Automated Sleep Apnea Detection Using CNNs: Insights into the Impact of FFT Feature Extraction on EEG Signals

Mera Kartika Delimayanti [1,*], Asep Taufik Muharram [1], Ayres Pradiptyas [1], Rinaldito Ahmad Ryanari [1], Raditya Arya Prasetyo [1], Rizky Adi [1], Mohammad Reza Faisal [2], Rizqi Fitri Naryanto [3], and Haralampos Hatzikirou [4]

[1] Computer and Informatics Engineering Department, Politeknik Negeri Jakarta, Depok, Indonesia
[2] Computer Science Department, Lambung Msangkurat University, Banjarbaru, Indonesia
[3] Mechanical Engineering Department, Engineering Faculty, Universitas Negeri Semarang, Semarang, Indonesia
[4] Mathematics Department, Khalifa University, Abu Dhabi, United Arab Emirates
Email: mera.kartika@tik.pnj.ac.id (M.K.D.); asep.muharram@tik.pnj.ac.id (A.T.M.); ayres.pradiptyas@tik.pnj.ac.id (A.P.); rinaldito.ahmadryanari.tik19@staff.pnj.ac.id (R.A.R.); raditya.aryaprasetyo.tik19@mhsw.pnj.ac.id (R.A.P.); rizky.adi.works@gmail.com (R.A.); reza.faisal@ulm.ac.id (M.R.F.); rizqi_fitri@mail.unnes.ac.id (R.F.N.); haralampos.hatzikirou@ku.ac.ae (H.H.)
*Corresponding author

***Abstract*—Sleep apnea, a serious sleep disorder characterized by interrupted breathing during sleep, poses significant health risks such as cardiovascular disease and diabetes. Traditional diagnostic methods, such as polysomnography, are cumbersome and expensive, creating a demand for automated solutions. Previous sleep apnea detection research relied on Multi-Layer Perceptrons (MLPs), which, while functional, can be limited in capturing the complex temporal dependencies within Electroencephalogram (EEG) signals. Our study introduces a novel approach by combining Convolutional Neural Network (CNN) with Fast Fourier Transform (FFT) to detect sleep apnea. The FFT was implemented as feature extraction to capture frequency-domain characteristics of the EEG signals. Two versions of the model were developed: one trained on raw EEG data and the other on FFT-processed data. The Physionet Sleep-EDF Database served as the source for EEG recordings, labeled as "normal" or indicative of sleep-disordered breathing. Performance metrics, including accuracy, precision, recall, and F1-Score, were used to evaluate the models. The CNN trained on raw EEG data achieved superior results, with 92% accuracy, a precision of 0.86, recall of 1.00, and an F1-Score of 0.92, outperforming previous studies utilizing Multi-Layer Perceptrons (MLP). However, the result shows that the approach using FFT produces worse results. This suggests that, in the context of sleep apnea detection using our specific dataset, the most discriminative features may not reside solely in the frequency domain as extracted by FFT. The results demonstrate the potential of CNNs in developing low-cost, accessible diagnostic tools. Future efforts should address dataset limitations and explore alternative feature extraction methods to improve generalizability.**

***Keywords*—Convolutional Neural Network (CNN), deep learning, Electroencephalogram (EEG), sleep disorder, sleep apnea**

Manuscript received February 1, 2025; revised March 26, 2025; accepted May 8, 2025; published September 5, 2025.

I. INTRODUCTION

Sleep apnea is a sleep disorder characterized by interrupted breathing during sleep. It is prevalent among adults and also affects a small portion of the younger population [1]. Individuals with sleep apnea experience episodes of either no breathing or insufficient breathing while asleep. The initial condition, characterized by temporary pauses in breathing, is called apnea, whereas the subsequent condition, involving intervals of reduced airflow or shallow breathing, is known as hypopnea. Both conditions can lead to associated medical issues, underlining their harmful impact on an individual's overall health [2]. A specific kind of sleep apnea, Obstructive Sleep Apnea (OSA), is typified by repeated episodes of partial or complete upper airway blockage, which causes intermittent hypoxia and disturbance of sleep [3]. The main negative effect of OSA includes hypoxia, which is a condition of decreasing oxygen levels, and disrupted sleep patterns. Additionally, research has demonstrated that OSA affects the body's sleep regulation and has negative effects on metabolic factors, insulin sensitivity, and cardiovascular health. This research suggests a correlation between sleep disorders and metabolic conditions, including obesity, metabolic syndrome, and diabetes mellitus [4].

In addition to Obstructive Sleep Apnea (OSA), there are two other forms of sleep apnea: Central Sleep Apnea (CSA) and Mixed Sleep Apnea (MSA). Each individual is marked by recurrent cases of infection of the upper respiratory tract, accompanied by a stable respiratory rhythm. In CSA, respiration may be reduced or entirely absent, whereas MSA is characterized by a combination of two distinct types of apnea [5]. During the diagnostic procedure, the circumstances of the sleep apnea event is used to determine the specific kind of sleep apnea that a patient has. Patients with MSA frequently combine central and obstructive events; those diagnosed with OSA usually experience obstructive events mostly; those identified with CSA mostly face central events [5, 6].

The physiological symptoms of this sleep disorder encompass snoring, experiencing abrupt breathlessness during sleep, waking up with a parched mouth, and, overall, encountering subpar sleep quality. These factors contribute to reduced concentration, insomnia, declining cognitive abilities, mishaps, memory impairment, and feelings of despondency. Apart from the diminished quality of life due to inadequate rest and weariness, sleep apnea can also trigger serious complications such as diabetes, cardiovascular disorders, high blood pressure, neurological challenges, and liver ailments. Given the widespread prevalence of sleep apnea and its associated long-term repercussions, both direct and indirect, it holds significance to accurately diagnose and address this condition [7–9].

