Detection of OSA Through the Application of Deep Learning on Polysomnography Data

Hasan Ulutas[1], Recep Sinan Arslan[2], Muhammet Emin Sahin[1], Halil Ibrahim Cosar[3,*], Cagri Arisoy[1], Ahmet Sertol Koksal[1], Mehmet Bakir[1], Bulent Ciftci[4]

[1]*Department of Computer Engineering, Faculty of Engineering and Architecture, Yozgat Bozok University, Yozgat, 66200, Turkiye*

[2]*Department of Computer Engineering, Faculty of Engineering, Architecture and Design, Kayseri University, Kayseri, 38000, Turkiye*

[3]*Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Yozgat Bozok University, Yozgat, 66200, Turkiye*

[4]*Department of Chest Diseases, Faculty of Medicine, Yuksek Ihtisas University, Ankara, Turkiye*

*hasan.ulutas@yobu.edu.tr; recepsinanarslan@kayseri.edu.tr; emin.sahin@yobu.edu.tr; *halil.cosar@yobu.edu.tr; cagri.arisoy@yobu.edu.tr; ahmet.koksal@yobu.edu.tr; mehmet.bakir@yobu.edu.tr; bulent.ciftci@yobu.edu.tr*

***Abstract*—This paper presents a comprehensive study on the application of deep learning techniques to accurately detect sleep apnea. The study leverages a dataset obtained from the Sleep Laboratory of the Department of Chest Diseases of Yozgat Bozok University, with the aim of developing an effective decision support system capable of identifying cases of sleep disorders with high accuracy. The proposed methodology focusses on the use of deep neural networks (DNNs) to enhance the accuracy and reliability of sleep apnea detection. By employing meticulous data collection, preprocessing, and analysis, the study demonstrates the potential of DNNs to capture intricate and high-dimensional features within complex sleep data, allowing precise and reliable diagnosis. The experimental results showcase the effectiveness of the proposed DNN-based classifier design, achieving an accuracy of 96.48 %. The study's contributions lie in the enhancement of sleep disorder diagnosis through the integration of deep learning techniques, offering promising implications for clinical practice. Early detection of sleep disorders has the potential to significantly improve patient outcomes and overall quality of life and lays the foundation for further advancements in the field of sleep medicine.**

***Index Terms*—Apnea; DNN; Classification; Preprocessing; Sleep disorder.**

I. INTRODUCTION

Obstructive sleep apnea (OSA) is one of the most important sleep-disordered breathing syndromes seen in the upper airways during sleep [1]. This syndrome is characterised by snoring and increased respiratory efforts to overcome resistance to the upper airways [2]. According to previous related works, more than one billion people are affected by OSA [3]. OSA patients are at risk for heart-related diseases that can lead to death [4]. Therefore, early diagnosis and treatment are essential for public health [5]. The guidelines of the American Academy of Sleep Medicine (AASM) are the gold standard method for OSA detection applied in sleep clinics [6]. Electrocardiogram (ECG), electroencephalogram (EEG), oxygen saturation, respiratory effort, and airflow sensors are connected to polysomnography (PSG), and the whole night's sleep data of a patient are recorded for analysis and sleep scoring [7]. Since every patient's data are composed of around 800 epochs (30 seconds each), analysing sleep and calculating the apnea-hypopnea index (AHI) is a time-consuming process that must be performed by a sleep doctor or sleep expert [8]. The apnea-hypopnea index can be calculated by 60 × (apneas + hypopneas)/total sleep in minutes. In recent years, the use of artificial intelligence in biomedical applications has grown extensively [9]–[11]. Over the last decade, computer engineers have engaged in productive collaborations with sleep doctors to advance sleep stage scoring, the identification of obstructive sleep apnea, and the accurate calculation of the AHI [12]–[16]. In pursuit of this goal, certain researchers adopted machine learning techniques, such as the hidden Markov model, for the purpose of feature extraction and classification [17]. Machine learning techniques such as supervised, unsupervised, and reinforcement learning can be employed in OSA detection [18]. Most studies are based on supervised learning. Deep learning approaches are considered subtypes of machine learning [19]. They can be categorised into several types, including deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks. With the proliferation of methods, deep learning-based models are being increasingly used for OSA detection [20], [21]. Although deep learning approaches have shown improvements in detecting OSA, they still have some limitations, including the use of a single channel for detection, achieving only limited levels of accuracy, and relying on publicly available datasets [22]. In this study, we

Manuscript received 14 July, 2024; accepted 6 November, 2024.

have employed a feedforward neural network with multiple hidden layers as a typical example [23] to detect OSA events in PSG data. Deep learning enables the model to learn relevant features directly from raw data through neurons, convolution, and pooling layers. Several works have been done to detect sleep apnea using deep learning. Morillo and Gross [24] used neural networks as a classifier to analyse OSA using pulse oximetry. Khondoker, Gubbi, and M. Palaniswami [25] decomposed the ECG signal into eight levels of detailed coefficients. Then, a feedforward neural network structure with a hidden layer of 30 neurons was fed into the event detection stage. Ozdemir *et al.* proposed a support vector machine and deep learning classifier that uses respiratory signals for OSA detection. According to previous works, our proposal differs from similar studies in having its own dataset, higher accuracy values, and using all channels instead of a single one [8]. OSA datasets are generally composed of extremely large data, sometimes with tens of millions of lines for each patient. Since our model is based on supervised learning and the supervisor plays an important role in classification and marking OSA and non-OSA events, we prepared our dataset from Yozgat Bozok University Sleep Clinics, which is composed of data from 50 patients. To compare the proposed model with similar works, Section II is prepared. The training and testing phases are the two main phases for generalisation capability. To validate our model, some of the patients' data have been used for testing. Therefore, data preprocessing must be performed accordingly to automatically discriminate episodes of night apnea from normal sleep periods. Data preparation, model creation, and testing methods are explained in Section III. The results obtained are presented in Section IV, and the conclusions and discussion phase are given in Section V.

