Detection of Sleep Apnea using 1D-CNN-BiLSTM: Fine-Tuning Convolutional Neural Network and Long Short-Term Memory Architecture for Improved Performance

Hritam Ghoshal, Anurag Dhar, Krishnadas Saha*

* Department of Computer Science and Engineering, Institute of Engineering and Management, Kolkata 700091, India

*krishnasas.saha@iem.edu.in

* Corresponding author

doi: https://doi.org/10.21467/proceedings.7.6.13

Abstract

Sleep Apnea is a commonly undiagnosed disease characterized by repeated interruptions in breathing during sleep, which leads to fragmented sleep and reduced blood oxygen levels. Handcrafted single- lead Electrocardiogram (ECG) signals are very complex in nature due to their large dimensions and not so robust nature. In this research, a 23 layered deep-learning based hybrid model was made by combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) with an attention mechanism module to process such complex ECG signals efficiently. The dataset was also preprocessed and segmented into time series data to improve the model training efficiency. Then, the Apnea-Hypopnea Index (AHI) was applied to identify normal and apnea subjects and categorize them into normal, mild-apnea, moderate-apnea and severe-apnea categories. These experiments were conducted on the publicly available Apnea-ECG dataset to assess the method's usefulness. Our proposed model succeeds in achieving a higher accuracy of 95.79% in comparison to other sleep apnea detection methods.

Keywords: Sleep Apnea, Deep Learning, CNN-LSTM

1 Introduction

Sleep apnea is a frequent but life-threatening sleep disorder that occurs in millions of individuals around the world. It is a common disorder where patients experience decreased airflow to the lungs for more than 10 seconds while sleeping [1]. The duration of sleep apnea can vary between 10s and 30s and can occur multiple times during a person's sleep [2]. Apnea causes either complete or partial obstruction of the airways [3][4]. There are three main types of sleep apnea: Obstructive Sleep Apnea (OSA), Central Sleep Apnea (CSA) and Complex Sleep Apnea. OSA, which is the most common among the three, is caused when the throat muscles get relaxed during sleep, which blocks the airways, for atleast 10 seconds of sleep [5][6]. If sleep apnea is left undiagnosed or untreated, it can cause severe health complications like cardiovascular diseases, hypertension, stroke, diabetes and even sudden death caused due to stroke [7].

Detection and diagnosis of sleep apnea usually involve overnight observation in a sleep laboratory with polysomnography (PSG). PSG is the gold standard for the diagnosis of sleep apnea, but it requires specialized equipment and personnel, thus making it costly, time-consuming and inaccessible for most patients [8]. Thus, increasing interest in creating automated, low-cost, and non-invasive techniques for the detection of sleep apnea. ECGs are very commonly used to detect sleep apnea as they are non-invasive in nature and contain a lot of information about the cardiovascular system. Heart Rate and Heart Rate Variability (HRV) are noted during apnea events. Various studies have indicated that ECG-based apnea detection techniques can provide high accuracy levels, which are as good as or even better than those from more invasive methods. The benefit of employing ECG signals is that they are readily recorded using handheld devices, facilitating continuous monitoring and real-time identification of apnea events. Thus, ECGs are a rich and accurate source of data for sleep apnea detection.

Deep learning methods, especially Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, have demonstrated great potential in the detection of sleep apnea from physiological signals such as electrocardiograms (ECGs) [9]. The CNN-LSTM model is used in this research because it combines the

Proceedings DOI: 10.21467/proceedings.7.6; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-989164-9-5

capabilities of both convolutional and recurrent networks, which not only makes the sleep apnea diagnosis more accurate, but also more reliable [10][11]. The CNN layers help to detect the patterns of the raw ECG signals (such as QRS complexes and P-waves) and the LSTM layers help in detecting the temporal patterns between consecutive heartbeats.

This hybrid model works extremely well for processing time-series data (here, ECG signals) as it is simultaneously able to learn temporal and spatial features. It also helps the model learn from future and past time steps, thus improving the apnea detection performance even further. This model is capable of learning sophisticated patterns and temporal relationships in the data, thus being particularly apt for time-series analysis. The CNN layers are used to extract spatial features from the raw single-lead ECG signals, whereas the LSTM layers are used to learn the temporal relationships required for detecting apnea events.

