International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

CONVOLUTIONAL VISION TRANSFORMER AND GRU BASED FRAMEWORK FOR ACCURATE SLEEP APNEA DETECTION FROM ECG SIGNALS

[1]Soorya Gayathri J,[2]Dr. S. Thomas George,[3]Dr. K. Martin Sagayam

[1]Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India, 92soorya@gmail.com

[2]Professor, Department of Biomedical Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India, thomasgeorge@karunya.edu

[3]Assistant Professor, Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India, martinsagayam.k@gmail.com

Abstract

Sleep apnea (SA), characterized by recurrent disruptions in breathing during sleep, poses critical health risks such as cardiovascular complications and cognitive dysfunction, highlighting the critical nature of this commonly undiagnosed disorder. The early and accurate detection hence becomes essential for timely intervention and effective management. This study proposes a hybrid deep learning model integrating Convolutional Vision Transformer (CvT) and Gated Recurrent Unit (GRU) architectures to classify SA using the ECG Data for Sleep Apnea Detection dataset. The CvT module extracts rich spatial features from the ECG time series through convolution-based patch embeddings and hierarchical transformer layers, enhancing the model to capture local and global dependencies. The GRU layer further processes these features to model temporal dynamics and sequential dependencies essential for identifying apnea patterns. The synergistic integration of CvTâ□□s attention-driven spatial encoding with GRUâ□□s sequential learning captures morphological as well as temporal features, critical in the sleep apnea classification. The proposed framework maintains a lightweight and scalable design, supporting robust inference even under resource-constrained conditions. The CvT-GRU hybrid model demonstrates superior classification performance, with an accuracy of 98.70% and an F1-score of 0.99 for both apnea and normal classes. The results confirm the modelâ□□s efficacy in automated apnea detection from ECG signals and demonstrate significant potential for clinical applications and remote health monitoring.

Keywords: Sleep Apnea, Electrocardiogram, Convolutional Vision Transformer, Gated Recurrent Unit, Deep Learning, Artificial Intelligence

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

INTRODUCTION

Sleep apnea, characterized by episodic reductions or cessations in breathing, presents a major health concern due to its association with multisystem health issues [1]. These disruptions, commonly referred to as apneic events can occur in durations of seconds to over a minute and can extend in frequency to even hundreds of times a night. The most common form, Obstructive Sleep Apnea (OSA), arises from partial/complete obstruction of the upper airway, generally caused by the relaxation of throat muscles during sleep [2]. Other variants include Central Sleep Apnea (CSA), stemming from neurological signalling issues; and the third one: Mixed Sleep Apnea is a mix of both cases in terms of features [3]. Typical symptoms include loud snoring, gasping for air, excessive daytime sleepiness, morning headaches, poor concentration, mood disturbances etc. Beyond sleep disruption, untreated sleep apnea is closely linked to severe comorbidities such as hypertension, atrial fibrillation, stroke, metabolic syndrome and type 2 diabetes, elevating both cardiovascular and all-cause mortality risk [4]. Figure 1 represents different health risks associated with sleep.

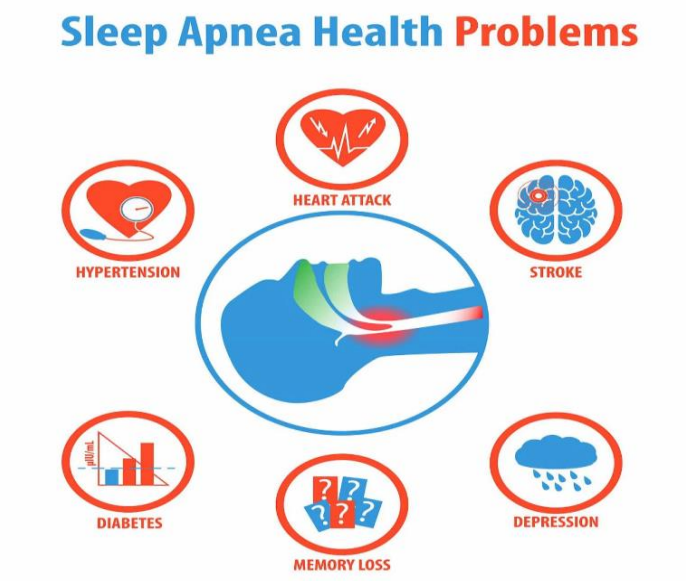

Fig.1. Different Health Risks Associated with Sleep Apnea

The clinical urgency in addressing sleep apnea is based not only from its physiological consequences but also from its insidious underdiagnosis. A substantial proportion of affected individuals remain undiagnosed particularly in low-resource settings and among asymptomatic or minimally symptomatic patients. Early diagnosis followed by suitable clinical and non-clinical approaches including continuous positive airway pressure (CPAP) therapy, lifestyle modification, surgical correction etc. can markedly improve sleep quality, cognitive function and cardiovascular health outcomes [5] [6]. Thus, the identification of OSA, particularly in its early or subclinical stages, constitutes a public health priority.

Received: August 07, 2025

Traditionally, the widely accepted technique for OSA diagnosis is polysomnography (PSG), a comprehensive overnight evaluation involving the simultaneous recording of various physiological signals such as EEG (Electroencephalogram), EOG (eye movements), ECG, airflow, thoracoabdominal movements and oxygen saturation [7]. While PSG offers high diagnostic precision, it is inherently resource-intensive, requiring overnight clinical supervision, expensive equipment and manual scoring by trained technicians. The discomfort associated with multiple sensors further jeopardize the sleep architecture, inadvertently affecting diagnostic accuracy [8]. Overnight pulse oximetry that detects oxygen desaturation events, but lacks the ability to distinguish between apnea subtypes or assess sleep stages. Questionnaire-based tools like STOP-BANG, the Berlin Questionnaire and the Epworth Sleepiness Scale offer quick risk assessments but are prone to subjective bias and low diagnostic precision [9]. Home Sleep Apnea Testing (HSAT) provides portable alternatives by recording limited physiological signals, but often underestimates severity due to the absence of EEG-based sleep staging. Actigraphy offers insights into sleep-wake cycles but cannot detect respiratory disturbances directly. Even though the methods offer better accessibility than PSG, they frequently suffer from limited accuracy, insufficient temporal resolution and reduced generalizability across diverse populations.