Remarkable progress in technology and artificial intelligence with machine learning or deep learning has proven to be useful in healthcare field [8–11]. For example, machine learning has been used for depression detection during COVID-19 crisis [10, 12]. Conversely, deep learning significantly influences healthcare by facilitating physician fixation prediction and the classification of pre-cancerous cervical lesions [13, 14]. This means machine learning and deep learning are also applicable for detecting sleep apnea. Since CNNs could automatically extract local temporal and spatial features from time-series data such as EEG signals, they were selected [15]. CNNs are more useful for EEG waveform analysis than MLPs as, through convolutional layers, they record local dependencies. CNNs are also computationally more efficient and more suited for fixed-length segments—as employed in this work—than Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTMs). CNNs beat MLPs in biomedical signal categorization applications, according to earlier studies [10, 16–18]. For the objective of creating an accurate, automated sleep apnea detection model using EEG, CNNs thus provide an ideal mix between accuracy, efficiency, and interpretability.

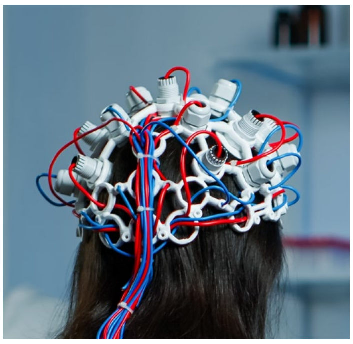

Fig. 1. The process of real time EEG examination.

Utilizing brain physiological signals or an Electroencephalogram (EEG), sleep apnea can be detected using deep learning model [19]. Electroencephalography (EEG) is a non-invasive technique employed to document the brain’s electrical activity, recording neural signals with high temporal precision [11]. The recorded brain impulses yield significant data that may be processed and analyzed by machine learning techniques, especially deep learning, to reveal patterns and insights for many applications [9]. Fig. 1 shown the efficacy of the EEG device in acquiring brain signal recordings.

II. Related Works

Previous research has demonstrated that the use of the Multi-Layer Perceptron (MLP) deep learning model achieves promising results, with an accuracy rate of 86% in detecting sleep apnea [19, 20]. However, this accuracy can be further enhanced. One of many approaches was changing the deep learning model. One of many deep learning models is Convolutional Neural Networks (CNN) which proved to be better than MLP on breast cancer detection or using other machine learning methods [13]. Besides changing the model, another approach is to employ feature extraction on the dataset. The feature extraction on the dataset can be done using Fast Fourier Transform (FFT), which believed could enhance the model performance than dataset that did not go through the FFT [11].

Recently, CNN has become a dominant and highly effective model in EEG classification task space [10, 21–23]. Because of that, we opted for CNN architecture over MLP for this sleep apnea detection study. Numerous prior works in EEG analysis have demonstrated the advantages of CNNs in automatically learning relevant features and achieving state-of-the-art performance [10, 22, 23].

The aim of this study is to develop an automated model for sleep apnea detection with improved performance over previous studies. We focus on introducing a new architecture approach compared to previous works which solely use MLP which is deemed to be outdated. This research introduces a novel methodology that leverages deep learning techniques, specifically Convolutional Neural Networks (CNNs), in combination with Fast Fourier Transform (FFT) for feature extraction from EEG datasets. While CNNs and FFT have shown effectiveness in enhancing model performance in other applications [9, 20, 24]. This study focuses on evaluating their combined efficacy within the context of sleep apnea detection. Additionally, the study compares the proposed approach to prior work and assesses the impact of FFT-based feature extraction by contrasting the performance of models trained on raw EEG data against those trained on FFT-processed datasets.

III. Materials and Methods

This article presents a method for detecting Sleep Apnea using Electroencephalogram data with Convolutional Neural Networks. This method outlines a systematic, sequential routine to follow during the detection process. This process is fully depicted in Fig. 2.

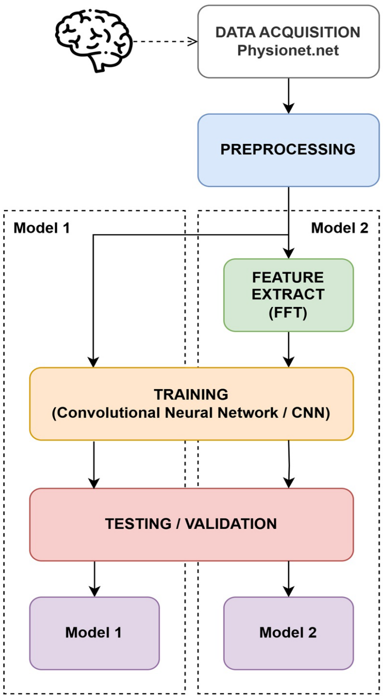

Fig. 2. Block diagram of sleep apnea detection development procedure.

In summary there are four steps for performing sleep apnea detection model development contained on Fig. 2, which are data acquisition, data preprocessing, training model, and testing/validating the trained model. On data acquisition steps, the brain physiological signals or an Electroencephalogram (EEG) data are collected, loaded, and annotated. Then those datasets are pre-processed to transform data into a format that is suitable for training process, in this step the dataset also splits into training and testing purposes data. In the training step, the data for training purposes from previous steps was then used to train the Convolutional Neural Network model. The trained model then tested or evaluated using testing purposes data to see the performance of the model.

A. Dataset

The dataset utilized in this research is derived from the Physionet Database, documented by the Sleep Disorders Center at Ospedale Maggiore in Parma, Italy [25]. Within this dataset, there are 108 polysomnographic records, each containing at least 3 or more channels of brain physiological signals per record, along with annotations of sleep events with other signals. Out of 108 records, only 4 of them are labelled as “sleep-disorder-breathing” (sdb) records and 16 of them labelled as “normal” records.

In Physionet database, there are 20 records available, including 4 records with subjects with apnea and 16 records of subjects with no pathology (normal). However, we only use 5 out of 20 records. This is because, among 4 records with apnea, 2 records are unable to open. Thus, resulting only 2 of the apnea records are usable for the research. Additionally, the other 3 records come from records with no pathology (normal). The chosen of the records is following the sequence naming of the sample. The amount of the subject with no pathology record is from the balance amount of segmented sample from the apnea and no pathology records. The 14 unused normal records were excluded solely to balance the apnea/normal ratio at the segmented sample level, not due to quality concerns.

