INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 6, June 2026 | ISSN: 2456-4184 | IJNRD.ORG

Hybrid CNN–Transformer-Based Sleep Apnea Detection Using Physiological Data

Maddula Meghala(M.Tech),
Student of Computer Science and Engineering,
MGIT Autonomous, TG
Hyderabad, India
mmeghala_pg24csea907@mgit.ac.in

Dr.K.Satish kumar,
Associate professor of Computer science and Engineering,
MGIT Autonomous, TG
Hyderabad, India
ksatishkumar_cse@mgit.ac.in

Dr.C.R.K.Reddy,
Associate professor of Computer science and Engineering,
MGIT Autonomous, TG
Hyderabad, India
crkreddy_cse@mgit.ac.in

***Abstract*— Sleep apnoea is a frequent sleep-related breathing disorder that, if left untreated, can result in cardiovascular complications, sleep disturbances and chronic health risks. Manual diagnosis and clinical observation are still used in traditional diagnosis while modern automatic diagnosis techniques have some deficient points in the diagnosis of signal characteristics within a local area and the larger-scale time correlation in physiological data. In order to overcome the constraints mentioned above, we created a hybrid DL framework that combines CNN and Transformer-based attention mechanism for automated sleep apnoea identification utilising ECG inputs. That frame is trained using an ECG database of 2660 samples, where each sample has 2500 ECG points per sample. There are data preparation steps such as data verification, normalisation of features, division of the dataset and the construction of the sequential input. This design incorporated feature extraction using convolution and contextual modelling using transformer for binary classification. Experimental results showed the accuracy of 98.12%, demonstrating the effectiveness of using local representation learning and attention guided sequence modelling for reliable sleep apnoea detection from ECG.**

"Keywords— Sleep Apnea Detection, Electrocardiogram (ECG), Hybrid Deep Learning, Convolutional Neural Network (CNN), Transformer Encoder, Multi-Head Attention, Physiological Signal Analysis, Explainable Artificial Intelligence (XAI)."

I. INTRODUCTION

Sleep apnoea is a widespread sleep-related breathing disorder where air stops flowing or becomes blocked from the nose or mouth during sleep, causing disturbances in sleep and inadequate levels of oxygen saturation. The disease is now a serious public health concern due to its association with cardiovascular disease, metabolic dysfunction, reduced quality of life, and health risks in the long-term if undiagnosed [1, 3]. Epidemiological studies indicate that sleep-disordered breathing is a significant health problem in the adult population around the world and highlights the growing need for accurate and scalable diagnostic tools [1, 2].

PSG is the clinical gold standard for the diagnosis of sleep apnoea, and entails recording various physiological data across one or more nocturnal monitoring sessions [4]. PSG provides complete evaluation possibilities but specialized equipment, controlled laboratory conditions, specialized interpretation and substantial time and operational expenses are required [4] [5]. These constraints create accessibility issues and drive alternative screening and detection technologies that can offer more effective, earlier identification of sleep-related breathing disorders [6].

It has been proven in recent years that ECG signals contain useful information that is relevant to the autonomic and cardiovascular responses that are triggered by apnoea events. The development for advances in this area has been made easier by the availability of physiological repositories and benchmark datasets [7, 8] that have allowed for the development and assessment of automated systems for sleep apnoea analysis. With the advent of DL, representative patterns of raw data can be extracted automatically without handmade feature engineering [9], and the amount of physiological data has been growing.

However, existing AID techniques are unable to reliably represent the localised properties of the signal and relations/dependencies in physiological sequences. Methods primarily based on conventional feature extraction or isolated learning paradigms may lack flexibility in dealing with complicated temporal correlations contained in biological information. Attention-based learning has shown great potential in learning long-range dependencies, and selectively focusing on the relevant portion of the sequential data, which has enabled new understanding of physiological signals [10].

Motivated by these trends, this study intends to explore an intelligent framework for sleep apnoea diagnosis based on physiological signal analysis with the emphasis on the improvement of automatic classification between normal and apnoea situations. This is to investigate an integrated learning method to obtain effective feature representation and apply to the ECG-based diagnostic scenarios. It focuses on the binary classification of sleep apnoea from physiological signals and is targeted towards automatic analysis, better interpretation assistance and reduction of manual interpretation.