The AI based advanced sleep apnea detection methods usually employ machine learning (ML) and deep learning (DL) techniques [10]. Biomedical parameters including ECG, SpO2 waveforms, airflow recordings and even acoustic data has been utilised to identify apnea episodes in real time or offline settings. ML, DL and hybrid models have demonstrated promising accuracy. However, the frameworks fall short at times due to poor interpretability, high computational complexity, lack of generalizability, dependency on noise-prone signals and reliance on annotated training data or engineered features. Given the heterogeneous clinical presentation of sleep apnea and the technological shortcomings in current approaches, there remains a critical need to develop a robust, scalable and interpretable model that not only performs well across diverse populations and signal types but also ensures deployment feasibility in real-world clinical or home-based contexts. The core contributions of the study are as presented:

- Development of a sequential hybrid framework integrating CvT and GRU for reliable sleep apnea detection from ECG signals, enabling precise identification of subtle apnea patterns across diverse cardiac signal variations.
- High-efficiency and scalable framework for real-time sleep disorder screening, optimized for low computational complexity and possible deployment in clinical and wearable settings.

The subsequent portions of the research are laid out in the following sections: Section 2 offers a detailed analysis of the latest advances in sleep apnea detection, highlighting the current research constraints. Section 3 delineates the proposed methodology. The empirical findings are

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

highlighted in Section 4 along with a comprehensive assessment of the model's functionality. Section 5 concludes the research by summarising the main conclusions and highlighting potential directions for future research.

RELATED WORKS

Kolhar et al. [11] investigated the efficacy of DL models, particularly Convolutional Neural Networks (CNNs), in detecting OSA from ECG signals. A dual-branch CNN architecture processed ECG signals through two parallel convolutional pathways for learning both local and global temporal patterns simultaneously. Each branch extracted distinct feature representations that are later concatenated and fed to fully connected layers for final classification. The framework achieved an accuracy of 94%, outperforming traditional Random Forest (RF) and Decision Tree (DT) methods. Overfitting risk and limited generalizability due to dataset homogeneity constrained the study's potential.

Huang et al. [12] suggested an ML-based screening model for OSA by leveraging clinical biochemical indicators and demographic variables. The study utilized a dataset of 4,124 patients referred to the Sleep Medicine Center of Fujian Medical University for around three years and included demographic attributes, routine biochemical markers and PSG-confirmed OSA status. Both univariate and multivariate analyses were employed to identify significant predictors, followed by feature ranking using the Boruta algorithm. Ten key features were analyzed, utilizing four classification architectures: Gaussian Naive Bayes (GNB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Logistic Regression (LR), where LR outperformed others with an AUC of 0.732. The biomedical markers: triglyceride-glucose (TyG) index and anthropometric variables were identified as enhancing factors in OSA screening. The limited geographic scope, exclusion of genetic and socio-economic variables and overreliance on data from a clinical cohort with higher-than-average OSA prevalence hampered the generalizability of the study.

Liu et al. [13] suggested an ML based architecture for the detection and classification of OSA using ECG signals. Data of 1,656 patients from the China Medical University Hospital was collected and an EfficientNet architecture, selectively employing convolutional layers for feature extraction was designed in the study. Preprocessing steps such as overlapping slicing and sample-weight adjustments were evaluated, revealing that overlapping slicing increased apnea signal capture probability considerably. The EfficientNet model, combined with the selected preprocessing techniques, achieved an accuracy of 85.5% for segment-level apnea detection. For severe OSA screening, the model was integrated with XGBoost, yielding an accuracy of 92.8%. The absence of PSG-based sleep stage features restricted the study's potential to provide a more holistic analysis of sleep-related breathing disorders.

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Jiménez-García et al. [14] proposed an interpretable DL framework for paediatric OSA detection by integrating CNN and RNN with explainable AI (XAI) techniques. Data comprising 1,638 annotated PSG recordings from the Childhood Adenotonsillectomy Trial (CHAT) dataset and 974 samples from a proprietary clinical database were employed in the study. Using airflow (AF) and oximetry (SpO2) signals segmented into 30-minute windows, the framework estimated the apnea-hypopnea index (AHI) and classified OSA into four severity levels. To highlight discriminative features in the input signals, notably AF cessations and SpO_2 desaturation events, Gradient-weighted Class Activation Mapping (Grad-CAM) was utilised. The study achieved an intraclass correlation coefficient above 0.9 for AHI regression and 4-class classification accuracies of 74.51% (CHAT) and 62.31% (private dataset). For binary classification across increasing AHI thresholds, the framework achieved an accuracy of 84%. Overreliance on a single XAI method and absence of multicentric or ambulatory data hampered the scalability of the study.

Bhongade & Gandhi [15] suggested WIVIDOSA-Net, a DL framework for OSA identification from smoothed Wigner–Ville spectrograms (SWVSs) derived from single-lead ECG signals. The PhysioNet Apnea-ECG database, comprising 70 full-night ECG recordings from 32 individuals was utilized in the study. A time–frequency transformation pipeline was employed to convert the one-minute ECG segments into WVD-based spectrograms, which were later smoothed by Savitzky–Golay filtering. The resultant 2D representations were fed to WIVIDOSA-Net, comprising six CNN layers, four MPL layers and a fully connected layer. The framework achieved a classification accuracy of 90.09%, outperforming ResNet-18 and ResNet-50. The exclusive use of spectrogram-based inputs, that usually neglected temporal features in raw ECG signals hampered the depth and generalizability of the study.