1.1 Convolutional Neural Network Architecture (CNN)

In the realm of deep learning, Convolutional Neural Network (frequently abbreviated as CNN or ConvNet) represents a category of Artificial Neural Network (ANN) which are particularly adept at processing visual or time-series data [12]. CNN consists of convolutional, pooling, flatten and fully-connected layers which work together to extract spatial and temporal features. Convolutional layers (which are the core layers of CNN) use learnable filters (known as Kernels) which identify features like edges, textures, shapes and time-series. These filters are spread across the input space, which makes the convolutional layers extremely efficient.

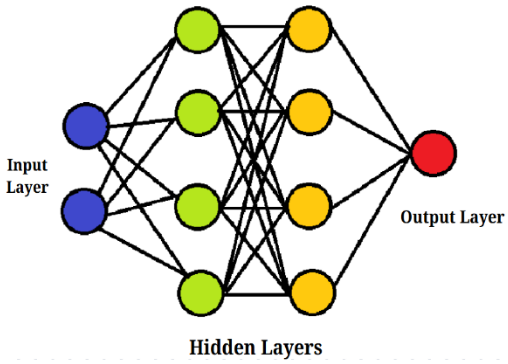

Fig 1. Visualization of Convolutional Neural Network architecture

Activation functions like ReLU (Rectifier Linear Unit) introduce non-linearity to facilitate the model training process and pooling layers such as max pooling or average pooling help to down-sample feature maps which decreases computational complexity while preserving critical information. Deeper CNN architectures stack several convolutional and pooling layers on top of one another to facilitate detection of complex patterns at multiple abstraction levels. The fully-connected layers which are present at the output end combine the extracted features to successfully give accurate prediction results. To minimize overfitting and stabilize the training procedure, methods like dropout and batch normalization are also used.

1.2 Long Short-Term Memory (LSTM)

A Long Short-Term Memory (LSTM) network is a particular type of recurrent neural network (RNN) that is specially designed for sequential data processing and learning of long-range dependencies. Unlike normal RNNs, which are plagued by exploding and vanishing gradients, LSTMs use specialized memory cells and gating mechanisms to control the flow of information. Each LSTM unit contains three gates, the forget gate, the input gate and the output gate. The forget gate helps to determine what to forget from the memory. Whereas the input gate helps to determine what new information to remember and finally, the output gate helps to determine the final output produced by the memory cell. These gates utilize sigmoid activation in order to have selective information storage and preservation of long-term dependencies. LSTMs also have a cell state which acts as a memory buffer retaining important information for many time steps. Such a specific design enables LSTMs to learn temporal patterns in time-series data, making LSTMs highly appropriate for such problems like speech recognition, natural language processing, and anomaly detection. LSTMs can even be configured in deep architectures for performance and feature extraction improvement. Moreover, they can also be combined with convolutional layers (CNN-LSTMs) for problems which involve learning of spatial and temporal features, i.e., video analysis and diagnosis of sleep apnea. Improvements like BiLSTMs (Bidirectional LSTMs) have even superior

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

performance by considering past and future context. Even though they are powerful, LSTMs consume much computational power and are generally outperformed by Transformer-based models in high-scale problems. Nevertheless, LSTMs remain a handy tool in sequential modeling due to their efficiency in preserving long-term dependencies.

The novelty of our methodology is expressed in points below:

- Combines Convolutional Neural Networks (CNN) for spatial feature extraction and Bidirectional LSTMs for capturing temporal dependencies in ECG signals
- Integrates an Attention layer to highlight the most relevant time-steps, improving detection of apnea patterns.
- Stacks multiple convolutional and recurrent layers, enabling the model to learn both local and global features effectively.
- Processes raw, segmented ECG signals without manual feature engineering, making the approach robust and scalable.

The paper is structured with a concise flow. The first section has the 'Introduction', the second section contains 'Dataset Preprocessing', third section has the 'Methodology' of our novel model architecture, fourth section contains the 'Experimental Result' section, fifth section has 'Discussions' and the sixth section has the 'Conclusion'.

2 Dataset Preprocessing

The Apnea-ECG dataset [13] is a well-acknowledged physiological dataset that is used to identify sleep apnea. The dataset contains 70 long-night ECG and respiration signal recordings, consistently divided into a training set containing 35 recordings and a validation set with the same number of recordings. Recordings a01 to a20, b01 to b05, and c01 to c10 are the recordings included in the training set, while those labeled x01 to x35 are included in the validation set. Each of the recordings demonstrates variability in time, ranging approximately from 7 to 10 hours, and contains many of the physiological signals important for identifying apnea. Each record has a digitized ECG signal that is continuous and apnea annotations and QRS annotations that are automatically generated. The apnea annotations were obtained by human annotators from the respiration and related signals that were recorded simultaneously, whereas the QRS annotations, which represent all the beats as normal, were automatically generated and could be incorrect.