The main contributions are as follows: (i) creation of a DL-based automatic sleep apnoea recognition framework, (ii) interconnection of complementary representation learning concepts for understanding physiological signals and (iii) quantitative evaluation of the effectiveness of the proposed framework. From this point of view, the framework provided could help facilitate the advancement of the accessible and data-driven solutions for screening and computer-aided clinical analysis of sleep apnoea.

II. RELATED WORK

Physiological signals are commonly used for automated sleep apnoea identification as standard nightly monitoring methods have their limitations and advanced signal processing and DL methods are increasingly adopted. Physiological modalities such as single-lead ECG signals are available and, since the acquisition techniques are relatively easy, have emerged as a useful alternative since they record cardiac and autonomic responses related to the occurrence of apnoea, while the signals are recorded simultaneously.

The initial studies on ECG-based methods mainly focused on hand-designed features extraction and conventional classification techniques. Zarei and Mohammadzadeh Asl [11] introduced a wavelet transform method and entropy-based features extracted from ECG signals for sleep apnoea detection and demonstrated the effectiveness of extracting frequency and nonlinear features for classification. These techniques were, however, quite complex to preprocess and were based on hand-crafted representations causing them to be limited in their use in different physiological settings.

The automatic representation learning is achieved using DL techniques to overcome these limitations in the field of ECG analysis. To gain the temporal information and boost the detection performance, Li et al. [12] combined deep neural network with hidden Markov model. Unsupervised feature learning was investigated by Feng et al. [13] to automatically obtain the latent ECG representations and reduce the reliance on manually designed processing. These approaches indicated improved learning ability, but were not without their challenges when it comes to modelling more complex sequences.

CNNs have been widely applied to the analysis of physiological signals, as a feature of the CNN is that it can automatically extract hierarchical features. For sleep apnoea diagnosis, Wang et al. [14] experimented with a modified LeNet-5 architecture and demonstrated its viability for learning with ECG data to detect sleep apnoea. Bahrami and Forouzanfar [15] investigated several DL algorithms, emphasising the effect of model design on the prediction performance. The CNN based frameworks enhanced the classification results but the localised receptive fields restricted the ability to catch long range temporal correlations found in ECG signals.

Issues of sequence modeling, attention techniques, and Transformer architectures were all found to be effective solutions. In [16] Hu et al. demonstrated that self-attention can be used to improve adaptive representation learning by attending to informative components of the signals. Based on that, Hu et al. [17] proposed a Hybrid Transformer-based approach for Obstructive Sleep Apnoea identification, and demonstrated improved contextual dependency learning across the ECG sequences.

Lately, a combination of convolution and Transformer techniques is used to utilise complimentary learning capabilities. Liu et al. [18] proposed a CNN–Transformer framework for sleep apnoea identification utilising ECG signals, showing that the combination of local and global features might boost the representation power. Apart from apnoea detection, Natarajan [19] showed the potential of Transformer networks for classification of ECGs, while Dosovitskiy et al. [20] showed the general performance of Transformer topologies for representation learning with the help of attention-based modelling.

Although much work has been done, there are still some problems with the existing techniques. The conventional approaches rely on features handcrafted by a human expert and complex pre-processing, single CNN network might not be able to extract long-range dependencies, and Transformer-based approaches might miss local signal patterns. Furthermore, there is a major challenge in balancing good predictive performance with interpretability. Therefore, the reliable ECG-based sleep apnoea detection is based on an integrated framework that includes local feature extraction, contextual sequence interpretation and interpretable decision support.

III. Methodology

A. Overall Framework

A sequential DL process was used to develop the methodological approach for automated sleep apnoea detection from physiological ECG signals. The ECG dataset was acquired and data preparation steps were implemented, such as data inspection, preprocessing and normalisation, and the input samples were structured for sequence-based learning. The processed data was then input to a hybrid CNN-Transformer network, where the convolutional network was employed for capturing the local signal representations, while the Transformer encoder captured wider temporal correlations in the ECG sequence. This was then trained with supervised learning, and optimized using the training control mechanism based on the validation. The performance was assessed with classification and discriminating metrics and an explainability analysis was carried out to comprehend the decision behaviour of the model and find the signal regions with influence. The whole process of the proposed system is shown in Fig. 1.