Thompson et al. [16] suggested a 1D CNN architecture for automated identification of OSA from ECG signals, benchmarked against conventional ML classifiers including SVM and RF. The PhysioNet Apnea-ECG database was employed in the study, comprising 35 overnight ECG recordings. The 1DCNN architecture incorporated a convolutional layer, max pooling, a flattened dense layer and a final multilayer perceptron (MLP) with SoftMax activation. Among the 15 trained models, the 1DCNN-500 configuration outperformed others with 96.99% accuracy, followed by RFC-500 (91%), while SVM-500 lagged with only 72% accuracy and exhibited unbalanced classification. The usage of a small dataset with just 35 ECG recordings hampered the generalizability across varied populations and real-world clinical scenarios.

Arslan [17] proposed a two-layer hybrid model combining DL and ML techniques to detect sleep apnea and classify its subtypes using multi-sensor PSG data. The first layer consisted of four distinct DL architectures: Deep Neural Network (DNN), Gated Recurrent Unit (GRU), RNN and Long Short-Term Memory (LSTM) each trained to make preliminary predictions based on a 23-dimensional feature vector comprising snoring, oxygen saturation, arousal index and sleep stage

Received: August 07, 2025

Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

data. A second-layer meta-learner processed the resultant outputs and acted as a decision-level classifier, verifying or correcting the initial results. The framework achieved 95.74% accuracy in detecting sleep apnea events. The complex two-layer architecture integrating multiple DL models increased computational overhead and reduced feasibility for real-time deployment in resource-constrained clinical environments.

Kang et al. [18] proposed an AI-driven model for automated sleep stage classification and OSA risk identification using power spectral density (PSD) features extracted from EEG signals. The study utilized standard PSG data collected from 139 participants aged 18–65 at Ewha Womans University Mokdong Hospital. Three separate models were generated based on the age category: general, younger age-specific and older age-specific models. Feature extraction from EEG was followed by classification using SVM, KNN and MLP algorithms. MLP emerged as the top performer with 73% accuracy in both sleep stage classification and OSA risk identification, while the younger age-specific model outperformed the general model in certain sleep stages, while the older age-specific model underperformed highlighting discrepancies stemming from demographic stratification. The reliance on manually engineered PSD features, where the EEG signal's complex temporal and non-linear dynamics are not captured completely, limited the scalability and adaptability of the study.

Wang et al. [19] suggested OSAnet, a deep CNN to detect sleep apneic events and estimate AHI from noncontact audio recordings of sleep sounds. Recordings from 135 participants with habitual snoring or heavy breathing, simultaneously monitored using PSG were utilized in the study. A naturalistic room setting without noise attenuation formed the platform for capturing sleep sounds and OSAnet performed event-by-event detection. The framework achieved an accuracy of 93.2% in detecting severe OSA (AHI $\geq$ 30). Across multiple AHI cutoffs, the model maintained high diagnostic accuracy, with values of 91.5% (AHI $\geq$ 5), 81.3% (AHI $\geq$ 10) and 91.5% (AHI $\geq$15). The reliance on clearly audible respiratory sounds limited applicability in cases where snoring or breathing is minimal or obscured by background noise.

Lin et al. [20] proposed RAPIDEST, an AI-based framework designed to detect OSA by analyzing rare transitions in sleep stages derived solely from EEG signals. A rarity score was generated that quantified the unusualness of full-night sleep stage sequences. Three datasets: Sleep-EDF, UCDDB and Wisconsin Sleep Cohort (WSC) were utilized in the model evaluation and OSA detection accuracy reached 80.00% on the UCDDB dataset and 69.15% on the WSC dataset. By relying exclusively on EEG data, the model significantly reduced the complexity of signal acquisition and enabled more accessible sleep monitoring. The inability to localize or identify the type of sleep disorder, as the rarity score is computed over entire sleep sessions rather than per-epoch segments, increased false positives in healthy individuals that exhibited isolated rare sleep patterns.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Molnár et al. [21] suggested the use of AI for preliminary screening of OSA using easily obtainable anthropometric, demographic and questionnaire-based data. Data from 100 patients from Semmelweis University, categorized into non-OSA, mild OSA and moderate-to-severe OSA groups based on PSG diagnosis was utilized in the study. A supervised learning algorithm was trained to classify OSA presence and severity using structured tabular data, with feature weights derived from correlations with known diagnostic outcomes. The predictive model mapped the input features to categorical OSA labels through a decision-boundary learning process. The framework achieved a prediction accuracy of 81% using only BMI, gender and age and improved to 83% when questionnaire data was included. The Epworth questionnaire alone yielded 75% accuracy, whereas the Berlin questionnaire achieved 62%. The primary limitation was the model's susceptibility to overfitting with the small and demographically imbalanced sample.

Urtnasan et al. [22] suggested a DL-based model to predict the incidence of OSA from ECG signals extracted from PSG data. The study utilized a subset of the publicly available MrOS Sleep Study, comprising 55 elderly male participants aged 65 and above, with 30 healthy controls and 25 diagnosed OSA patients. The model architecture consisted of a five-layer CNN that incorporated convolutional, pooling and fully connected layers for end-to-end learning. ECG recordings were segmented into 10-second intervals and directly input into the CNN, which automatically extracted temporal features and classified each sample via a softmax-activated output layer. Evaluation revealed that the framework predicted the OSA events with 82.2% accuracy. The exclusive use of ECG signals without incorporating other complementary physiological signals limited the study in differentiating apnea subtypes or in detecting complex cases.

Paul et al. [23] suggested an AI based real-time sleep apnea identification model leveraging SpO2 and ECG signals, both separately and combined. Using the PhysioNet Apnea-ECG dataset, the study extracted R-R intervals from ECG signals and processed 1-minute-long sequences through a feed-forward deep artificial neural network (ANN). Three separate models were trained: SpO2-only, ECG-only and a combined model. The combined model outperformed others with 91.83% accuracy, followed by the SpO2-based model (90.78%) and the ECG-based model (80.04%). The fusion of both signal types improved robustness, as the complementary characteristics of oxygen desaturation and heart rate variability captured more reliable apnea patterns. The overreliance on machine-generated QRS annotations for R-R interval extractions, which are prone to inaccuracies, affected the model performance adversely.