The primary contributions of this study are as follows.

– One of the key contributions of this study is the acquisition of an original and diverse dataset from the Sleep Laboratory of the Yozgat Bozok University Department of Chest Diseases. This dataset includes sleep data from 50 patients, captured through 19 distinct sensors, offering a comprehensive overview of sleep patterns and disorders. The uniqueness of this dataset adds value to the study by presenting real-world scenarios and allowing for more accurate and relevant model training and testing.

– Addressing the challenge of class imbalance, this study contributes by employing undersampling techniques based on the "majority" class. This strategy effectively balances the dataset, ensuring that the model is trained and tested on representative instances of each class.

– This study contributes by showcasing the successful application of deep learning techniques, specifically deep neural networks (DNNs), in the field of detection of sleep apnea. The use of DNN offers a novel approach to accurately classify cases of sleep disorders, enriching the landscape of diagnostic tools available to healthcare professionals.

– By achieving an impressive accuracy rate of 96.48 % in the detection of sleep apnea, this study contributes to improving the diagnostic accuracy of sleep disorders. The proposed model's ability to consistently deliver accurate results in diverse patient profiles reflects its potential to enhance clinical decision making and patient care.

– The results and insights presented in this study contribute by guiding future research directions in the detection of sleep disorders.

II. METHODOLOGY

A. Dataset

The sleep dataset was provided by the Sleep Laboratory at the Department of Chest Diseases of Yozgat Bozok University. The ethical permissions required for data collection were duly obtained. The sleep data collected comes from patients of different ages and genders. Their general distribution has been determined to occur depending on the number of patients coming to the hospital and the prevalence of the disease in the community. No patient selection was made with respect to a certain age and gender in the creation of the dataset, and it includes the data of all patients admitted to the hospital within a certain period. In the training and testing stages of this study, data from 50 patients were used. In a nutshell, as shown in Table I, there are 761,000 Apnea records, approximately 1.8 million Hypopnea records, and around 226 million normal records for 50 patients. However, a substantial number of normal records were omitted. Even if an individual has severe apnea, instances of displaying apnea or hypopnea symptoms during an eight-hour sleep session are comparatively scarce compared to the normal state. This circumstance negatively impacts both the learning and classification phases of the model's training, leading the trained model to exhibit a tendency to label all cases as normal during testing. To resolve this issue and rectify the considerable class imbalance, undersampling was carried out based on the "majority" class, resulting in a reduction in the number of normal instances for each patient. As a result, the average number of normal instances per patient, initially around 4.5 million, was reduced to 12,000 instances. This strategy effectively rectified the sample imbalance across classes.

TABLE I. DATASET DETAILS.

Before Undersampling Data Size				
Patient ID	**Apnea**	**Hypopnea**	**Normal**	**Total**
Total Data	761000	1843200	226651000	229255200
Average	15220	36864	4533020	4585104
After Undersampling Data Size				
Patient ID	**Apnea**	**Hypopnea**	**Normal**	**Total**
Total Data	761000	1843200	617000	3221200
Average	15220	36864	12340	64424

Even if humans have severe apnea during sleep due to their nature, the number of epochs in which they show signs of apnea in an eight-hour sleep period is quite low and this is reflected in the collected data. When the average values were examined, it was observed that the number of apnea and hypopnea samples was very low compared to the normal period. This will result in the following.

1. Biased learning situation. In other words, while it will not have difficulty recognising the common class because it creates a bias against the majority class, it will have difficulty recognising the less common class. This situation will have a positive impact on the classification performance in terms of our problem, but it will cause difficulty in recognising the disease, as it will have

difficulty in determining the critical apnea and hypopnea situations.

2. Misleading accuracy situation will occur. In other words, the overall performance of the model will be high because it recognises the class well with a large number of samples, but it will fail to evaluate the disease states that need to be determined. For these reasons, it is necessary to solve the problem of sleeping sickness with a serious data imbalance. Various methods such as oversampling, undersampling, ensemble learning, and error-correction mechanism can be applied to solve this problem. In terms of this problem, if oversampling is performed, synthetic data must be generated 30 times for each minority class example, and this causes a serious overfitting problem. In ensemble learning models, approximately two-thirds of the data are reserved for training, but this is not enough to balance the data for the sleep apnea problem, where there is an approximately 30-fold difference between classes. With error-correction mechanisms, erroneous situations caused by convergence can be corrected. However, this did not make it possible to avoid overfitting for this problem with serious data imbalance. For these reasons, the undersampling method was used in this study, considering data loss.