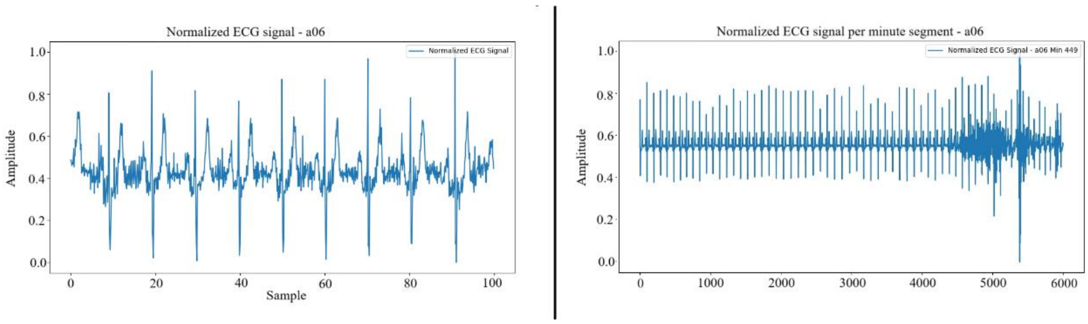

Fig. 2. Normalization of signal file a06 representing a patient. Left Graph shows apnea peak detection for 7-8 hours of patient a06. Right graph shows apnea peak detection per minute of patient a06

Aside from the ECG signals and the annotations, eight particular recordings—designated as a01 through a04, b01, and c01 through c03—hold other respiration-related signals. These are Resp C, which is chest respiratory effort as detected by inductance plethysmography, and Resp A, which is abdominal respiratory effort as detected by the same technique. Another is Resp N, which is oronasal airflow detected by nasal thermistors, and SpO2, which is blood oxygen saturation.

This dataset provides several files with different functions to accomplish the tasks of storing and processing data. The rnn.dat stores the ECG signal records in 16-bit, Least Significant Byte First (LSBF) for each pair and is sampled at 100 samples per second. A nominal value of 200 A/D per millivolt is given to represent amplitude values of the signal. To facilitate proper data interpretation, a .hea file is provided for each recording: a text header describing the format and organization of signal files to make them compatible with the analysis tools. In case of the training set's recordings, there is a .apn file that represents the minute-by-minute binary annotation of apnea presence or absence. Furthermore, there is a .qrs file that contains QRS annotations created by the algorithm with machine generation, that should be validated by a person. Besides, there are another eight recordings, including respiratory signals, with an additional rnnr.dat file containing all four respiration-related signals into one file. The presence of .rnnr.hea and .rnner.hea header files in the dataset facilitates joint evaluation of ECG and respiration signals for simultaneous visualization and interpretation of both physiological signals.

The preprocessing was done in this manner. First, a Chebyshev bandpass filter with cutoff frequencies at 0.5 and 48 Hz was designed to suppress very low frequency baseline wander and power line interference noise. The ECG signals were first segmented into 1-minute-long segments. After that, the noisy segments were to be automatically discarded via an auto-matic weight calculation algorithm. This algorithm measures the similarity of the autocorrelation functions (ACF) of the segments with the time lags of 60 s. Finally, a symmetric parameter matrix mapping similarity was produced. Then, the vector for weights formed in the next step, whereby each of the columns in the AM has been summed. Noise segments identified with a threshold value of 0.8 were removed. This value was chosen via try and error after running different experiments. At this point, approximately 5% of the segments (1717 segments of the Apnea-ECG dataset) were removed from the analysis.