Hemrajani et al. [24] proposed an efficient DL-based framework for detecting OSA using ECG signals, employing the PhysioNet Apnea-ECG dataset. Three architectures: MobileNet V1 alone, MobileNet V1 combined with LSTM and MobileNet V1 combined with GRU were evaluated. MobileNet V1 served as a lightweight CNN for spatial feature extraction, while the LSTM and

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

GRU components captured temporal dependencies in ECG sequences. The MobileNet V1 model achieved an accuracy of 89.5%, while the MobileNet V1 with LSTM and MobileNet V1 with GRU configurations yielded accuracies of 90% and 90.29%, respectively. High dependence on accurate real-time ECG signal quality, signal artefacts or noise during acquisition significantly degraded classification performance, particularly in ambulatory or wearable settings.

Javeed et al. [25] developed a hybrid ML framework for predicting OSA using electronic health records (EHR) from the Swedish National Study on Aging and Care (SNAC), comprising 10,765 samples and 75 features. The framework integrated an XGBoost module for feature importance ranking and classification by a Bidirectional Long Short-Term Memory (BiLSTM) network, enabling both efficient dimensionality reduction and temporal sequence modeling. The framework achieved a classification accuracy of 97% employing top six most influential features, which included type 2 diabetes, respiratory disorders, psychological stress indicators and behavioral metrics. The absence of validation across longitudinal or time-separated patient records affected the model stability and reliability, hampering its long-term applicability.

2.1 Research gap

Despite notable advances in the application of AI for SA detection, several critical research gaps persist. Many studies have achieved high classification accuracy using physiological signals such as ECG, SpO2, airflow and EEG; however, a lack of model generalizability due to homogenous or demographically constrained datasets remains a recurring concern [10] [14] [20]. A number of models rely heavily on manually engineered features or black-box representations, which limit clinical interpretability [18]. Several frameworks also suffer from dependence on external preprocessing tools (e.g., unaudited QRS detectors) or high-complexity architectures that hinder real-time deployment, especially in wearable or resource-limited settings [23] [24]. Furthermore, existing studies rarely incorporate temporally dynamic or evolving patient profiles, overlooking the importance of longitudinal data validation for chronic conditions like OSA [25]. Hence, there is a need for an explainable, multi-modal, low-complexity framework that supports generalizable, real-time and longitudinal OSA risk assessment across heterogeneous populations.

MATERIALS AND METHODS

The proposed study integrates a Convolutional Vision Transformer with a Gated Recurrent Unit to detect SA. The publicly available ECG Data for Sleep Apnea Detection, from Kaggle that comprises annotated ECG recordings labeled as normal or sleep apnea is utilized in the study. The standardized input sequences are initially passed through CvT layers, that extract spatial and frequency-aware representations. These learnt features form the input of the GRU layer that models the temporal dependencies present in the sequential data. Batch normalization is applied to stabilize the training dynamics, followed by a dense classification head that outputs binary labels corresponding to normal or apnea classes. This architecture is designed to leverage both

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

local pattern extraction and long-range temporal learning for robust sleep apnea classification. Figure 2 illustrates the basic architecture of the proposed model.

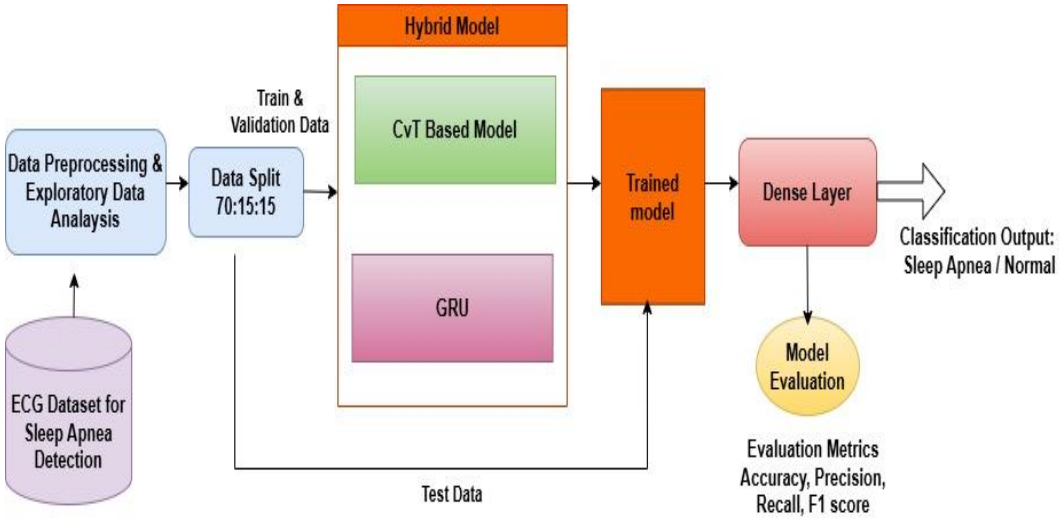

Fig.2. Basic Architecture of the Proposed Model

3.1 Dataset Description

The study employs the publicly available ECG Data for Sleep Apnea Detection dataset, from Kaggle, designed to support the automated identification of SA events using ECG signals [26]. Given the time-series nature of ECG signals, the dataset provides a valuable modality for inferring physiological disruptions associated with sleep apnea. Each instance in the dataset comprises sequential ECG signal recordings captured over time, representing the dynamic heart activity of an individual across various sleep stages. The data points are sampled at regular intervals and contain two primary dimensions: the time index and the corresponding ECG signal amplitude. This facilitates the extraction of temporal and morphological features that are crucial for discriminating between normal and apneic states. The dataset is structured for binary classification tasks, with annotations for the presence or absence of SA in each sample. The dataset's compatibility with lightweight neural architectures and its open-access availability makes it a practical resource for developing deployable apnea detection systems, particularly in wearable or real-time monitoring contexts.