The dataset follows the 10–20 international system, which consists of Fp1-F3, F3-C3, C3-P3, P3-O1 and/or Fp2-F4, F4-C4, C4-P4, P4-O2 [25]. The location of electrodes in 10–20 international system are shown Figs. 3 and 4 illustrated the raw EEG signals obtained from the dataset. These signals serve as the primary input for the study, providing the data used for preprocessing, feature extraction, and training of the CNN models for sleep apnea detection.

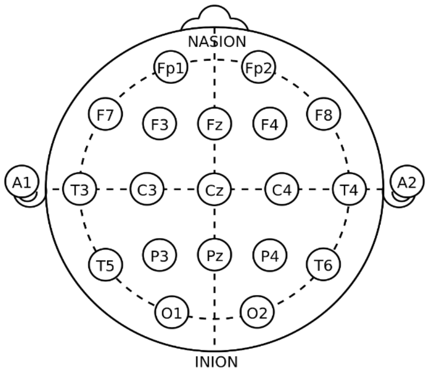

Fig. 3. Electrode locations of international 10–20 system of EEG recording [11].

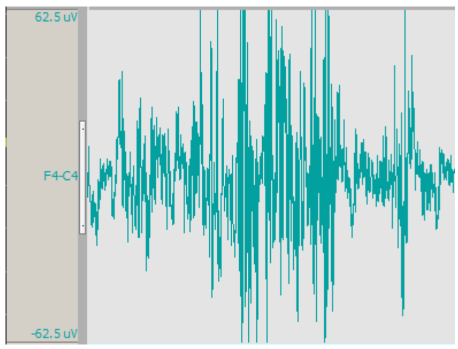

Fig. 4. The raw EEG signal generated.

B. Maintaining the Integrity of the Specifications

Data preprocessing step was done before the dataset can be used for training sleep apnea detection model. Preprocessing aims to clean, normalize, and prepare the data for the detection model’s utilization. The following Fig. 5 outlines the preprocessing stages.

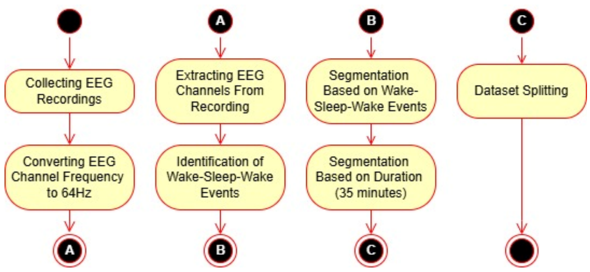

Fig. 5. Block diagram of data preprocessing.

1) Converting frequency

The frequency of the original brain physiological signal records ranging from 256 Hz to 512 Hz with a duration of 7 to 8 h. To make the EEG frequency consistence, reducing frequency down to 256 Hz, which is the lower frequency. Unfortunately, the usage of samples with 256 Hz is too large for the available memory. resulting in force restart during modelling or even unable to load the samples. To be able to do the modelling, the approach of reducing frequency of the sample is chosen, with 64 Hz as the highest frequency possible without resulting in any error due to limited memory. Fig. 6 illustrates the procedure for capturing EEG in cases of sleep apnea. All data from each channel of brain physiological signals is exported into text files.

2) Dataset segmentation

The segmentation process extracts 35-minute EEG data segments. The following Fig. 6 illustrate the segmentation process. It identifies sleep event using the database's 30-second interval sleep annotations, defining these events as periods starting from the last "Wake" annotation and ending at the next "Wake" annotation. Each identified event is segmented, and then trimmed to include only the final 35 min. This creates uniform-length input data for the CNN. Finally, each 35-minute segment is labeled as "normal" or "sleep-disordered breathing (sdb)" based on the original record's overall classification, providing a consistent, albeit simplified, dataset for model training and evaluation.The result of this segmentation process was a dataset with a total of 60 samples, with 30 samples labelled as "normal" and 30 samples labelled as "sdb". The data on sleep apnea occurrences were collected during sleep phases N2 and N3, as indicated in the highlighted box in Fig. 6.

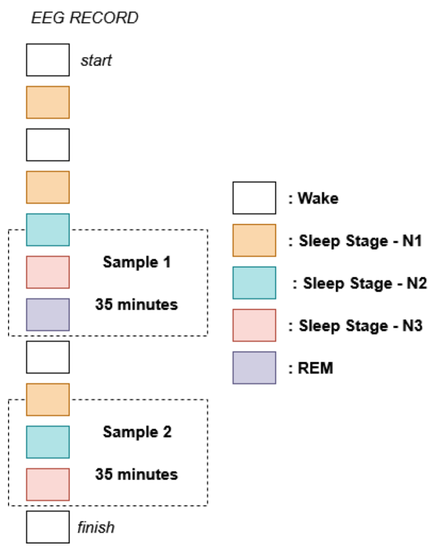

Fig. 6. Block diagram of dataset segmentation.

3) Dataset splitting

After the segmentation, the dataset is divided into 3 groups: training, validation, and testing. This division ensures that the dataset sufficiently fulfils the requirements of model design up to the testing phase. The training group comprises 70% or 42 samples, the validation group includes 10% or 6 samples, and the testing group consists of 20% or 12 samples.

C. Fast Fourier Transform (FFT)

Fast Fourier Transform (FFT) is one of the algorithms used to perform the calculation of Discrete Fourier Transforms (DFT) rapidly and efficiently. Fourier Transformation is a tool that can reconstruct periodic waveform signals using a series of harmonics [26]. Fourier Transforms can break down periodic waveform signals into their underlying harmonic components. Although FFT is not suitable for analyzing short-duration brain physiological signal waves, it is a method suitable for processing signals in the form of sine waves, such as EEG, because it is faster compared to other methods [24]. To evaluate the impact of FFT on our model performance, we performed two types of experiment. The first is an experiment using a normal dataset, and the second one using a dataset that undergoes feature extraction using FFT. The equation of FFT can be shown in Eq. (1).