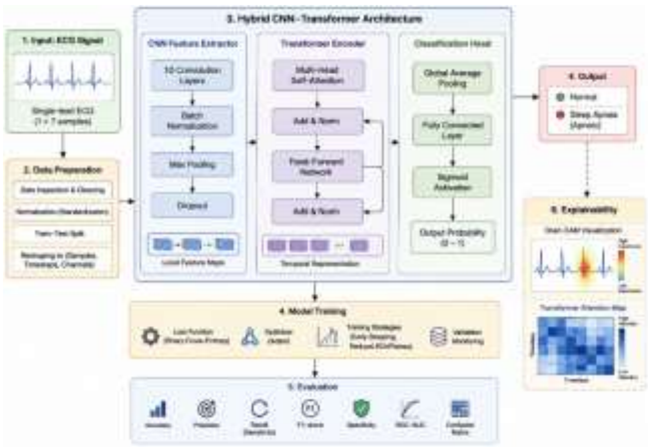

Fig.1 System Architecture

The overall system design for the proposed sleep apnoea detection system is shown in Figure 1. The workflow starts with collection of ECG physiological signal then data preparation and transformation to a suitable sequential representation. The extracted features of the local signals are then passed to a hybrid CNN-Transformer network, where the CNN component is responsible for extracting the local signal characteristics, and the Transformer component extracts the temporal correlation characteristics of the contextual signals. A binary classification of normal and sleep apnoea is performed using the learnt representation. Lastly, we add evaluation and explainability phases to assess the predicted performance and model decision behavior.

B. Dataset Description

The overall system design for the proposed sleep apnoea detection system is shown in Figure 1. The workflow starts with collection of ECG physiological signal then data preparation and transformation to a suitable sequential representation. The extracted features of the local signals are then passed to a hybrid CNN-Transformer network, where the CNN component is responsible for extracting the local signal characteristics, and the Transformer component extracts the temporal correlation characteristics of the contextual signals. A binary classification of normal and sleep apnoea is performed using the learnt representation. Lastly, we add evaluation and explainability phases to assess the predicted performance and model decision behavior.

Table 1. Dataset Characteristics

Parameter	Value
Signal Type	ECG
Samples	2660
Features	2500
Classes	2
Task	Binary Classification

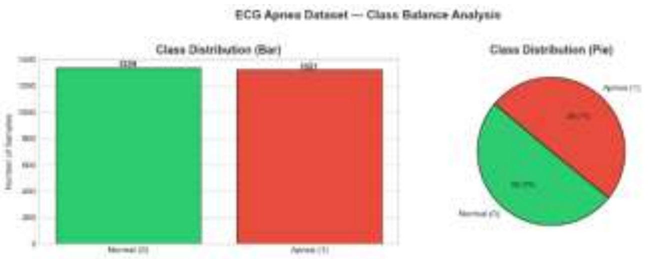

Fig.2 ECG Apnea Dataset — Class Balance Analysis

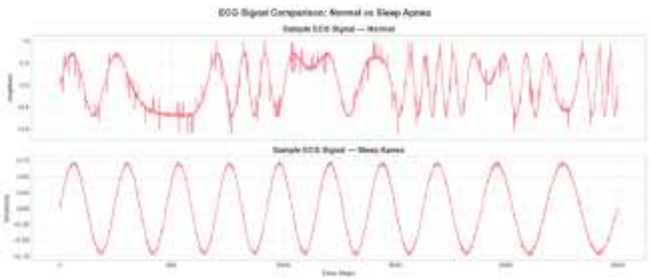

Fig.3 ECG Signal Comparison: Normal vs Sleep Apnea

C. Data Preprocessing

Pre-processing of ECG signals was performed to facilitate the effective analysis of the ECG via DL. Data quality checks, missing value treatment, feature engineering, normalisation, dataset splitting and reshaping operations were employed in the preprocessing procedure to prepare the data as structured inputs suitable for the sequence-learning architecture proposed.