B. Data Preprocessing

The initial and most extensive phase of the project involved collecting sleep data and having it evaluated by a specialised sleep doctor. For this research, sleep information was collected from 50 patients who visited the Yozgat Bozok University Sleep Centre. Data were captured using 19 distinct sensors, including seven EEG, three EMG, one ECG, two eye movement, one SpO2, two chest effort, one thermistor, one pressure sensor, and one body position sensor. These sensors were used to comprehensively assess each patient's sleep patterns and conditions. Subsequently, the manual evaluation of the obtained data constitutes the subsequent stage in the data preparation process. The involvement of sleep specialist doctors in this step is essential, given its utilisation in the learning model. The responsibility for conducting this phase was undertaken by the doctor involved in the study. The evaluation encompassed five distinct stages (WK, N1, N2, N3, and REM) according to the AASM guidelines. As stipulated in the standard protocol, a sleep duration of 30 seconds was defined as one epoch, and an average sleep span of 6 to 6.5 hours was marked as 800 epochs. Several preprocessing procedures were required to prepare this dataset for use in network-based modelling frameworks aimed with automated scoring. This necessity arises due to variations in sampling frequencies across diverse sleep sensors. A comprehensive explanation of this aspect can be found in the publication in [16], and analogous techniques are implemented in the current study.

C. DNN

Deep neural networks (DNNs), especially convolutional neural networks (CNNs), have been widely used in various image classification tasks, showing significant performance improvements since 2012 [26], [27]. Traditional neural networks, with only one hidden layer, are considered "shallow" models. Although they can effectively solve simple problems, they often struggle with complex and high-dimensional data because of their limited capacity to learn from the data. In contrast, DNNs consist of multiple hidden layers, which enhances their ability to learn higher-level and abstract features. This makes them capable of extracting more meaningful and abstract features from complex data structures. Consequently, DNNs are better suited for large and intricate datasets, achieving superior results.

DNNs have proven to be effective in detecting sleep apnea using wearable technologies. Various types of DNN, including deep transfer learning models, have been used to classify sleep stages and detect SA using wrist-worn consumer sleep technologies (CST) [28]. These models process accelerometer and photoplethysmography (PPG) signals, training on clinical datasets and validating on both clinical and wrist-worn device data. Although the results showed significant performance improvements when trained on clinical data, the models performed slightly less well on wearable device data compared to clinical data.

Another study in this field presents an algorithm to detect sleep apnea using deep learning models on single-lead ECG signals. The proposed LSTM-CNN model combines CNN and long short-term memory (LSTM) networks to extract spatial and temporal features from ECG signals, automatically distinguishing apnea events from normal segments. Tested on the Apnea-ECG and UCDDB datasets, the model achieved high accuracy, sensitivity, and specificity, with average accuracies of 97.21 % per segment and 100 % per recording on the Apnea-ECG dataset, and 93.70 % accuracy, 90.69 % sensitivity, and 95.82 % specificity on the UCDDB dataset [29].

D. Evaluation Metrics

In sleep apnea detection studies, various evaluation metrics such as accuracy, precision, recall (sensitivity), and F-score are used to assess the performance of artificial intelligence-based machine learning and deep learning algorithms. These metrics play a crucial role in evaluating the algorithms' accuracy, sensitivity and balance [9]. To calculate these performance metrics, a confusion matrix is used. Accuracy represents the ratio of correctly predicted instances to the total number of examples, while precision represents the ratio of true positive predictions among all positive predictions. Recall (or Sensitivity) represents the proportion of true positive predictions among all actual positive instances. F-Score is a measure that combines precision and recall to provide a balanced assessment.

III. EXPERIMENTAL RESULTS

In our experimental setup, we used a high-performance workstation to process and analyse the collected data. The workstation was equipped with 64 GB of RAM, an Intel Xeon Silver 4114 CPU operating at 2.2 GHz with 40 cores, and an HP Z6 G4 server. Additionally, the system featured an NVIDIA GeForce RTX 3080 Ti graphics card, providing substantial computational power for intensive data processing tasks.

The operating system used was Ubuntu 20.04.5 LTS, which ensured a stable and efficient environment for running our data analysis and machine learning algorithms. Our aim is to achieve high classification performance with the proposed model and develop an effective decision support system capable of detecting sleep disorders. To this end, the designed model architecture is comprehensively illustrated in Fig. 1. This model is primarily built upon the DNN method. The DNN method allows for a thorough processing of the data, aiding in a better comprehension of the intricacies and features of the sleep data. Consequently, it is possible to detect and classify the indicators of sleep disorders with greater precision. The flexible structure of the DNN method, owing to its ability to capture deep relationships and patterns within data, particularly empowers us to attain effective outcomes in such intricate and multidimensional datasets. During the training phase of our model, the DNN method's learning capacity has been enhanced through the use of large and diverse datasets. This enables better learning and understanding of the wide range of features embedded in the sleep data.