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} \quad (1)$$

- $X(k)$ is the transformed signal in the frequency domain,
- $x(n)$ is the input signal in the time domain,
- N is the total number of samples,
- k is the frequency index,
- $e^{-j\frac{2\pi}{N}kn}$ represents the complex exponential basis functions.

The Fast Fourier Transform (FFT) is a computational algorithm used to efficiently transform a time-domain signal into its frequency-domain representation. This transformation enables the identification of the frequency components within the signal, facilitating further analysis and processing in various applications, including machine learning [11, 26].

D. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a distinct type of neural network. CNN is an artificial neural network utilized for recognition and processing activities [27]. The convolution layer, pooling layer, non-linearity layer, fully connected layer are the four primary parts of CNN. The convolution layer is an operational layer that integrates two sets of information. The input data undergoes filtration using a convolutional filter, yielding a feature map [27]. The non-linearity layer turns the input signal into a non-linear output signal. Examples of non-linear activation functions include sigmoid (logistic), Tanh, Rectified Linear Unit (ReLU), Exponential Linear Unit (ELU) and Parametric ReLU (PReLU) [28].

The pooling layer is the third primary component, and it is categorized into two types: Max Pooling and Average Pooling. The feature map is summarized by Max Pooling, which determines the highest value in the feature map. Conversely, Average Pooling describes the feature map by determining the average value extracted from the feature map [27]. Finally, the Fully Connected Layer is an essential part of Deep Neural Networks that is responsible for the development of predictions that are intended for use in regression or classification [28]. This research utilizes a Convolutional Neural Network (CNN) as the principal model to categorize brain physiological inputs into sleep apnea or normal classifications. The designed architecture combined fully connected layer, convolutional 1D layer, max pooling 1D, ReLU. Fig. 7 shows the designed architecture illustration.

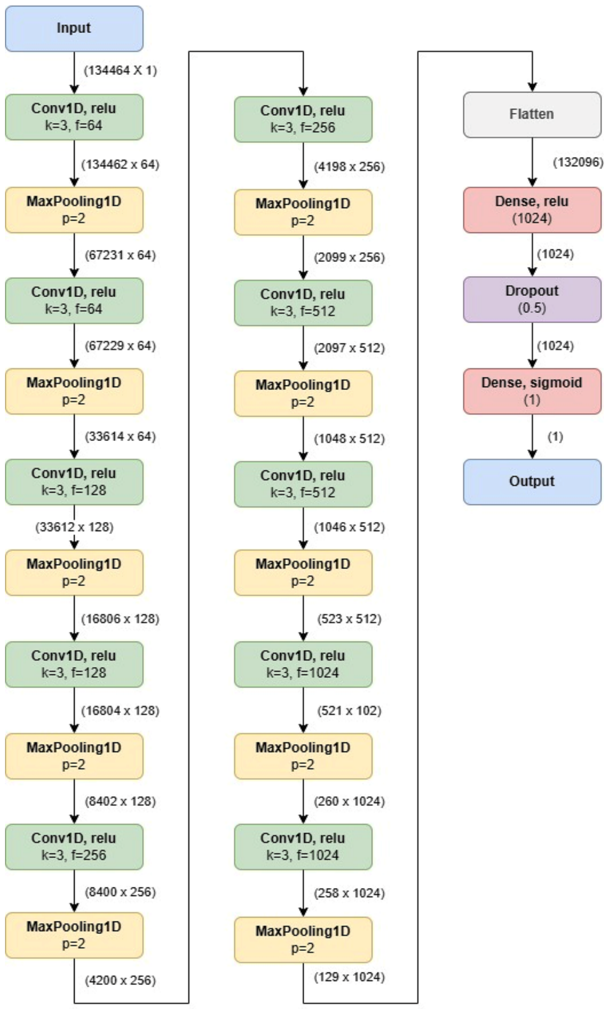

Fig. 7. The architecture of CNNs in deep learning model.

The CNNs model that is being employed, as illustrated in Fig. 6, comprises 10 layers of 1D convolutions, each of which is activated by a Rectified Linear Unit (ReLU). These layers are followed by max-pooling layers using a pool size of 2. The convolution layers are succeeded by flatten and dense layers, which are followed by a dropout layer and a final dense layer. The initial layer in the convolution layers is given an input of size (134464, 1), with a kernel size of 3 and a filter of 64. The kernel size of the second to tenth layers is 3, and the filters have values of 64, 128, 128, 256, 256, 512, 512, 1048, and 1048, respectively. Following the convolution layers, there is a Flatten layer. The flatten layer functions as a bridge within the convolutional layers and the fully connected layers, converting the multi-dimensional outcome of the convolutional layers into a one-dimensional format that is then passed to the fully connected layers [28]. After this, a Dropout layer is present, followed by a Dense layer with 16 neurons as well as Rectified Linear Unit (ReLU) activation. In order to prevent overfitting, the Dropout layer eliminates a portion of neurons during the training process. Lastly, a Dense layer with Sigmoid activation is implemented. Furthermore, PCA (Principal Component Analyst) may be implemented to mitigate the high dimensionality of the EEG dataset; however, it is inevitable that data loss will occur during the process. Consequently, the results of PCA may be insufficient [29, 30]. Consequently, we refrained from employing PCA in our methodology.

E. Evaluation

Confusion matrix, precision, recall, accuracy, and F1-Score were implemented as evaluation metrics in this investigation. The test dataset was employed to evaluate the final trained CNN model, which consisted of 12 data samples. These samples were divided into six normal samples and six sleep-disorder-breathing samples. The evaluation result then compared model 1 which trained using normal data, model 2 which trained using FFT applied data, and previous work that used MLP. While doing the evaluation, we did not employ any form of cross-validation techniques, this is due to our hardware limitations. Cross-validation techniques are known to be computationally expensive [31].