Missing Value Analysis: The data set used was checked for the completeness and consistency of the data set before the construction of the model. All the ECG signal properties and target labels were subjected to a missing value assessment done column wise to identify missing or null observations. A median based imputation technique was selected as a powerful pre-processing technique, and during the inspection no missing values were detected. Median substitution was used in this study to reduce the effect of outliers and to maintain the stable physiological signal value statistics.

Feature–Label Separation: Then the data set was split into features to be fed into the model and target labels. The data put into matrix was the values of ECG signal at each observation. There were 2500 sequential signal points measured at each sample. The binary class labels (Normal and Sleep Apnoea) were designated as the target variable in the last column of the dataset. The separation allows for the possibility of processing the predictor variables and output labels separately during training of the model.

Feature Normalization: To ensure consistent convergence during model optimisation and to remove the difference in magnitude across ECG readings, the Normalisation technique called StandardScaler was used. Each feature is scaled to have a mean of 0 and a variance of 1 in the standardisation process. This method is helpful for improving numerical stability and effective learning behaviour of deep neural networks. In mathematical terms the normalisation is:

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

where x is the original value, μ is the mean of the feature, and σ is the standard deviation.

Dataset Splitting and Reshaping: The dataset was split into two sets: training set and testing set, to verify the generalisation ability of the normalised dataset. An 80:20 train-test ratio was used. In addition, a validation subset was also created from the training data set to evaluate the model performance during the model learning process. Next, the input data was re-shaped into the 3D tensor form that is well suited for convolutional processing of sequences. The final representation was organized as:

$$samples, timesteps, channels \tag{2}$$

where timesteps correspond to ECG sequence length and channels represent the univariate ECG input dimension.

D. Proposed Hybrid CNN–Transformer Architecture

The proposed architecture integrates the convolutional-based representation learning and Transformer-based sequence modelling to enable automatic sleep apnoea detection using the ECG data. This approach combines local feature extraction and global contextual comprehension into one DL process, enabling the model to learn discriminative physiological patterns and temporal correlations from sequential ECG data. The overall design of the proposed framework is shown in Fig. 1.

CNN-Based Local Feature Extraction: Local temporal features of the ECG sequences are extracted using one-dimensional convolutional neural networks (Conv1D) in the initial step. Convolution techniques give feature maps which represent the morphology of the signals and temporal variations of the neighbouring signals. When training, batch normalisation helps stabilise training and helps the network converge better, and max pooling helps reduce the dimensionality of the features and helps retain the predominant features. To overcome the overfitting and to improve the generalisation power in learning, drop out regularisation is followed.

The convolution operation can be expressed as:

$$y(t) = \sum_{i=0}^{k-1} x(t+i) \cdot w(i) + b \tag{3}$$

where (x) denotes the input signal, (w) represents convolution kernel weights, (k) is kernel size, and (b) indicates the bias term.

Transformer Encoder Module: After convolutional features extraction, the resultant feature representations are then fed to the Transformer encoder to capture longer temporal dependencies over ECG sequences. The Transformer module has multi-head self-attention to capture informative regions and simulate interactions between different parts of a sequence. We also add residual connections to maintain the learnt information and make the gradient flow easier. Useful tricks such as feed-forward layers and layer normalisation to tweak features and to regularise training. This technique enables contextual learning outside of localised receptive areas.

The self-attention operation is defined as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (4)$$

where (Q), (K), and (V) denote query, key, and value matrices respectively, and (d_k) represents the attention scaling factor.

Classification Layer: The final layer of the architecture makes the prediction on top of the compact learned representations from the transformer module. The temporal features are merged into a lower dimensional representation using global average pooling, without losing the general information in the signal. The outputs of the two are then fed into a fully connected dense layer to perform non-linear feature refining. Finally, the probability score for binary classification of Normal and Sleep Apnoea cases is obtained from the sigmoid activation function which will aid in automated end-to-end decision making.

E. Model Training Strategy

The training technique is designed to optimize the learning of the model, to enhance the stability of convergence and to enhance the generalisation ability for sleep apnoea classification from ECG signal sequence.