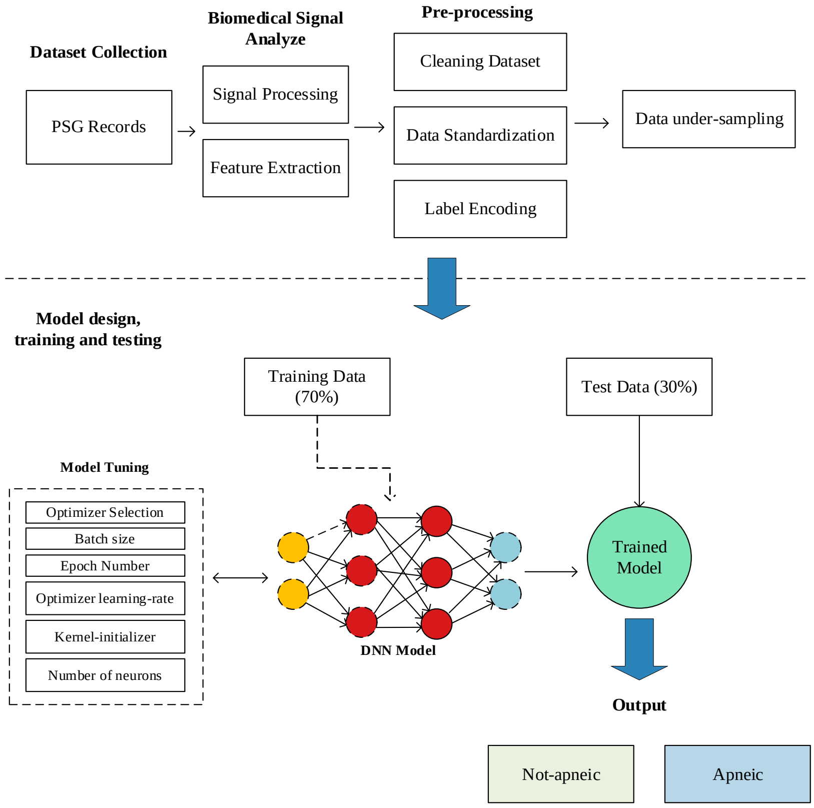

Fig. 1. Designed model architecture.

In this study, both the training of deep learning models and all testing processes are performed with a desktop computer equipped with 128 GB RAM and Intel(R) Processor Xeon W-1350P (6C/12T, 4.0/5.1 GHz, 12 MB). Experimental studies were carried out using the Python programming language and Keras library environment and matplotlib, sklearn, imblearn, numpy, pandas libraries. The deep learning neural network model used is shown in Fig. 2. The model has one input, hidden layers, and one output layer. The number of neurons in each layer is shown in the figure. There are 64 neurons in the input layer, 32, 16, 8 neurons in the hidden layers and two neurons in the output layer, respectively. "reLu" activation function is used in the input layer and hidden layers and "sigmoid" activation function is used in the output layer. "Adam" was selected as the optimiser. "uniform" was used as kernel initialiser. To avoid overfitting in the model, early stopping is performed, and any dropout layer is used.

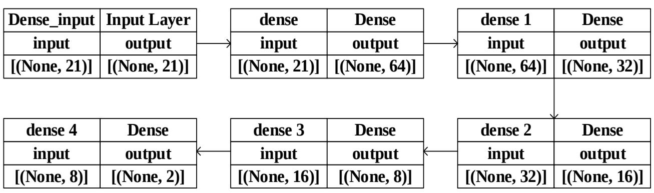

Fig. 2. Architecture of proposed DNN model.

For the determination of the structure of the DNN model and parameter selection, the grid search algorithm is used to automatically select the parameters at certain intervals, and the selected values are given in Table II. Within these selections, the highest classification performance is achieved. Early stopping is used to prevent the model from overfitting. With a proactive approach, the aim is to increase the security of the model by intervening and stopping a possible overfitting at an early stage. The model uses sigmoid as an activation function and man as optimiser in the training process. The learning rate is chosen as 0.001 and the batch size is set to 50. Uniform is used as a kerner_initilizer at the beginning of the model. Model training is terminated at 100 epochs. As a result, with these choices, the objective is to keep the performance of the DNN model at the best value.

In this study, upon analysis of the average, maximum, and minimum outcome values obtained for the proposed two-layer classifier design developed using the DNN deep learning model for 50 patients, it is evident that the suggested model outperforms other classifiers in terms of all parameters, including accuracy, precision, recall, and F-score. These results include the highest and lowest accuracy rates among patients, as well as the average accuracy among all patients, and are comprehensively presented in Table III. Due to the utilisation of deep learning models in the context of deep learning, the proposed model has gained the ability to extract complex features from the data. This ability is crucial to accurately discern subtle differences between cases of sleep disorders and normal conditions. As a result, the model, through its enhanced ability to understand subtle variations and patterns within the data, achieves higher precision, recall, and F1-score values.

TABLE II. HYPERPARAMETERS OF THE PROPOSED SYSTEM.