3 Methodology

The two models proposed in this study incorporate hybrid deep learning architectures through several modalities including Convolutional Neural Networks (CNNs), bidirectional long short-term memory (BiLSTM) networks, and an attention mechanism for the detection of sleep apnea. The design of the model incorporates an input of one-dimensional electrocardiogram signals represented as sequences of 6000 samples per input window. An input layer for this model takes raw ECG signals fed in a one-dimensional format before the feature extraction performed by three different 1-dimensional Conv (Conv1D) convolution blocks. In this way the convolutional layers will be set with ReLU activation of the rectified linear unit and use batch normalization to stabilize training. Max pooling layers followed by different pool sizes (4, 4, and 3) have been put into place to reduce dimensionality while conserving the essential features. Increasing dropout layers will be {0.2, 0.3, 0.4}-overfitting as a result of the dropout. The dependence of sequential ECG signals is captured using two bidirectional LSTM layers after feature extraction. The first BiLSTM layer comprises 128 units and functions by returning sequences, while the second BiLSTM layer has 64 units. Together, the two cover the entire range of short- and long-term dependencies while learning the ECG data. Between every BiLSTM layer, a dropout regularization value (0.3) is provided to reduce overfitting.

An attention mechanism is employed to enhance the feature representation on the BiLSTM outputs. Different weights are assigned to each of the time steps by the attention layer so that the model learns to focus on the most relevant portions of the signal concerning apnea detection. The attention-weighted feature representations are flattened before being passed through fully connected layers. A classification is made through these dense layers having 128 and 64 neurons in their respective layers and using the ReLU activation as dropout with rates of 0.4 and 0.3, respectively. An output layer with sigmoid activation shall classify whether apnea occurs or not. The compiled model uses an Adam optimizer with a learning rate of 0.001 and also gradient clipping (clipnorm=1.0) for improving stability. The loss function followed is binary cross-entropy and the evaluation metric is accuracy. This methodology ensures that feature extraction, the sequential learning process, and attention mechanisms are all incorporated effectively into the model-built strength automation and robustness for sleep apnea detection through ECG signals. Figure shows the model architecture.

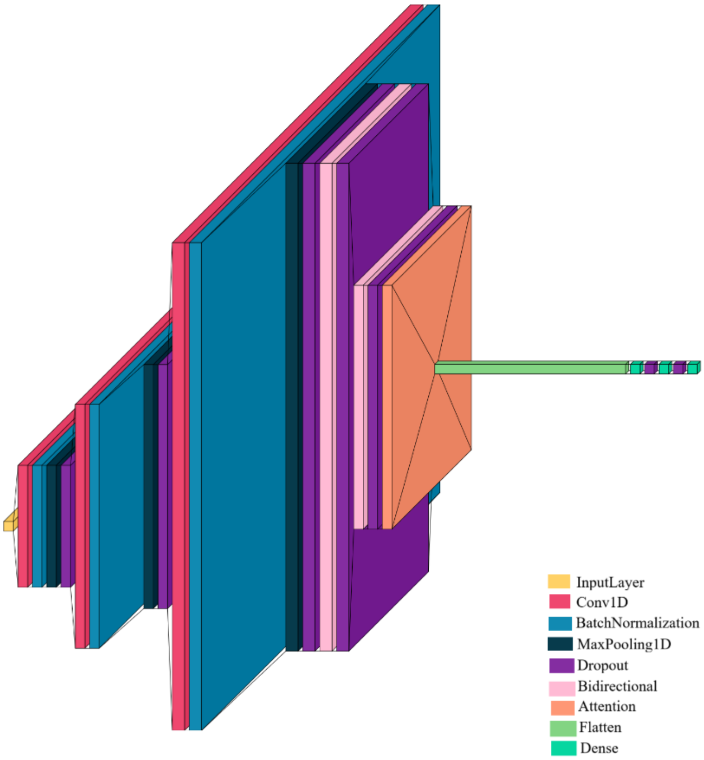

Fig. 3. Proposed CNN-LSTM Model Architecture in layers

4 Experimental Results

The test of the current CNN-LSTM model was done on the Apnea-ECG dataset in order to assess the capacity of the designed model at classifying sleep apnea series based on the ECG recordings. Its successive improvement of accuracy, over 150 training epochs, explicitly proved that the convolutional feature extraction combined with sequential learning through LSTMs was effective. This suggests that this model learned the discriminative patterns in ECG signals for the accurate detection of apnea events.