Optimization Setup: Adam optimiser was used for the optimisation of the model because to its adaptive learning ability and efficient gradient based parameter updates. The learning rate was fixed to 0.001 for uniform convergence during training. The loss function used was BCE because there are only two classes of outcomes related to a prediction task: Normal and Sleep Apnoea. The BCE measures the difference between the predicted probability and the true class labels, and is used to adjust the weights during optimisation.

The binary cross-entropy loss is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \qquad (5)$$

where (y_i) represents the true label and (\hat{y}_i) denotes the predicted probability.

Training Control Mechanisms: To optimise the training efficiency and avoid overfitting, various training control methods were used in training the model. We applied early stopping based on validation loss, and restarted training when no improvement was observed in the validation loss, using the best parameters from the previous training. In addition, the ReduceLROnPlateau method automatically adjusted the learning rate when the validation performance was flat, allowing more accurate optimisation updates. The maximum epochs set to 50 and batch size set to 32. A validation set was also produced when training, which is continually used to test learning behaviour and to aid in model selection.

F. Performance Evaluation Metrics

It was comprehensively tested in terms of prediction performance of the proposed framework through several classification and discrimination criteria, class-level behaviour and reliability of decisions.

Accuracy: It is the percentage of samples correctly predicted from all the samples.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \qquad (6)$$

Precision: Precision is the percentage of positive samples correctly classified by comparing the predictions with the actual values, and it's the precision of the predictions.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \qquad (7)$$

Recall: Recalls the ability of the model accurately to classify true positive cases of sleep apnoea.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad (8)$$

F1-score: The F1-score is a balance of the precision and recall.

$$F1\ Score = 2 * \frac{Precision \ \mathrm{X} \ Recall}{Precision + Recall} \qquad (9)$$

Sensitivity: Sensitivity is true positive rate and indicates the efficiency of apnoea detection.

$$Sensitivity = \frac{TP}{(TP + FN)} \qquad (10)$$

Specificity: The ability to properly identify normal ECG samples.

$$Specificity = \frac{TN}{(TN + FP)} \qquad (11)$$

ROC–AUC: ROC–AUC is a measure of the ability to discriminate between the different classification thresholds.

$$AUC = \int TPR(FPR)d(FPR) \qquad (12)$$

G. Explainability Analysis

The addition of explainability techniques to explore key ECG areas and to train attention patterns further increases the interpretability of the model and allows to understand the model's decision-making behavior.

Grad-CAM for ECG Interpretation: The contribution of the ECG segments to the model predictions was visualised using Grad-CAM for ECG interpretation. This approach is based on using intermediate convolutional feature maps to compute saliency representations, which are then projected into the input signal space. The produced visualisation points out physiologically informative temporal regions, and gives an insight into the model’s prioritisation of ECG parameters during sleep apnoea categorisation.

Transformer Attention Visualization: Using Transformer Attention visualisation, contextual learning behaviour in sequential ECG representations was explained. We performed attention analysis on the locations we found that were emphasized more during the prediction generation. The visualisation allows us to comprehend the dependencies modelled by the Transformer encoder and to choose emphasise informative signal patterns during the entire decision-making process by looking at the intensity of the attention across the temporal positions.

IV. Results And Discussion

A. Experimental Setup

The experimental evaluation was performed to evaluate the performance and learning behaviour of the proposed hybrid CNN–Transformer framework for sleep apnoea identification based on ECG signals. The model training was optimized for the steady convergence and reliable performance estimate, so the model training was designed with supervised optimisation and preset hyperparameter. The Adam optimiser is used with an initial learning rate of 0.001 and adaptive parameter changes were performed. Binary cross entropy was used as the loss function for the binary classification. A maximum of 50 epochs were used for the training, and the batch size was set to 32. We performed early stopping and dynamically adjusted the learning rate using the ReduceLROnPlateau scheduling technique by monitoring the validation.Early stopping and dynamic learning rate adjustment using ReduceLROnPlateau scheduling technique was used to monitor validation to improve generalisation and to avoid overfitting.

B. Training Performance Analysis

The learning stability and convergence features during the optimization process were studied by analysing the behavior of the proposed hybrid CNN-Transformer model training. The training and validation accuracy have been steadily increasing with the number of epochs, as seen in Fig. 4, reflecting the gradual improvement of the feature learning capabilities. Slow reduction and stabilisation of corresponding loss curves were observed, signifying reduction in the prediction error and optimisation effect in between. The validation performance was similar to that of training, indicating steady generalisation during training. Controlled convergence was used to minimise overfitting effects while ensuring efficient training progression, using early stopping and adaptive learning rate scheduling.