Parameter	Tuning Range	Selected Parameters for Each DL Models
Kernel_initializer	Uniform, lecun_uniform,normal, zero, glorot_normal, glorot_uniform, he_normal, he_uniform	uniform
Optimizer	SGD, RMSProp, Adagrad, Adadelta, Adam, Adamax, Nadam	DNN: adam
Learning_rate	[0.001, 0.005, 0.008, 0.01, 0.02, 0.1, 0.3]	0.001
Batch_size	[5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100]	50
Epoch	[10, 20, 50, 100, 150, 160, 170, 200, 300]	100
Neuron activation function	Softmax, softplus, softsign, relu, tanh, sigmoid, hard_sigmoid, linear	DNN: sigmoid
Number of neurons	[1, 5, 10, 15, 20, 50, 100, 200, 500]	DNN: 64, 32, 16, 8, 2
Early stopping (patient)		32 epochs

TABLE III. EVALUATION METRICS RESULTS OF PROPOSED SYSTEM FOR PATIENTS.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Patient 1 (Max)	0,9835	0,9895	0,9774	0,9834	0,9835
Patient 2 (Min)	0,9320	0,9320	0,9310	0,9320	0,9317
All Patients (Avg.)	0,9648	0,9769	0,9519	0,9641	0,9648

In terms of accuracy, the proposed model demonstrates exceptional performance in different patient scenarios. For the patient achieving the highest accuracy, the model achieves an impressive accuracy rate of 98.35 %, alongside noteworthy precision, recall, and F-score values of 98.95 %, 97.74 %, and 98.34 %, respectively, highlighting its proficiency in accurately identifying the cases of sleep disorders. Even for the patient with the lowest accuracy, the model maintains a commendable accuracy rate of 93.20 %, showcasing consistent performance. This robustness is further substantiated by precision, recall, and F-score metrics of 93.20 %, 93.10 %, and 93.20 % respectively. When considering the collective patient group, the model's performance remains strong, with an average accuracy of 96.48 %, accompanied by precision, recall, and F-score values of 97.69 %, 95.19 %, and 96.41 %, respectively. These outcomes underscore the model's reliability in consistently delivering accurate results in diverse patient profiles. The confusion matrix, accuracy loss, and area under the curve (AUC) for patients with the highest and lowest accuracy rates obtained by the classification of apnea using DNN method are presented in Figs. 3 to 10.

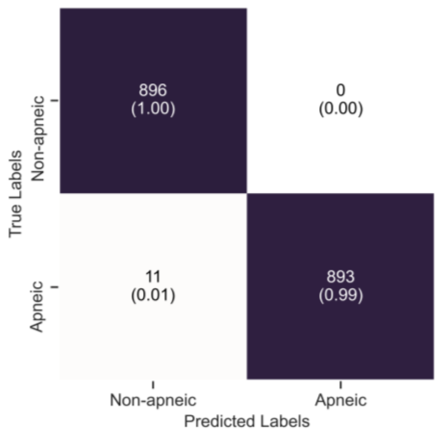

Fig. 3. Detailed confusion matrix of patient 1.

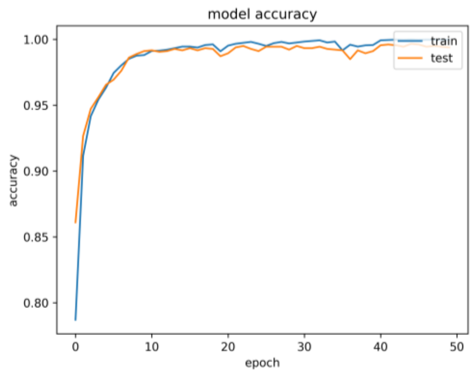

Fig. 4. Accuracy graph of patient 1.

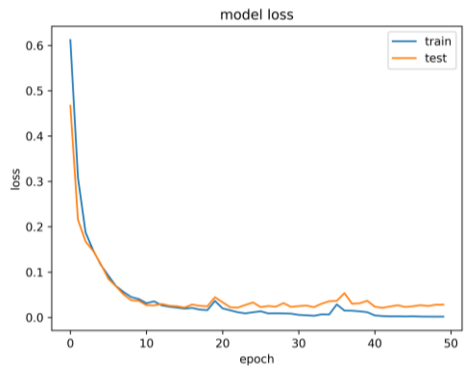

Fig. 5. Loss graph of patient 1.

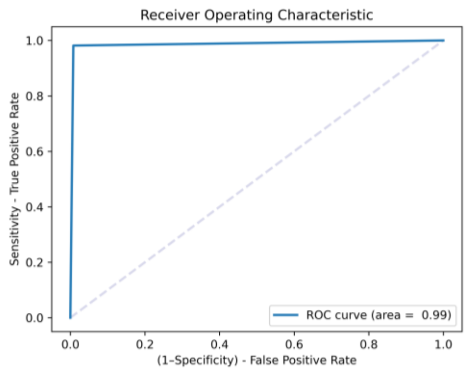

Fig. 6. ROC graph of patient 1.

Upon examining the confusion matrix in Fig. 3, it can be observed that the designed model misclassifies some apnea patients as nonapnea, but accurately distinguishes nonapnea cases. When examining the accuracy and loss graphs provided in Figs. 4 and 5, a consistent pattern of change in accuracy and loss over time is clear.

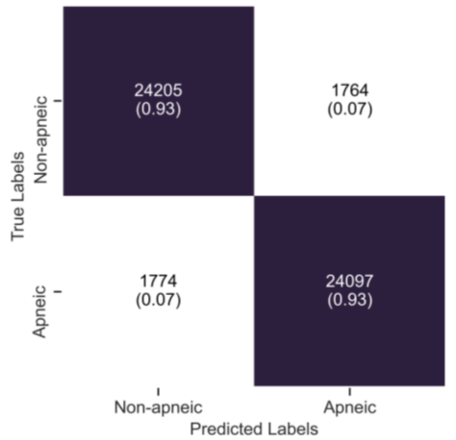

Fig. 7. Detailed confusion matrix of patient 2.