1) Confusion matrix

Confusion matrix usually used as a measure for determining the performance of model in classification or pattern recognition task [32]. By observing the diagonal line in the confusion matrix, we can determine whether the model is good or bad in classifying or recognizing pattern on the data [32]. The concept of the confusion matrix involves comparing the model's detection results with the actual detection results for each data. This comparison is categorized into True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) categories [33]. True Positives (TP) denote the count of positive samples accurately predicted by the model, True Negatives (TN) signify the count of negative samples accurately predicted by the model, False Positives (FP) indicate the count of negative samples erroneously predicted as positive by the model, and False Negatives (FN) represent the count of positive samples erroneously predicted as negative by the model [33].

2) Accuracy, recall, precision, F1-Score

The performance of trained models is evaluated using Accuracy, Recall, Precision, and F1-Score. Accuracy is a metric that determines the proportion of correctly predicted class instances compared to the total number of evaluated samples. Precision serves as a metric for determining the number of positive patterns that were accurately predicted out of all positive predictions in the positive class. Recall is a metrics that is employed to determine the number of positive patterns that are accurately classified. The F1-Score is the measure that determines the harmonic relationship between precision and recall. The F1-Score is most advantageous at 1 and most detrimental at 0. TN, TP, FN, and FP were utilized to calculate each of them, as represented in the confusion matrix. The formulas for calculating each metric are illustrated in Eqs. (2)–(5).

$$Accuracy = \frac{TP+TN}{(TP+TN)+(FP+FN)} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (5)$$

IV. Result and Discussions

The resulting CNN models for sleep apnea detection exhibited robust performance metrics. Model 1, trained on raw EEG signals, attained a training accuracy of 0.9762 and a validation accuracy of 1.000, with training and validation losses of 0.0849 and 0.0018, respectively. Model 2, employing FFT-based feature extraction for EEG signal modeling, attained a training accuracy of 0.9762 and a validation accuracy of 1.000, while demonstrating slightly elevated training and validation losses of 0.0945 and 0.1714, respectively. The training process of both models is depicted in Figs. 8 and 9. The figures use hyperparameter settings based on the 10-layer 1D CNN architecture, with ReLU activation functions in the hidden layers and a Sigmoid activation function in the output layer for the selected model.

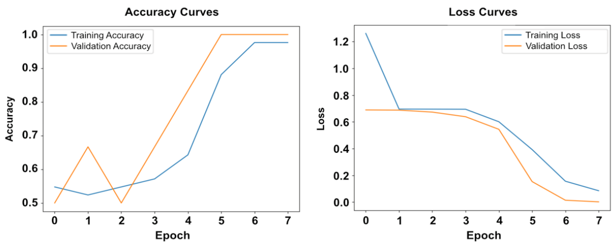

Fig. 8. Model 1 result on train and validation performance plot with the 10 layers of CNN 1D and ReLu Sigmoid: (left) accuracy (right) loss.

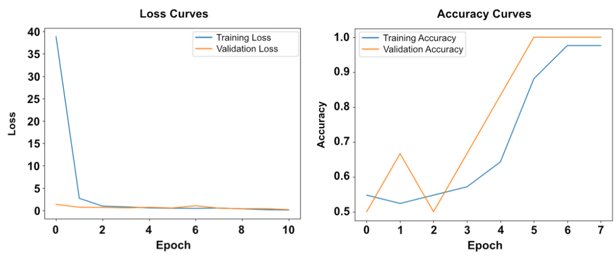

Fig. 9. Model 2 result on train and validation performance plot with the 10 layers of CNN 1D and ReLu Sigmoid: (left) accuracy (right) loss.

While the training and validation performance shows a similar result between model 1 and model 2, the performance plot on Fig. 8 and Fig. 9 shows that model 1 is more stable than model 2 in training and validation performance. Besides the training and validation performance, we also conducted evaluation using test data. The test results of the models are presented in confusion matrix and table form. Figs. 10 and 11 are the confusion matrix of both model 1 and model 2.

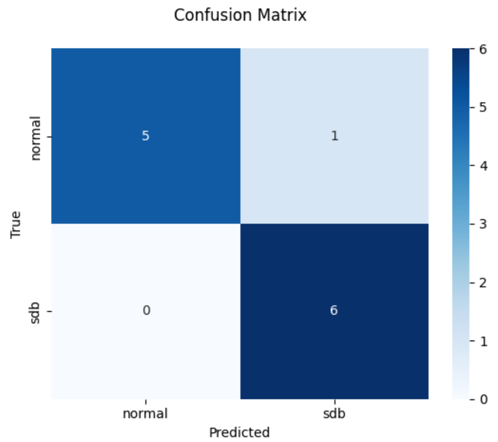

Fig. 10. Confusion matrix of model 1 (Raw EEG Signal).

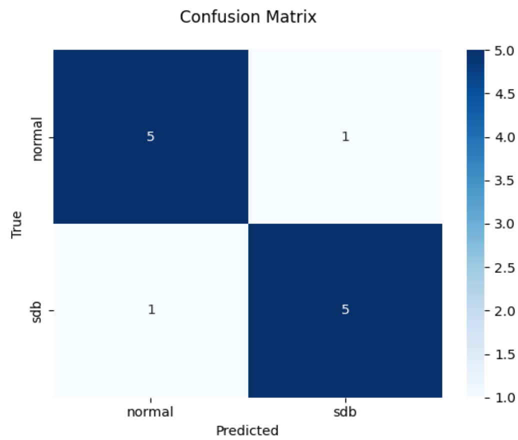

Fig. 11. Confusion matrix of model 2 (FFT-based feature extraction for EEG signal modeling).

Fig. 10 showed that model 1 almost has a perfect score, it only misses one sample, which is a normal sample classified as sleep-disorder-breathing. While Fig. 11 depicted two misclassified samples from both normal and sleep-disorder-breathing. From these confusion matrices, we can see briefly that model 1 is better than model 2. For better comparison of each model, we can see on Table I for the test result.