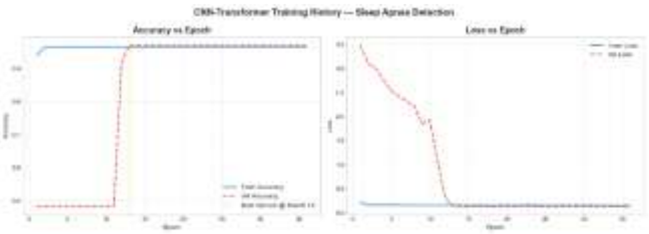

Fig. 4. CNN–Transformer Training History

C. Classification Performance

The ability of the proposed hybrid CNN-Transformer framework to classify was evaluated using a confusion matrix and various quantitative metrics. The confusion matrix (Fig. 5) indicates a good separation between the two classes (Normal, Sleep Apnoea) with a high percentage of correctly classified samples and very few misclassifications. The model achieved 245 true negatives and 277 true positives, while the number of false positives and false negatives turned out to be moderate - 5 samples in each category, which illustrates the balancing of predicting behaviour in both classes. The efficiency of the framework is again confirmed from the quantitative results of Table III. The accuracy, precision, recall and F1 scores are high demonstrating a reliable prediction of the apnoea state and the normal state, with high sensitivity and specificity demonstrating a good capability in prediction of apnoea and the normal state. Moreover, the ROC–AUC value shows that there is good overall class separability.

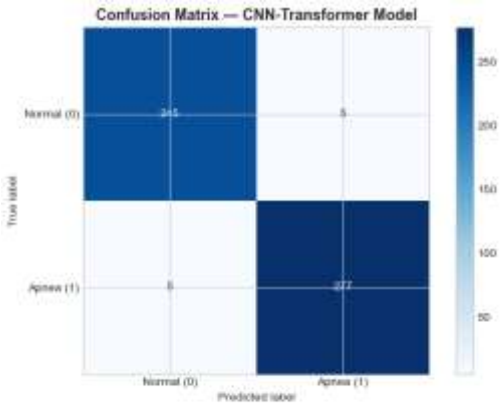

Fig. 5. Confusion Matrix – CNN–Transformer Model

Table 3. Final Model Performance

Metric	Value
Accuracy	98.12%
Precision	0.9812
Recall	0.9812
F1-score	0.9812
Sensitivity	0.9823
Specificity	0.9800
ROC–AUC	0.9870

D. ROC–AUC and Discrimination Analysis

For the assessment of the threshold independent discrimination power the proposed model was analysed using the ROC curve. The ROC profile is good, as seen in figure 6, for all decision thresholds between the Normal and Sleep Apnoea classes. The curve shows a good balance between the TPR and FPR, suggesting that the classification performance remains stable across different operating conditions. The obtained ROC-AUC value of 0.9870 demonstrates a very good discriminative performance and confirms the model's ability to correctly set the decision boundaries between ECG patterns associated with apneas.

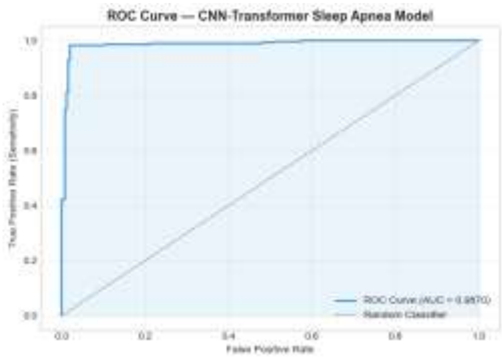

Fig.6 ROC Curve

The experimental results confirm a regular learning behaviour and an effective convergence for the entire training. The proposed hybrid CNN-Transformer system demonstrated high classification performance and high interpretability using the explainability analysis, indicating its applicability for reliable sleep apnoea detection using ECGs.