Figure 3 illustrates the dataset head comprising numerical time-series segments of ECG signals, where each row corresponds to a 2500-sample window extracted from a continuous ECG recording. As illustrated in the figure, each column represents a specific timestamped ECG amplitude value, indicating the dynamic electrical activity of the heart. The final column labelled "Target" provides the ground truth annotation for each corresponding signal segment, with binary class labels: either Sleep Apnea or Normal.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Dataset Head:

	0	1	2	3	4	5	6	\
0	0.012976	0.022008	0.049401	0.007309	0.080700	0.084465	0.053040	
1	0.007665	0.011024	0.014001	0.115614	0.045236	0.053854	0.144119	
2	0.044957	0.028612	0.085881	0.018910	0.078694	0.103297	0.046348	
3	-0.011676	0.027831	0.029627	0.021658	0.068194	0.075705	0.095863	
4	-0.008188	0.001010	-0.009165	0.061274	0.087704	0.055419	0.120823	

	7	8	9	...	2491	2492	2493	2494	\
0	0.110527	0.147975	0.158048	...	-0.086520	-0.076036	-0.073070	-0.075017	
1	0.093378	0.178889	0.133928	...	-0.590053	-0.585818	-0.504592	-0.412495	
2	0.148435	0.148251	0.140560	...	-0.663029	-0.596072	-0.501529	-0.432133	
3	0.123471	0.155526	0.150709	...	-0.085069	-0.117560	-0.108299	-0.083278	
4	0.107706	0.133526	0.166235	...	-0.114268	-0.061457	-0.097293	-0.085431	

	2495	2496	2497	2498	2499	Target
0	-0.077458	-0.051301	-0.029699	-0.044448	0.015390	Sleep Apnea
1	-0.410130	-0.247508	-0.543017	-0.138173	-0.020498	Normal
2	-0.396986	-0.308740	-0.195283	-0.085512	0.001185	Normal
3	-0.067660	-0.010860	-0.030781	-0.007939	0.019734	Sleep Apnea
4	-0.065468	-0.049709	-0.013976	-0.011729	0.014483	Sleep Apnea

Fig.3. Dataset Sample

3.2 Exploratory Data Analysis (EDA)

EDA was performed to analyse the structure and distribution of ECG signals and their relationship to sleep apnea. Key steps included examining class balance, analysing statistical properties across features and visualizing representative signal patterns for both 'Normal' and 'Sleep Apnea' classes. The analysis also helps in identifying potential noise, outliers and inconsistencies that informed the preprocessing strategy. Figure 4 illustrates the comprehensive statistical analysis conducted on the ECG signal dataset, encompassing all 2,500 signal columns. Each column represents a sequential time point across the ECG waveform and the summary statistics: mean, standard deviation, minimum, quartiles and maximum, offer valuable insights into the amplitude variation across these points.

Summary Statistics:

	0	1	2	3	4	\
count	2660.000000	2660.000000	2660.000000	2660.000000	2660.000000	
mean	-0.000081	0.015979	0.035886	0.051289	0.068840	
std	0.067291	0.066560	0.058968	0.063981	0.065107	
min	-0.444399	-0.444715	-0.420871	-0.421869	-0.394914	
25%	-0.016200	0.001065	0.019272	0.036224	0.053729	
50%	0.000471	0.017652	0.035223	0.052665	0.070564	
75%	0.016875	0.034704	0.052325	0.070078	0.087606	
max	0.378275	0.402286	0.380593	0.410794	0.455771	

	5	6	7	8	9	...
count	2660.000000	2660.000000	2660.000000	2660.000000	2660.000000	...
mean	0.085575	0.104296	0.121735	0.138715	0.157789	...
std	0.065125	0.066458	0.058638	0.057902	0.060768	...
min	-0.395065	-0.349534	-0.328775	-0.316848	-0.313061	...
25%	0.070032	0.089174	0.106737	0.124190	0.141844	...
50%	0.087321	0.106638	0.123258	0.139788	0.158176	...
75%	0.104585	0.122310	0.139650	0.157845	0.175342	...
max	0.447308	0.456807	0.487887	0.501368	0.530445	...

	2490	2491	2492	2493	2494	\
count	2660.000000	2660.000000	2660.000000	2660.000000	2660.000000	
mean	-0.395727	-0.373458	-0.342772	-0.306010	-0.264624	
std	0.283938	0.275371	0.258188	0.234781	0.209291	
min	-1.120864	-1.104580	-1.043668	-0.992015	-0.949979	
25%	-0.673397	-0.636857	-0.590583	-0.530016	-0.460279	
50%	-0.343775	-0.300568	-0.264135	-0.222023	-0.146276	
75%	-0.116823	-0.103511	-0.091596	-0.077365	-0.063412	
max	-0.048903	-0.037853	-0.007640	-0.017860	-0.002640	

Fig.4. Statistical Analysis

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

An essential preliminary step in data validation involved checking for missing values across all features and the target label. As illustrated in Figure 5, the dataset is entirely free of null or missing entries in all 2,500 signal columns and in the target column. This indicates excellent data integrity and ensures that no imputation or row elimination is required prior to model training.

```
Missing Values:
 0          0
1           0
2           0
3           0
4           0
           ..
2496        0
2497        0
2498        0
2499        0
Target      0
Length: 2501, dtype: int64
```

Fig.5. Missing Values

The target variable distribution plot illustrated in Figure 6 shows a near-perfect balance between the two classes: Sleep Apnea and Normal. Each class comprises approximately half of the total samples, with minimal discrepancy in frequency. This balanced distribution ensures that the classification model does not suffer from class imbalance bias during training, thereby supporting unbiased learning across both conditions.