TABLE I. Test Result with 0 (Zero) Desibel

Model	Precision	Recall	F1-Score	Accuracy
EEG MLP [19]	-	-	-	0.86
Model 1 CNN	0.86	1.00	0.92	0.92
Model 2 CNN + FFT	0.83	0.83	0.83	0.83

As we can see in Table I with the same 0 db, the highest recall calculation value is found in model 1, where the recall reaches 1.00, indicating a low False Negative value or the absence of detected errors in actual labels. On the other hand, the lowest recall value is in the model where the recall is only 0.83, signifying a high value of False Negative or a significant number of errors in actual labels detected. As for the highest precision calculation value, the highest is model 1, with a precision of 0.86. This indicates a low value of False Positive, while model 2 has a False Positive value of 0.83. The highest F1 calculation value is observed in model 1, with an F1-Score of 0.92, signifying good precision and recall. Conversely, the lowest recall value is in the model where the recall is only 0.83, indicating poor precision and recall values. Similarly, the highest accuracy calculation value is found in model 1, where the accuracy reaches 0.92 beating the previous works that use MLP, and the lowest is in model 2, with an accuracy of only 0.83. From the results we can conclude that our model was better than the previous work using MLP [19]. While Model 1, trained on raw EEG data, demonstrated superior performance with a recall of 1.00 and an F1-Score of 0.92, Model 2, trained on FFT-based features, exhibited reduced performance metrics. This suggests that for sleep apnea detection within our dataset, the most discriminative features may not be primarily located in the frequency domain as isolated by the FFT. It is plausible that crucial information for accurate sleep apnea detection resides within the temporal dynamics and morphology of the raw EEG waveforms. The FFT, while effective for frequency analysis, might have inadvertently discarded or obscured these critical time-domain features that are essential for CNN to effectively learn and classify sleep apnea events. Furthermore, the direct use of raw EEG might allow the CNN to automatically learn complex time-frequency representations that are more nuanced and task-specific than those derived from a fixed FFT implementation.

In recent times, deep learning algorithms have been extensively employed across several domains to accomplish categorization or classification task [19]. It also applies to the healthcare and medical field. Where deep learning has been applied for classifying bone age, cervical pap smear, electrocardiogram signal noise, and others [1, 3, 34]. With extensive usefulness of deep learning on healthcare and medical field, in this paper we tried to apply deep learning for detecting sleep apnea using EEG signal. The same approach has been carried out using MLP showing a promising 86% accuracy as a result. The research findings align with prior studies utilizing deep learning CNN to analyze physiological signal data [15, 35]. The FFT-based model (Model 2)'s diminished performance may be attributed to the model's limited dataset size, which could limit its capacity to generalize. In order to improve the robustness of the model and alleviate the constraints identified in this research, future research should investigate alternative feature extraction techniques, such as wavelet transformations, and larger datasets [36].

TABLE II. COMPARISON WITH OTHER STUDIES

Model	Precision	Recall	F1-Score	Accuracy
Model 1 CNN	0.86	1.00	0.92	0.92
Moridani *et al.* [19] MLP	-	0.84	-	0.86

As we see on Table II, the same approach has been carried out using MLP showing a promising 86% accuracy as a result. The research findings align with prior studies utilizing deep learning CNN to analyze physiological signal data [15, 35]. The FFT-based model (Model 2)'s diminished performance may be attributed to the model's limited dataset size, which could limit its capacity to generalize. In order to improve the robustness of the model and alleviate the constraints identified in this research, future research should investigate alternative feature extraction techniques, such as wavelet transformations, and larger datasets [36].

The impact of the small sample size was mitigated by two important factors, even though the total number of original EEG recordings used in this study was limited to five subjects [37]. First, every EEG record was split into several 35-minute epochs, producing 60 samples total—30 assigned as "normal", and 30 as "sleep-disordered breathing". This segmentation technique assures that the CNN model is trained and assessed on a sufficiently large number of input instances, therefore enabling the model to learn temporal and morphological patterns efficiently across many segments [22]. Second, the dataset was carefully balanced between the two classes—normal and apnea—so avoiding class imbalance from distorting the performance or evaluation measures of the model [15]. Although the variety of topics is modest, the great number of uniformly spaced segments offers a rich training set for the CNN, therefore supporting significant learning and dependable performance evaluation within the parameters of this work [10, 21]. However, we recognize that the generalizability of the model could be further improved by broadening the subject pool in future research.

In this paper, we proposed a new deep learning approach for automatic detection of sleep apnea using EEG signal by utilizing CNN combined with and without FFT algorithm as a feature extraction algorithm for the dataset. This study emphasizes the efficacy of raw EEG data when processed with CNNs, attaining higher accuracy and recall, in contrast to previous studies that primarily relied on MLPs or FFT-enhanced features. This emphasizes the significance of customizing preprocessing techniques to suit particular applications [9, 11, 15, 19, 38]. The findings of this study demonstrate the potential of CNNs for sleep apnea detection using raw EEG signals, paving the way for developing portable, low-cost diagnostic tools. Such advancements could reduce reliance on traditional polysomnography, making sleep apnea diagnosis more accessible and efficient [39–41].

A significant limitation of this study is the relatively small size of our dataset. Training deep learning models, particularly Convolutional Neural Networks, effectively requires substantial amounts of data to learn robust and generalizable features. With a limited dataset, there is a heightened risk of overfitting, where the model learns to perform exceptionally well on the training data but fails to generalize new, unseen data. This means that while our model might achieve promising results on our specific dataset, its performance could degrade significantly when applied to different datasets or in real-world clinical settings [25].

This study recognizes the dataset's limitation regarding the number of original subjects. The division of EEG recordings into 35-minute epochs yielded 60 balanced samples, offering adequate data instances for training and assessment. The dataset was extensively chosen to guarantee equitable class representation and eradicate bias, while standardized performance criteria were employed to assess the model [42]. Although we acknowledge that larger datasets could enhance generalizability, the present results are consistent and offer a reliable basis for future research with more varied and comprehensive EEG datasets. To improve the generalization of our sleep apnea detection model, future work should prioritize the exploration of domain adaptation techniques. Specifically, adversarial learning presents a promising direction for training domain-invariant features and mitigating dataset bias. Beyond adversarial approaches, research could also investigate other domain adaptation methods such as domain-invariant feature normalization or transfer learning [19]. Implementing and evaluating these techniques will be critical steps towards developing a more robust and widely applicable sleep apnea detection system.