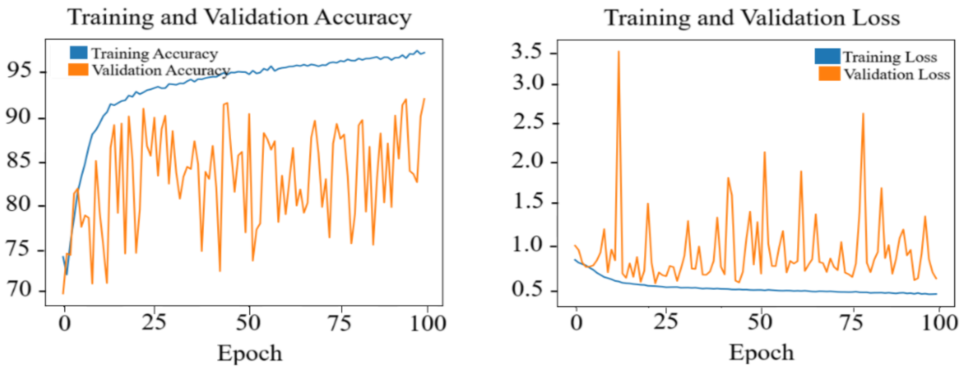

Fig. 4. The left graph shows the training and validation accuracy of proposed model. Right graph shows the training and validation loss of the proposed model

So, what was observed in the model as things went on was a gradual increase in accuracy. The initial training accuracy was 62.19% around epoch 10, then increased to 68.75% at epoch 25. As the model improves the features learned, it became 74.63% at epoch 50 and increased to 89.17% at epoch 100. The highest training accuracy was achieved at epoch 150, hence, reaching a peak of 95.79%. A similar corresponding trend could be followed by the validation accuracy; it achieved a final value of 93.62%, thus proving the generalizability and ro-

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

bustness of the model. Most significantly, the coupling of CNN layers and BiLSTM layers becomes an essential part in detecting apnea events exactly in the ECG signals.

Table 1. Accuracy improvement over 150 epochs

Epoch	Accuracy (%)
10	62.19
25	68.75
50	74.63
100	89.17
150	95.79

The local patterns in the ECG waveform were captured by the convolution layers, emphasizing some critical features such as heart rate or rhythm changes correlated with apnea episodes. These spatial features are then passed to BiLSTM layers modeling temporal dependencies and sequential variations in sleep segments. The bidirectional nature of the LSTMs made sure that the network would provide input to each segment from both past and future information, improving the classification accuracy. The attention mechanism proved to be one of the most significant parts of that architecture, as it allowed the model to attend to the time steps that are most relevant within the sequence. Since an apnea episode manifests as changes occurring over some stretching periods of time, the attention mechanisms helped enhance critical parts marking heart rate and respiratory effort changes, making the network more interpretable.

We use these procedures to assess our model's performance against established state-of-the-art methods for sleep apnea detection based on deep learning models. Our 1D-CNN-BiLSTM-based model recorded the highest final training accuracy of 95.79% and validation accuracy of 93.62%, although still outperforming some of the published works-an LSTM-based model [14] achieved this accuracy of 82.1% and such performance is drastically lower than our method's. This performance indicates that the sole usage of LSTMs is not enough for efficient extraction of spatial ECG features. Our model extends the information to LSTM layers with superior classification accuracy, which is ensured by CNN layers before this transfer.

Table 2. Our model compared to SOTA architectures

Model	Accuracy (%)
LSTM	93.62
CNN	95.57
1D CNN + BiLSTM (ours)	**95.79**

Table 3. Performance of our proposed model under different metrics

Performance Metrics	Proposed Model
Accuracy	95.79
Specificity	94.87
Sensitivity	89.32
ROC	98.00
F1-Score	92.00

5 Discussion

Adding multiple convolutional layers enabled the gradual extraction of more complicated features from ECG signals. With three different convolutional blocks, a hierarchy was formed by the model to represent apnea-specific patterns and capture low-level features together with high-level features relevant to classification. The bidirectional LSTM layers further enhanced the capabilities of the model by allowing it to capture information about both past and future dependencies from the time-series data. Particularly, this was useful for the detection of apneas, which tend to have very subtle and gradual onset characteristics that may escape immediate observability. Thus, the LSTM proved to be very useful in detecting such instances, thereby making the model smart in identifying considered apneas.

To generalize the training more on out-of-distribution subjects, dropout regularization has been applied at several points through the network. With dropout rates of between 20 and 40%, the net was encouraged to work more generally by not depending too much on certain features, thereby allowing it to learn more general patterns from the data. An attention mechanism was also included to further refine the model's attention to the most significant time intervals within each ECG segment. Differentiating weights over time steps not only aid in increasing model interpretability but also help in increasing the model's acuity when it comes to detecting apnea episodes. The training would be further improved through the usage of Adam as an optimizer. Compared to constant learning rates, dynamically adapting the learning rate onto convergence fairly stable to ensure that improvement is realized throughout the course of training.