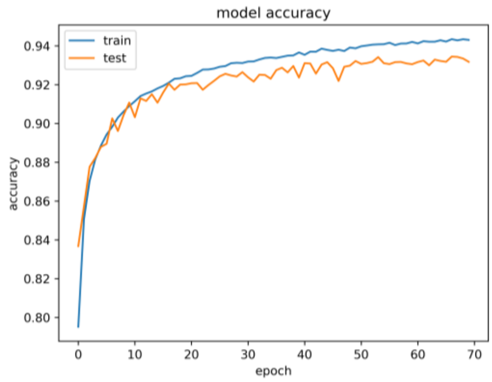

Fig. 8. Accuracy graph of patient 2.

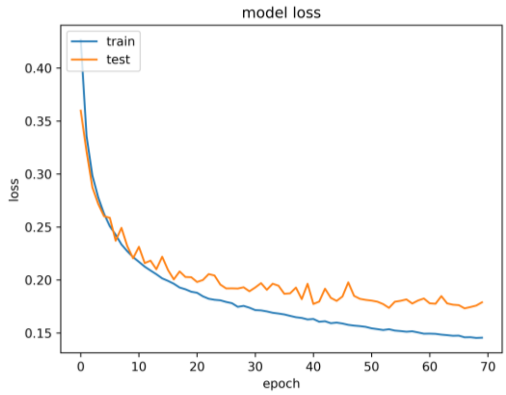

Fig. 9. Loss graph of patient 2.

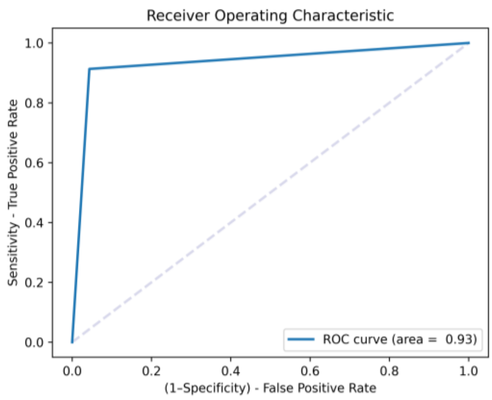

Fig. 10. Loss graph of patient 2.

Considering the test results of the study, it is evident that the proposed model does not suffer from overfitting issues during apnea detection and consistently achieves high levels of accuracy. Furthermore, the designed model underwent early stopping during the training phase, with an early stopping value of 32 epochs. This was implemented as a precaution against overfitting and to enhance the generalisability of the model.

IV. Discussion

This work focusses on the application of deep learning methodologies in the field of detection of sleep apnea. It was conducted using a comprehensive dataset obtained from Yozgat Bozok University Department of Chest Diseases Sleep Laboratory. Ethical approvals are obtained and valuable insights about the performance of the DNN model are obtained by performing training and testing phases on data obtained from 50 patients. The results of the study highlight the effectiveness of the DNN-based classifier design in accurately classifying cases of sleep apnea. Efforts to reduce the imbalance between normal and apnea cases come to the fore during the data preprocessing stage. This affects the training and evaluation phases of the model, as apnea cases are less common than normal cases. The undersampling strategy corrected for this imbalance and contributed to a more balanced training process. Despite the reduction in the dataset, the proposed model demonstrated superior performance and reaffirmed its ability to accurately detect sleep disorders. Deep learning architecture, specifically the DNN method, proved to be a suitable choice for this study due to its ability to learn complex and high-level features from complex data structures. The potential of DNN in the field of sleep apnea detection is encouraging advances in this field. However, the limitations of the study and areas on which future research could focus should also be considered. Over the last decade, numerous studies have focussed on detecting sleep apnea, and research continues today. Table IV highlights a selection of these studies.

TABLE IV. SIMILAR WORKS RELATED TO SLEEP APNEA DETECTION.

Reference	Method	Dataset	Input Type	Classifiers	Accuracy
Surrel, Aminifar, Rincón, Murali, and Atienza 2018 [30]	ML	PhysioNet Apnea-ECG	Single channel ECG	SVM	88.2 % (max)
Li, Pan, Li, Jiang, and Liu 2018 [19]	DL	Apnea-ECG dataset	Single channel ECG	Decision fusion	85 %
Wang. Lu, Shen, and Hong 2019 [21]	DL	PhysioNet Apnea-ECG	Single channel ECG	LeNet-5	87.6 %
Erdenebayar, Kim, Park, Joo, and Lee 2019 [11]	DL	Own dataset	Single channel ECG	CNN	93.1 %
Shen, Qin, Wei, and Liu 2021 [22]	DL	Apnea-ECG database	ECG	CNN + weighted-loss time-	89.4 %

Reference	Method	Dataset	Input Type	Classifiers	Accuracy
				dependent classification	
Sheta *et al*. 2021 [31]	DL and ML	Physionet's CinC challenge-2000 database	ECG signals	CNN + LSTM	86.2 %
Yang, Zou, Wei, and Liu 2022 [32]	DL	Apnea-ECG dataset	ECG	Multi-model fusion	90.3 %
Wang, Xiao, Fang, Li, Wang, and Zhao 2022 [33]	DL	Own dataset	EEG channels	LSTM	92.7 %
Zarei, Beheshti, and Asl 2022 [29]	DL	Apnea-ECG and UCDDB	ECG	CNN + LSTM	97.2 %
Hu *et al*. 2024 [34]	DL	PAEDB	Single Lead ECG	Hybrid Transformer	90.3 %
Our study	**DL**	**Own dataset occupied from 50 patients**	**All channels**		**96.48 %**

Analysis reveals that recent research predominantly uses various datasets. Classification tasks are typically performed in pairs using algorithmic solutions, mainly relying on machine learning and deep learning techniques. The classification performance ranges from 85 % to 97.21 %.