V. Conclusion

This study demonstrates the significant potential of Convolutional Neural Networks (CNNs) for the accurate detection of sleep apnea using EEG signals. Notably, a CNN model without Fast Fourier Transform (FFT) preprocessing achieved a 92% accuracy, surpassing both CNN models incorporating FFT and traditional Multi-Layer Perceptron (MLP) approaches. In contrast, Model 2 exhibited a slightly diminished performance, achieving an accuracy of 83% across key metrics. While FFT is theoretically beneficial for feature extraction, it did not yield performance improvements in our experiments. This suggests that, in the context of sleep apnea detection using our specific dataset, the most discriminative features may not reside solely in the frequency domain as extracted by FFT. The superior performance of the CNN models in comparison to previous studies that implemented Multi-Layer Perceptron (MLP) further emphasized the reliability and robustness of CNNs for this task. The evaluation underscores the potential of CNN-based systems as a foundation for portable and efficient sleep apnea diagnostic instruments, thereby reducing reliance on traditional, overpriced Polysomnography (PSG).

Nevertheless, there were a number of constraints encountered in the research that could affect the generalization and scope of its findings. For instance, the dataset used was very small-five records from the Physionet Sleep-EDF database-and this greatly limited the diversity of the training data. Besides, working with segmented EEG data instead of full-length records might have resulted in the lack of precious temporal patterns that are very important for sleep apnea detection. Another limitation is the age of the dataset since real-world EEG data may currently be different due to changes in diagnostic standards and recording techniques. Furthermore, FFT-based feature extraction did not improve the performance of the model, although it had a theoretical justification.

To enhance the model performance, exploring alternative feature extraction techniques beyond frequency domain analysis, such as time-domain features (statistical, morphological, nonlinear) or Wavelet Transform maybe beneficial. Additionally, optimizing the hyperparameters through systematic hyperparameter tuning are also crucial. Future research should include model evaluation on larger datasets. Additionally, alternative EEG signal preprocessing techniques can be used to improve the performance of the model.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

MKD provided supervision and guidance throughout the project, reviewed and edited the final version of the paper. ATM and RAR were responsible for data analysis, while AP and RAP authored the paper and conducted reviews and edits on the final version. RA executed the experiment. MRF and RFN were involved in research and data collection, and RA also handled data preprocessing and experiment execution. HH reviewed and edited the final version of the paper. All authors approved the final version.

Funding

This paper is supported by Ministry of Education, Culture, Research and Technology, Indonesia on behalf of the Center for Research and Community Service of Politeknik Negeri Jakarta number 450/PL3.18/PT.00.06/2023.

Acknowledgment

We would like to express our highest gratitude to Politeknik Negeri Jakarta for supporting this research through International Partnership Collaboration Research Grant Schema.

References

[1] S. S. Mostafa *et al.*, "A systematic review of detecting sleep apnea using deep learning," *Sensors*, vol. 19, no. 22, 4934, Nov. 2019.

[2] A. R. Spector, D. Loriaux, and A. E. Farjat, "The clinical significance of apneas versus hypopneas: Is there really a difference?" *Cureus*, vol. 1, no. 4, e4560, 2019.

[3] A. Sheta *et al.*, "Diagnosis of obstructive sleep apnea from ECG signals using machine learning and deep learning classifiers," *Appl. Sci.*, vol. 11, no. 14, 6622, July 2021.

[4] G. Pugliese *et al.*, "Sleep apnea, obesity, and disturbed glucose homeostasis: Epidemiologic evidence, biologic insights, and therapeutic strategies," *Curr. Obes. Rep.*, vol. 9, pp. 30–38, Jan. 2020.

[5] X. X. Lin *et al.*, "Rapidest: A framework for obstructive sleep apnea detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 387–397, 2023.

[6] X. Zhao *et al.*, "Classification of sleep apnea based on EEG sub-band signal characteristics," *Sci. Rep.*, vol. 11, no. 1, 5824, Mar. 2021.

[7] A. Ramachandran and A. Karuppiah, "A survey on recent advances in machine learning based sleep apnea detection systems," *Healthcare*, vol. 9, no. 7, 914, July 2021.

[8] A. M. Maitin *et al.*, "Survey of machine learning techniques in the analysis of EEG signals for Parkinson's disease: A systematic review," *Appl. Sci.*, vol. 12, no. 14, 6967, July 2022.

[9] M. K. Delimayanti *et al.*, "Evaluation of sleep stage classification using feature importance of EEG signal for big data healthcare," *Int. J. Adv. Intell. Inform.*, vol. 9, no. 3, pp. 418–428, Nov. 2023.

[10] M. K. Delimayanti *et al.*, "Implementation of convolutional neural network for COVID19 screening using x-rays images," in *Proc. 2023 International Conf. on Computer Science, Information Technology and Engineering (ICCoSITE)*, 2023, pp. 398–403.

[11] 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, 1797, Mar. 2020.

[12] M. Changizi and S. Rashidi, "Diagnosis of depression based on new features extractive from the frequency space of the EEG," in *Proc. 2023 13th International Conf. on Computer and Knowledge Engineering (ICCKE)*, 2023, pp. 67–72.

[13] M. Desai and M. Shah, "An anatomization on breast cancer detection and diagnosis employing Multi-layer Perceptron neural network (MLP) and Convolutional Neural Network (CNN)," *Clin. EHealth*, vol. 4, pp. 1–11, 2021.

[14] Z. Cai *et al.*, "Machine learning for multi-omics data integration in cancer," *iScience*, vol. 25, no. 2, 103798, Feb. 2022.

[15] M. H. Abdelaziz, W. A. Mohamed, and A. S. Selmy, "Hand gesture recognition based on electromyography signals and deep learning techniques," *J. Adv. Inf. Technol.*, vol. 15, no. 2, pp. 255–263, 2024.

[16] M. R. Faisal *et al.*, "3D word embedding vector feature extraction and hybrid CNN-LSTM for natural disaster reports identification," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 22, no. 5, pp. 1196–1208, Oct. 2024.