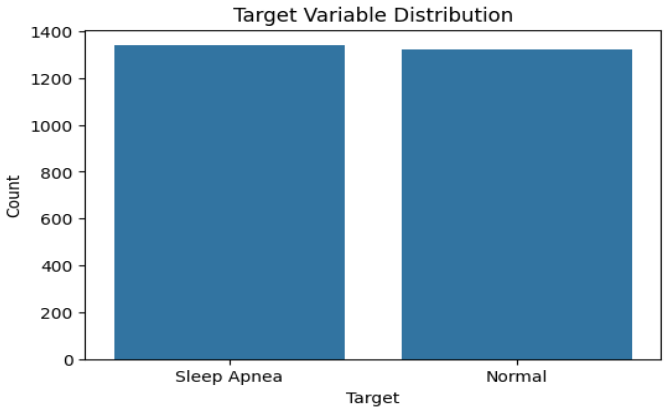

Fig.6. Target Distribution

The time series line plot illustrated in Figure 7 visualizes the temporal variation of the first five ECG-derived features across the 2,500-time indices. Each colored line represents a distinct feature, capturing the signal amplitude fluctuations over time. The plot reveals consistent rhythmic patterns interspersed with sharp peaks and troughs, which are characteristic of

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

physiological ECG signals. The overlapping signals indicate the complexity and high dimensionality of the dataset, emphasizing the importance of advanced modeling techniques to capture the temporal dependencies and feature correlations.

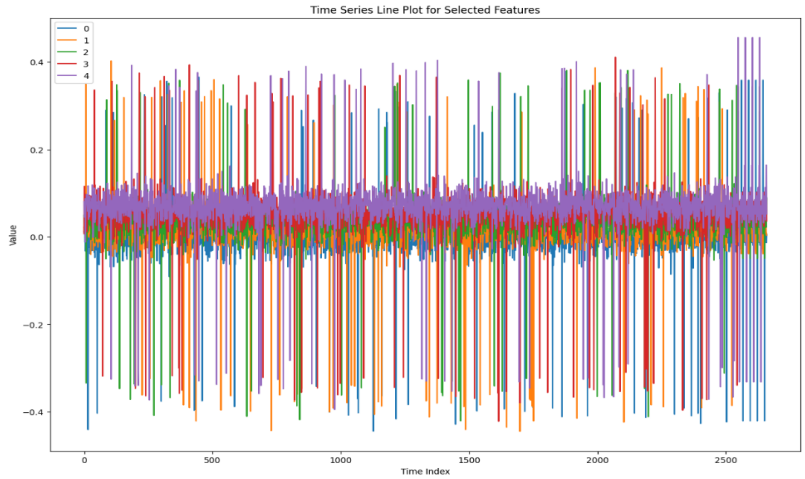

Fig.7. Time Series Line Plot for First 5 Features

The temporal distribution of the target variable is illustrated in Figure 8 using a binary line plot, where each point represents the classification of a sample as either normal or indicative of sleep apnea. The y-axis is labeled with categorical values 'Normal' corresponding to label 0 and ‘sleep Apnea' to label 1 mapped across the entire time index on the x-axis. The consistent interspersing of both categories throughout the dataset indicates a well-balanced temporal representation of the two classes.

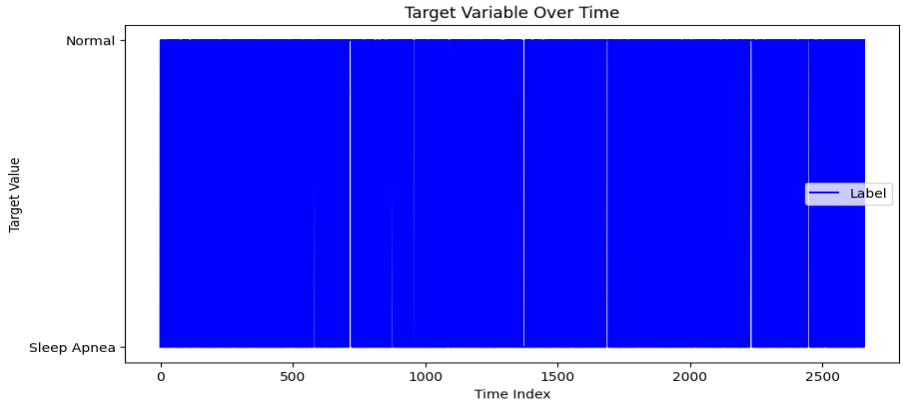

Fig.8. Target Variable over Time

Figure 9 illustrates the autocorrelation plots for features 0, 1 and 2 to assess the temporal dependence of these variables across various lags. All three plots exhibit autocorrelation values that remain close to zero across the full lag range, indicating a lack of significant periodicity or time-dependent patterns in the signal. This suggests that the features do not exhibit significant

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

temporal self-similarity, which is consistent with the nature of physiological signals captured in short windows.

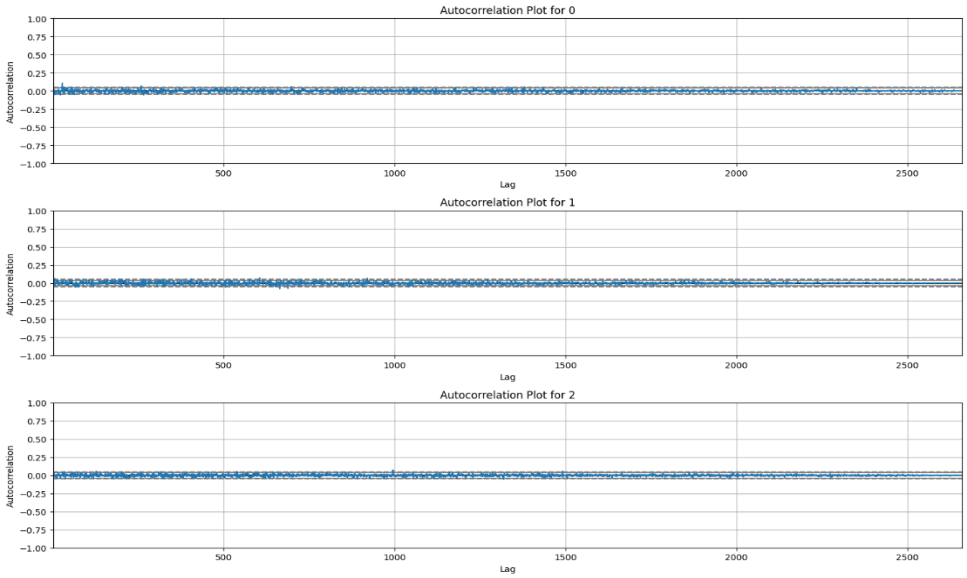

Fig.9. Autocorrelation Plot

The rolling statistics plot in Figure 10 illustrates the temporal evolution of the mean and standard deviation for the first three selected features using a sliding window of size 50. Solid lines represent the rolling mean, while dashed lines indicate the corresponding rolling standard deviation. The rolling means for all three features exhibit mild fluctuations around zero, suggesting a relatively stationary behavior in terms of central tendency. In contrast, the rolling standard deviation curves demonstrate more pronounced variability, capturing dynamic changes in signal volatility over time. These patterns highlight localized bursts of variance, which may contain diagnostic cues relevant to distinguishing sleep apnea episodes from normal states.