Analysis of the model show that improvements were observed in accuracy at around epoch 10 where its accuracy improved to about 62.19%, with steady increase to at least 68.75% by epoch 25. By epoch 50, it climbed to 74.63. The most improvement occurred from epochs 50 to 100, with accuracy jumping from 74.63% to 89.63%. Training accuracy peaked at an epoch of 150 at 95.79%, validation accuracy standing at 93.62%. That swift increase suggested that during that epoch the model refinement was important and it even learned quite specific patterns related to apnea, considering a high F1-score of 92%.

6 Conclusion

On the improved path of CNN-LSTM for sleep apnea detection, the model showed very high effectiveness in that it extracted both spatial and temporal features from ECG signals. Apnea-related intricate characteristics with high accuracy were captured due to a combination of convolutional layers for hierarchical feature learning and bidirectional LSTM layers for sequential pattern recognition. This model is supported by dropout regularization and an attention mechanism to increase generalization and interpretability. Training for 150 epochs, the model achieved the best training accuracy at 95.79% and validation accuracy at 93.62%. Improvements in performance were huge between epoch 50 and 100 when accuracy shot up from 74.63% to 89.63%. This would serve to further confirm that extended training is paramount when it comes to refining learned representations. The final accuracy thus affirms the model's competitiveness over existing deep learning methods, thereby providing a robust solution in apnea detection. Our hybrid CNN-LSTM architecture considered spatial and temporal dependencies and outperformed previous CNN and LSTM models in apnea classification. The ability to simultaneously look at local and long-range feature processing within the ECG signal presented an added advantage, leading to more accurate and reliable detection. Way forward, we may wish to enhance and build upon this work as future works. These enhancements may include other physiological signals and the feature-extraction-based transformer model. The integration of the model with real-time clinical deployment would facilitate early diagnosis and monitoring for sleep apnea, resulting in better patient outcomes.

7 Declaration

The authors declare that there is no conflict of interest regarding the research and publication of this paper. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. 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, p. 103125, Jan. 2022.
2. M. Bahrami and M. Forouzanfar, “Sleep apnea detection from single-lead ECG: A comprehensive analysis of machine learning and deep learning algorithms,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.
3. F. Mendonca, S. S. Mostafa, A. G. Ravelo-Garcia, 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, Mar. 2019.
4. A. Zarei and B. M. Asl, “Automatic classification of apnea and normal subjects using new features extracted from HRV and ECG-derived respiration signals,” *Biomedical Signal Processing and Control*, vol. 59, p. 101927, May 2020.

5. H. D. Nguyen, B. A. Wilkins, Q. Cheng, and B. A. Benjamin, “An online sleep apnea detection method based on recurrence quantification analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1285–1293, Jul. 2014.
6. A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, “Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 37–48, Jan. 2009.
7. M. O. Mendez, A. M. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, “Sleep apnea screening by autoregressive models from a single ECG lead,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 12, pp. 2838–2850, Dec. 2009.
8. 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, p. 105001, Oct. 2019.
9. D. Dey, S. Chaudhuri, and S. Munshi, “Obstructive sleep apnoea detection using convolutional neural network based deep learning framework,” *Biomedical Engineering Letters*, vol. 8, no. 1, pp. 95–100, Feb. 2018.
10. D. Novak, K. Mucha, and T. Al-Ani, “Long short-term memory for apnea detection based on heart rate variability,” in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, BC, Canada, Aug. 2008, pp. 5234–5237.
11. R. K. Pathinarupothi, R. Vinaykumar, E. Rangan, E. Gopalakrishnan, and K. P. Soman, “Instantaneous heart rate as a robust feature for sleep apnea severity detection using deep learning,” in *Proceedings of the IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Orlando, FL, USA, Feb. 2017, pp. 293–296.
12. S. Goswami, U. Srivastava, P. M. Chinni, and S. Dhanalakshmi, “Enhancing brain MRI tumor detection: Exploring vision transformers and fine-tuned convolutional neural network architecture for improved performance,” in *Proceedings of the International Conference on Advanced Computing and Intelligent Technologies*, Singapore, Dec. 2023, pp. 303–316.
13. T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, “The apnea-ECG database,” in *Computers in Cardiology 2000*, vol. 27, Cambridge, MA, USA, Sep. 2000, pp. 255–258.

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922