using validation data taken from training data. The purpose of model validation is to ensure that the architecture is able to produce good performance without *overfitting*.

The implementation of 1D-CNN uses a configuration consisting of a *convolutional layer* (conv layer), *max-pooling*, *dropout*, and *fully-connected (FC)* with different amounts. All convs use Relu activation function, but the output part of the FC uses sigmoid activation. In the compilation section, we use *sparse_categorical_crossentropy* as a loss function because it is used to classify two classes. Meanwhile, for CNN parameter optimization, we use Adam's optimization function of 0.0001 with 50 epochs and batch size=32.

Table 1 is the CNN architecture settings for scenario 1. Grounded in the summary of the model, a more complete architecture with more layers and the appropriate parameter settings results in a better performance measure than scenario 2. The CNN architecture of scenario 1 uses 655,730 configurable parameters. Table 2 shows the results of the CNN scenario 1 performance measurement with an accuracy of 83.03%, a precision score of 81.78%, a sensitivity/recall score of 87.78%, a Cohens Kappa score of 65.73%, and a ROC AUC calculation of 82.68%.

3.2 Scenario 2: This Schema Uses a 20 sec Dataset

In line with the implementation of scenario 1, the dataset used has been preprocessed. The test uses an architecture as shown in Figure 5. This test uses a dataset with divisions, i.e. 12,903 training data, 2581 validation data which are part of the training data, and 3141 test data. Each data has 2000 samples or a duration of 20 seconds. The result of the model construction shows that the parameters that can be set are 168,892.

Figure

Hereafter, the parameter setting for model optimization is using the Adam function with a learning rate of 0.001 and the number of epochs = 100. Meanwhile, the architectural construction for scenario 2 is translated into a summary model as shown in Table 3.

Table

Based on the test, the application of scenarios in each tested model produces different accuracy. This is due to many factors, including the number of data and the number of samples, dataset complexity, data normalization and standardization, different architectures, and parameter optimization. The task of classifying, apnea and non-apnea of the ECG recording is a challenging one. However, what is required is the right architecture configuration with various scenarios such as setting the learning rate, the use of the adaptive gradient method, a large-scale distributed training process, and the use of iteration algorithms that do not converge quickly during the training process and model testing.

Parameter setting is undertaken by repeated testing until the most optimal results are obtained. Changes to one parameter do not necessarily produce the best accuracy. In one condition the results of the settings require a very long computation time. For instance, setting the learning rate is very important to optimize inasmuch as it can improve CNN performance, but if the value is too large it will make CNN work harder to extract information so that the process is unstable. On the other hand, if it is too small, the learning process will take longer.

In addition, the architectural complexity and the number of datasets also affect CNN's performance. Scenario 1 produces better accuracy than scenario 2 because the number of datasets is larger, the number of layers is more complex, and the parameter settings are more in line with the characteristics of the data. Furthermore, to get better results, such corrective actions are needed as increasing the amount of data using several schemes, e.g. augmenting the dataset using the oversampling method and Synthetic Minority Over-sampling Technique (SMOTe) which maximizes the minority class to generate synthetic data. In addition, it is also necessary to try other normalization and standardization techniques such as maximum absolute scaling, min-max scaling, and robust scaling so as to overcome differences in feature values to the same scale by re-scaling the dataset.

Figure 6 shows the curve of the relationship between accuracy and error performance. The lower the error level will result in better accuracy. In order to maintain the shape of the curve with high accuracy and low error level, attention must be paid to parameter settings and data validation so that there is no overfitting, whereby performance during training is better than performance during testing. This is because the model is not able to recognize the test data optimally so that the level of accuracy does not increase at the same time the error rate is still high.

4. CONCLUSION

The testing shows that the CNN model with parameter settings according to the first scenario has the highest performance value compared to the second scenario. Scenario 1 sequentially has detection performance with accuracy, precision, sensitivity, Cohens kappa, and ROC UAC of 83.03%, 81.78%, 87.78%, 65.73%, and 82.68%. To increase accuracy in deep learning, additional datasets with sufficient balance of training data, validation data, and test data for both apnea and non-apnea classes are required. In addition, deep learning requires innovation through transfer learning techniques that function to complete a specific task using part or all of a model that was pre-trained on a different task. This approach is expected to improve the performance of the classification method. The use of large datasets in deep learning causes more attention to datasets, including how to add data, how to make

Commented [L1]: Ini dicak lagi pak, karena file Bahasa indonya sepertinya salah tulis

existing data more abundant, how to rescaling data, how to normalize data, and how to select data. The scenario in the dataset aims to increase accuracy without overfitting.

REFERENCES

emit-1

ORIGINALITY REPORT

9%

SIMILARITY INDEX

PRIMARY SOURCES

1	creativecommons.org Internet	202 words — 6%
2	"Medical Image Understanding and Analysis", Springer Science and Business Media LLC, 2017 Crossref	17 words — 1%
3	WWW.MDPI.COM Internet	16 words — < 1%
4	jtd.amegroups.com Internet	13 words — < 1%
5	www.ncbi.nlm.nih.gov Internet	11 words — < 1%
6	ajec.smartsociety.org Internet	10 words — < 1%
7	docplayer.net Internet	9 words — < 1%
8	archive.org Internet	8 words — < 1%
9	Khalil Ahmad, Abdullah. "Chapter 5 Applications in Signal Processing", Springer Science and Business Media LLC, 2018 Crossref	7 words — < 1%

EXCLUDE QUOTES OFF

EXCLUDE SOURCES OFF

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES OFF