V. CONCLUSIONS

This study used deep learning techniques to detect sleep apnea, with the aim of establishing a robust decision support system for precise diagnosis. Using deep neural networks (DNNs), the research achieved a remarkable accuracy rate of 96.48 %, surpassing benchmarks from similar studies. This breakthrough significantly enhances early detection capabilities and has the potential to improve patient outcomes in sleep medicine. The application of deep learning enables accurate pattern recognition, promising advancements in real-time monitoring technologies. Ensuring model interpretability through explainable AI methods is crucial for building confidence among medical professionals. Ongoing clinical validation and longitudinal studies are underway to refine the model for practical implementation in medical settings.

In this study, we explored the application of deep learning techniques in the detection of sleep apnea. Our aim was to develop a robust decision support system to accurately identify sleep disorder cases. Leveraging DNNs, we focussed on effectively distinguishing between normal sleep patterns and sleep apnea instances, thus improving diagnostic accuracy. Through meticulous data collection and analysis, we demonstrated DNNs' potential in capturing complex sleep data patterns. Our results highlight the success of the DNN-based classifier, achieving an accuracy of 96.48 %, surpassing similar studies. This study contributes to sleep medicine by enhancing early diagnosis, which can significantly impact patient outcomes. The use of deep learning enables precise pattern recognition, promising advancements in real-time monitoring applications. Ensuring model interpretability through explainable AI methods is crucial for medical practitioners' trust. Clinical validation and longitudinal studies will further refine the model for real-world medical settings.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Y. Yeghiazarians *et al.*, "Obstructive sleep apnea and cardiovascular disease: A scientific statement from the American Heart Association", *Circulation*, vol. 144, no. 3, pp. e56–e67, 2021. DOI: 10.1161/CIR.0000000000000988.

[2] A. V. Benjafield *et al.*, "Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis", *The Lancet Respiratory Medicine*, vol. 7, no. 8, pp. 687–698, 2019. DOI: 10.1016/S2213-2600(19)30198-5.

[3] L. F. Drager, R. D. McEvoy, F. Barbe, G. Lorenzi-Filho, and S. Redline, "Sleep apnea and cardiovascular disease: Lessons from recent trials and need for team science", *Circulation*, vol. 136, no. 19, pp. 1840–1850, 2017. DOI: 10.1161/CIRCULATIONAHA.117.029400.

[4] J. S. Floras, "Sleep apnea and cardiovascular disease: An enigmatic risk factor", *Circulation Research*, vol. 122, no. 12, pp. 1741–1764, 2018. DOI: 10.1161/CIRCRESAHA.118.310783.

[5] D. Linz *et al.*, "Associations of obstructive sleep apnea with atrial fibrillation and continuous positive airway pressure treatment: A review", *JAMA Cardiology*, vol. 3, no. 6, pp. 532–540, 2018. DOI: 10.1001/jamacardio.2018.0095.

[6] S. P. Patil, I. A. Ayappa, S. M. Caples, R. J. Kimoff, S. R. Patel, and C. G. Harrod, "Treatment of adult obstructive sleep apnea with positive airway pressure: An American academy of sleep medicine systematic review, meta-analysis, and GRADE assessment", *Journal of Clinical Sleep Medicine*, vol. 15, no. 2, pp. 301–334, 2019. DOI: 10.5664/jcsm.7638.

[7] J. Faber, C. Faber, and A. P. Faber, "Obstructive sleep apnea in adults", *Dental Press Journal of Orthodontics*, vol. 24, no. 3, pp. 99–109, 2019. DOI: 10.1590/2177-6709.24.3.099-109.sar.

[8] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, "A review of obstructive sleep apnea detection approaches", *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 825–837, 2019. DOI: 10.1109/JBHI.2018.2823265.

[9] M. E. Sahin, "Image processing and machine learning-based bone fracture detection and classification using X-ray images", *International Journal of Imaging Systems and Technology*, vol. 33, no. 3, pp. 853–865, 2023. DOI: 10.1002/ima.22849.

[10] H. Ulutas, M. E. Sahin, and M. O. Karakus, "Application of a novel deep learning technique using CT images for COVID-19 diagnosis on embedded systems", *Alexandria Engineering Journal*, vol. 74, pp. 345–358, 2023. DOI: 10.1016/j.aej.2023.05.036.

[11] U. Erdenebayar, Y. J. Kim, J.-U. Park, E. Y. Joo, and K.-J. Lee, "Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram", *Computer Methods and Programs in Biomedicine*, vol. 180, art. 105001, 2019. DOI: 10.1016/j.cmpb.2019.105001.