[17] R. A. Widodo, M. K. Delimayanti, and A. Wulandari, "Ddos attacks detection with deep learning approach using convolutional neural network," *J. Appl. Inform. Comput.*, vol. 8, no. 2, pp. 235–240, Aug. 2024.

[18] R. F. Naryanto *et al.*, "Fault detection in diesel engines using artificial neural networks and convolutional neural networks," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 2, pp. 680–692, 2024.

[19] M. K. Moridani, M. Heydar, and S. S. J. Behnam, "A reliable algorithm based on combination of EMG, ECG and EEG signals for sleep apnea detection: A reliable algorithm for sleep apnea detection," in *Proc. 2019 5th Conf. on Knowledge Based Engineering and Innovation (KBEI)*, 2019, pp. 256–262.

[20] S. K. Golla and S. Maloji, "Maximum Overlap Discrete Transform (MODT)—Gaussian Kernel Radial Network (GKRN) model for epileptic seizure detection from EEG signals," *J. Adv. Inf. Technol.*, vol. 14, no. 5, pp. 883–891, 2023.

[21] L. D. Barnes *et al.*, "Detection of sleep apnea from single-channel Electroencephalogram (EEG) using an explainable Convolutional Neural Network (CNN)," *PLOS ONE*, vol. 17, no. 9, e02721679, Sep. 2022.

[22] M. A. Bravo-Ortiz *et al.*, "Preprocessing strategy to improve the performance of convolutional neural networks applied to steganalysis in the spatial domain," *J. Adv. Inf. Technol.*, vol. 15, no. 1, pp. 33–39, 2024.

[23] H. M. Mohan and S. Anitha, "Real time audio-based distress signal detection as vital signs of myocardial infarction using convolutional neural networks," *J. Adv. Inf. Technol.*, vol. 13, no. 2, pp. 106–116, 2022.

[24] A. Yudhana *et al.*, "Human emotion recognition based on EEG signal using fast fourier transform and k-nearest neighbor," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 5, no. 6, pp. 1082–1088, 2020.

[25] M. G. Terzano *et al.*. (November 2001). CAP sleep database. *physionet.org*. [Online]. 2(6). pp. 537–553. Available: https://www.physionet.org/content/capslpdb/1.0.0/

[26] H. C. Lin and Y. C. Ye, "Reviews of bearing vibration measurement using fast Fourier transform and enhanced fast Fourier transform algorithms," *Adv. Mech. Eng.*, vol. 11, no. 1, pp. 1–12, Jan. 2019.

[27] J. Teuwen and N. Moriakov, "Convolutional neural networks," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, New York: Academic Press, 2020, pp. 481–501.

[28] A. Dhillon and G. K. Verma, "Convolutional neural network: A review of models, methodologies and applications to object detection," *Prog. Artif. Intell.*, vol. 9, no. 2, pp. 85–112, Dec. 2019.

[29] Y. Xie *et al*., "Least angle sparse principal component analysis for ultrahigh dimensional data," *Annals of Operations Research*, pp. 1–27, 2024. https://doi.org/10.1007/s10479-024-06428-0

[30] A. Palummo *et al.*, "Functional principal component analysis for incomplete space–time data," *Environ. Ecol. Stat.*, vol. 31, pp. 555–582, Mar. 2024.

[31] P. Burman, "A comparative study of ordinary cross-validation, r-fold cross-validation and the repeated learning-testing methods," *Biometrika*, vol. 76, no. 3, pp. 503–514, Sep. 1989.

[32] I. Düntsch and G. Gediga, "Confusion matrices and rough set data analysis," *J. Phys. Conf. Ser.*, vol. 1229, no. 1, 012055, 2019.

[33] I. Markoulidakis *et al.*, "Multi-class confusion matrix reduction method and its application on net promoter score classification problem," in *Proc. 14th PErvasive Technologies Related to Assistive Environments Conf.*, 2021, pp. 412–419.

[34] K. S. Lee *et al.*, "Evaluation of scalability and degree of fine-tuning of deep convolutional neural networks for COVID-19 screening on chest x-ray images using explainable deep-learning algorithm," *J. Pers. Med.*, vol. 10, no. 4, 213, Nov. 2020.

[35] H. C. Bhandari *et al.*, "Exploring non-euclidean approaches: A comprehensive survey on graph-based techniques for EEG signal analysis," *J. Adv. Inf. Technol.*, vol. 15, no. 10, pp. 1089–1105, 2024.

[36] A. Ahmadi, H. Bazregarzadeh, and K. Kazemi, "Automated detection of driver fatigue from electroencephalography through wavelet-based connectivity," *Biocybern. Biomed. Eng.*, vol. 41, no. 1, pp. 316–332, 2021.

[37] M. K. Delimayanti *et al.*, "Accurate deep learning-based sleep apnea detection from cardiac physiological signals," in *Proc. 2023 IEEE Asia Pacific Conf. on Wireless and Mobile (APWiMob)*, 2023, pp. 284–289.

[38] P. Huynh, G. Warner, and H. Lin, "Effects of EMD and feature extraction on EEG analysis," *J. Adv. Inf. Technol.*, vol. 11, no. 1, pp. 26–34, 2020.

[39] W. Aribowo, "Optimizing feed forward backpropagation neural network based on teaching-learning-based optimization algorithm for long-term electricity forecasting," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 1, pp. 11–20, 2022.

[40] S. Suryawanshi and S. B. Patil, "Efficient brain tumor classification with a hybrid CNN-SVM approach in MRI," *J. Adv. Inf. Technol.*, vol. 15, no. 3, pp. 349–354, 2024.

[41] A. Anbumozhi and A. Shanthini, "Gldiccnn model: Groundnut leaf diseases identification and classification for multiclass classification using deep learning," *J. Adv. Inf. Technol.*, vol. 15, no. 7, pp. 812–821, 2024.

[42] G. B. Moody, "PhysioNet: Research resource for complex physiologic signals," *Jpn. J. Electrocardiol*, vol. 29, pp. 1–3, 2009.