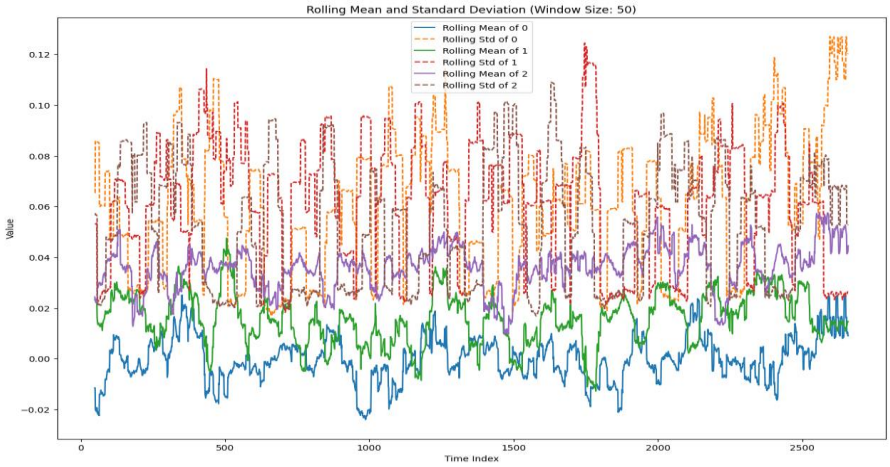

Fig.10. Rolling Mean and Standard Deviation

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.3 Data Preprocessing

The data preprocessing step ensures that the dataset is clean, standardized and suitable for effective model training. In the proposed framework, preprocessing includes normalization of input features to a common scale and encoding of categorical labels into numerical form. These transformations help stabilize the learning process and improve convergence during training. Subsequently, this processed dataset is split as training, validation and testing sets using stratified sampling to preserve the class distribution across all subsets. Specifically, 70% (1803) is allocated for training, 15% (386) to the validation; and the remaining 15% (387) for testing. The Min-Max normalization is applied to the dataset to maintain proportional influence from each feature in the learning process, also ensuring to avoid dominance by features with larger numerical ranges. This technique transforms the original feature values to a standardized scale, in the range [0,1] as given in Equation (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x, x_{min} and x_{max} represents original, minimum and maximum values respectively. Now the label encoding converts the categorical output variable into a numerical format suitable for classification. The labels indicating SA presence/absence are initially represented as textual classes such as “Normal” and “Apnea” in the dataset. These are encoded into binary integers: 0 to “Normal” and 1 to “Apnea” as per Equation (2).

$$Encoded\ Label\ y_i = \begin{cases} 0, & if\ y_i = Normal \\ 1, & if\ y_i = Apnea \end{cases} \tag{2}$$

Figure 11 illustrates the class distribution in the training, validation and test sets after stratified splitting. Both classes: Normal (Class 0) and Sleep Apnea (Class 1) are evenly represented across all subsets, confirming that the stratification effectively preserves class balance. Specifically, the training set contains nearly equal instances of both classes, while the validation and test sets also exhibit a balanced distribution. This balance is essential for unbiased model training and performance evaluation.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

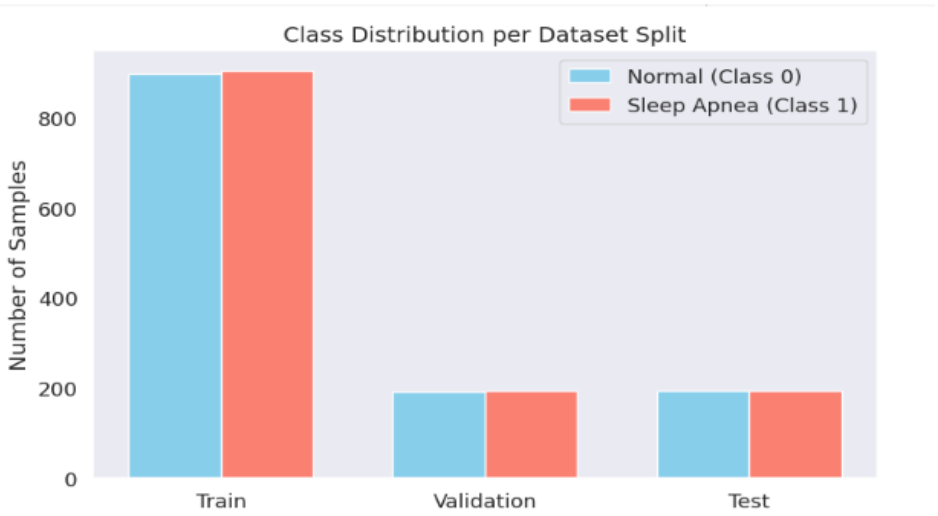

Fig.11. Class Distribution per Dataset Split

3.4 Model Development

Effective identification of SA from ECG signals requires a model capable of capturing spatial intricacies and temporal dependencies inherent in the data. The proposed architecture addresses this need by integrating the CvT for rich spatial feature extraction and GRU for learning sequential patterns. The hybrid architecture captures ECG dynamics in detail, improving the system's accuracy in detecting apnea-related events.