[12] R. S. Arslan, H. Ulutaş, A. S. Köksal, M. Bakır, and B. Çiftçi, "Automated sleep scoring system using multi-channel data and machine learning", *Computers in Biology and Medicine*, vol. 146, art. 105653, 2022. DOI: 10.1016/j.compbiomed.2022.105653.

[13] A. R. Hassan and M. A. Haque, "An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting", *Neurocomputing*, vol. 235, pp. 122–130, 2017. DOI: 10.1016/j.neucom.2016.12.062.

[14] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017. DOI: 10.1109/TNSRE.2017.2721116.

[15] H. Hilmisson, N. Lange, and S. P. Duntley, "Sleep apnea detection: Accuracy of using automated ECG analysis compared to manually scored polysomnography (apnea hypopnea index)", *Sleep and*

Breathing, vol. 23, no. 1, pp. 125–133, 2019. DOI: 10.1007/s11325-018-1672-0.

[16] R. S. Arslan, H. Ulutas, A. S. Köksal, M. Bakir, and B. Çiftçi, “Sensitive deep learning application on sleep stage scoring by using all PSG data”, *Neural Computing and Applications*, vol. 35, no. 10, pp. 7495–7508, 2022. DOI: 10.1007/s00521-022-08037-z.

[17] C. Song, K. Liu, X. Zhang, L. Chen, and X. Xian, “An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals”, *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1532–1542, 2016. DOI: 10.1109/TBME.2015.2498199.

[18] A. Ramachandran and A. Karuppiah, “A survey on recent advances in machine learning based sleep apnea detection systems”, *Healthcare*, vol. 9, no. 7, p. 914, 2021. DOI: 10.3390/healthcare9070914.

[19] K. Li, W. Pan, Y. Li, Q. Jiang, and G. Liu, “A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal”, *Neurocomputing*, vol. 294, pp. 94–101, 2018. DOI: 10.1016/j.neucom.2018.03.011.

[20] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, “Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network”, *Information Fusion*, vol. 53, pp. 174–182, 2020. DOI: 10.1016/j.inffus.2019.06.024.

[21] T. Wang, C. Lu, G. Shen, and F. Hong, “Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network”, *PeerJ*, vol. 7, no. 9, p. e7731, 2019. DOI: 10.7717/peerj.7731.

[22] Q. Shen, H. Qin, K. Wei, and G. Liu, “Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal”, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021. DOI: 10.1109/TIM.2021.3062414.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[24] D. S. Morillo and N. Gross, “Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry”, *Medical & Biological Engineering & Computing*, vol. 51, no. 3, pp. 305–315, 2013. DOI: 10.1007/s11517-012-0995-4.

[25] A. H. Khandoker, J. Gubbi, and M. Palaniswami, “Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings”, *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 1057–1067, 2009. DOI: 10.1109/TITB.2009.2031639.

[26] W. Rawat and Z. Wang, “Deep convolutional neural networks for image classification: A comprehensive review”, *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017. DOI: 10.1162/neco_a_00990.

[27] S. S. Yadav and S. M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis”, *Journal of Big Data*, vol. 6, art. no. 113, 2019. DOI: 10.1186/s40537-019-0276-2.

[28] M. Olsen *et al.*, “A deep transfer learning approach for sleep stage classification and sleep apnea detection using wrist-worn consumer sleep technologies”, *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 8, pp. 2506–2517, 2024. DOI: 10.1109/TBME.2024.3378480.

[29] A. Zarei, H. Beheshti, and B. M. Asl, “Detection of sleep apnea using deep neural networks and single-lead ECG signals”, *Biomedical Signal Processing and Control*, vol. 71, part A, art. 103125, 2022. DOI: 10.1016/j.bspc.2021.103125.

[30] G. Surrel, A. Aminifar, F. Rincón, S. Murali, and D. Atienza, “Online Obstructive Sleep Apnea detection on medical wearable sensors”, *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 4, pp. 762–773, 2018. DOI: 10.1109/TBCAS.2018.2824659.

[31] A. Sheta *et al.*, “Diagnosis of obstructive sleep apnea from ECG signals using machine learning and deep learning classifiers”, *Applied Sciences*, vol. 11, no. 14, p. 6622, 2021. DOI: 10.3390/app11146622.

[32] Q. Yang, L. Zou, K. Wei, and G. Liu, “Obstructive sleep apnea detection from single-lead electrocardiogram signals using one-dimensional squeeze-and-excitation residual group network”, *Computers in Biology and Medicine*, vol. 140, art. 105124, 2022. DOI: 10.1016/j.compbiomed.2021.105124.

[33] Y. Wang, Z. Xiao, S. Fang, W. Li, J. Wang, and X. Zhao, “BI - Directional long short-term memory for automatic detection of sleep apnea events based on single channel EEG signal”, *Computers in Biology and Medicine*, vol. 142, art. 105211, 2022. DOI: 10.1016/j.compbiomed.2022.105211.

[34] S. Hu *et al.*, “Semi-supervised learning for low-cost personalized obstructive sleep apnea detection using unsupervised deep learning and single-lead electrocardiogram”, *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 11, pp. 5281–5292, 2023. DOI: 10.1109/JBHI.2023.3304299.