V. Conclusion

This paper introduced an automatic method for sleep apnoea diagnosis using a hybrid learning approach consisting of CNN and Transformer network processing physiological ECG signals. The aim was to build an intelligent and reliable technique to identify normal and sleep apnoea states to reduce the need on manual interpretation and traditional feature-driven analysis. The proposed solution aims to be able to extract and interpret the meaningful temporal patterns in a sequence of ECG observations, by combining the principles of complementary representation learning in a single framework. The developed framework comprised a comprehensive end-to-end process from preparing physiological signals to automatic categorization, quantitative performance evaluation and explainable interpretation. The experimental results demonstrated that the proposed approach could yield a steady learning behaviour, good classification ability, and a consistent discrimination between the target classes. In addition to predictive effectiveness, the combination of explainability methods helped with enhanced transparency by offering interpretable insights into model decision behaviour and signal areas affecting forecasts. To sum up, the implementation illustrates the feasibility of hybrid DL approaches for physiological signal processing, revealing their significance in assisting the development of efficient, data-driven and user-friendly sleep apnoea detection systems in computer-aided healthcare settings.

The potential future directions of this framework include further strengthening of the framework, scalability and application of this framework in practice for automated sleep apnoea detection. The model can be adapted to accept multiple modalities of physiological inputs and add other biological signals to enhance the context of sleep-related phenomena. Future studies could explore a subject independent evaluation to support generalisation of the findings to other populations and clinical settings. The light-weighted deployment optimisation may make possible the application of the system in wearable and edge-based healthcare systems for continuous system monitoring. Besides, improved explainability methodologies and adaptive attention mechanisms may be researched to further improve the interpretability, dependability and decision support ability in intelligent sleep health applications.

References

[1] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New England Journal of Medicine*, vol. 328, no. 17, pp. 1230–1235, Apr. 1993.

[2] J. F. Benjafield *et al*., "Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis," *Lancet Respiratory Medicine*, vol. 7, no. 8, pp. 687–698, 2019.

[3] S. Javaheri, R. M. Barbé, and R. L. Campos-Rodriguez, "Sleep apnea: Types, mechanisms, and clinical cardiovascular consequences," *Journal of the American College of Cardiology*, vol. 74, no. 5, pp. 706–721, 2019.

[4] K. E. Bloch, "Polysomnography: A systematic review," *Technology and Health Care*, vol. 5, no. 4, pp. 285–305, 1997.

[5] W. Flemons, "Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research," *Sleep*, vol. 22, no. 4, pp. 667–689, 1999.

[6] T. Penzel and T. M. Schöbel, "New methods for screening and diagnosis of sleep apnea," *Frontiers in Digital Health*, vol. 3, 2021.

[7] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ECG database," in *Proc. Computers in Cardiology*, 2000, pp. 255–258.

[8] A. L. Goldberger *et al*., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[10] A. Vaswani *et al*., "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[11] A. Zarei and B. Mohammadzadeh Asl, "Automatic detection of obstructive sleep apnea using wavelet transform and entropy-based features from single-lead ECG signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1011–1021, 2019.

[12] 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.

[13] K. Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[14] 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, p. e7731, 2019.

[15] M. Bahrami and M. Forouzanfar, "Detection of sleep apnea from single-lead ECG: Comparison of deep learning algorithms," in *Proc. IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2021, pp. 1–5.

[16] S. Hu, W. Cai, T. Gao, J. Zhou, and M. Wang, "Robust wave-feature adaptive heartbeat classification based on self-attention mechanism using a transformer model," *Physiological Measurement*, vol. 42, no. 12, Art. no. 125001, 2021.

[17] S. Hu, W. Cai, T. Gao, and M. Wang, "A hybrid transformer model for obstructive sleep apnea detection based on self-attention mechanism using single-lead ECG," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[18] H. Liu, S. Cui, X. Zhao, and F. Cong, "Detection of obstructive sleep apnea from single-channel ECG signals using a CNN–Transformer architecture," *Biomedical Signal Processing and Control*, vol. 82, Art. no. 104581, 2023.

[19] A. Natarajan, "A wide and deep transformer neural network for 12-lead ECG classification," in *Proc. Computers in Cardiology*, 2020, pp. 1–4.

[20] A. Dosovitskiy *et al*., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.