3.4.1 Convolutional Vision Transformer (CvT)

CvT is a hybrid DL architecture that combines the global modelling capabilities of Vision Transformers (ViTs) with the spatial inductive biases of CNNs [27]. Unlike traditional ViTs that split images into flat, non-overlapping patches, CvT introduces convolutional token embedding to preserve local spatial continuity. For an input image $X \in \mathbb{R}^{H\times W\times C}$, a convolutional layer with kernel size k, stride s and padding p is applied to generate the initial token representation as shown in Equation (3).

$$Z_0 = Conv\ 2D_{k,s,p}(X) \in \mathbb{R}^{H'\times W'\times D} \tag{3}$$

where H' and W' denote the reduced spatial dimensions after downsampling and D, the embedding dimension. Each subsequent stage performs another convolutional embedding, followed by Layer Normalization and a modified Transformer block. To further enhance the

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

spatial encoding, CvT replaces the linear projections to compute the query (Q), key (K) and value (V) matrices in self-attention with convolutional projections as in Equation (4).

$$Q = DWConv_q(Z_0); K = DWConv_k(Z_0); \text{V} = DWConv_v(Z_0) \tag{4}$$

where $DWConv_*$ represents a depthwise separable convolution that factorizes a standard convolution into a depthwise convolution followed by a pointwise 1×1 convolution as in Equation (5).

$$DWConv(X) = Conv_{1\times1}\left(Conv_{s\times s}^{depthwise}(X)\right) \tag{5}$$

The projection preserves local spatial structure while reducing computational complexity. Once projected, the multi-head self-attention mechanism operates on it as in Equation (6).

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where d_k represents the dimensionality of the key vectors. To allow multiple types of relational modelling, CvT uses h parallel attention heads and concatenates the outputs as in Equation (7).

$$MHSA(Q,K,V) = Concat(head_1, \ldots, head_h)W^O \tag{7}$$

where each $head_i = Attention(Q_i, K_i, V_i)$ and W^O is a learnable output projection matrix. Finally, the output of each Transformer block is passed through a feed-forward network (FFN) composed of two linear layers with a GELU activation as in Equation (8).

$$FFN(x) = W_2 \cdot GELU(W_1 \cdot x) \tag{8}$$

where x is the input vector, and W_1 and W_2 represent the weight matrix for the first and second linear transformations respectively. By integrating convolution into both the token embedding and attention mechanisms, CvT benefits from local spatial feature learning, efficient downsampling and global sequence modelling offering a superior hybrid structure for visual tasks. Figure 12 illustrates the basic architecture of CvT.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

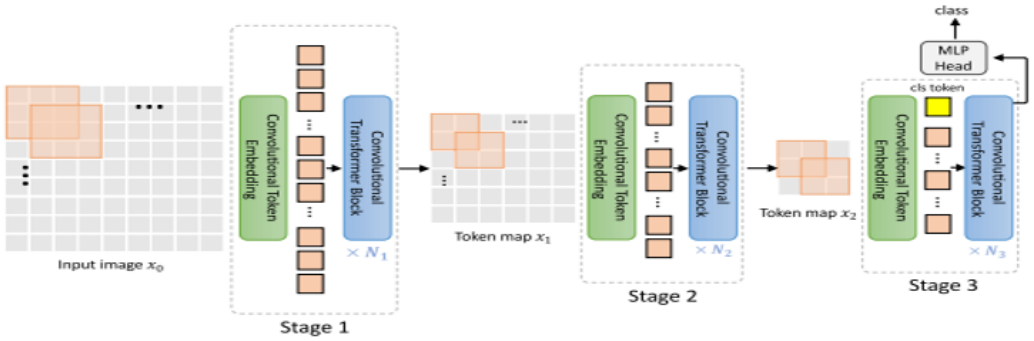

Fig.12. Basic Convolutional Vision Transformer Architecture

3.4.2 Gated Recurrent Unit

GRU is a streamlined version of RNN that efficiently models temporal dependencies in sequential data [28]. It simplifies the basic concept of LSTM networks by minimising the number of gates and parameters, making it computationally light comparatively while maintaining the ability to learn long-range patterns. Two primary gating mechanisms are employed: update gate and reset gate, which regulate the information flow in the network and mitigate the vanishing gradient problem. Figure 13 illustrates the basic architecture of GRU.

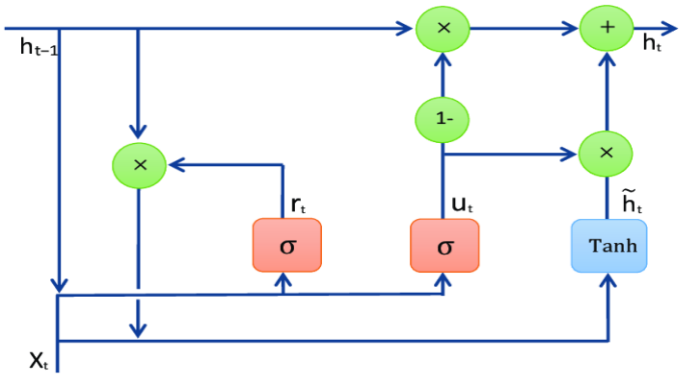

Fig.13. Basic Architecture of GRU

At each time step t, it receives an input vector x_t and the previous hidden state h_{t-1}. The update gate z_t determines how much of the previous memory to retain as per Equation (9).

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1}) \tag{9}$$

where W_z and U_z are the weight matrix for x_t and h_{t-1} respectively, σ represents the sigmoid activation function. The reset gate r_t regulates the impact of the previous hidden state when generating the candidate activation as in Equation (10).

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1}) \tag{10}$$

Using this reset gate, the model computes a candidate hidden state $\tilde{h}_t$ which represents the new memory content as in Equation (11).

$$\tilde{h}_t = \tanh\big(W \cdot x_t + U \cdot (r_t \odot h_{t-1})\big) \tag{11}$$

where $\odot$ denotes element-wise multiplication. The new hidden state h_t is updated by combining the previous hidden state h_{t-1} and the candidate $\tilde{h}_t$ based on the update gate z_t as in Equation (12).

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{12}$$

This formulation enables the GRU to adaptively remember or forget temporal patterns, rendering it suitable for tasks involving time series forecasting and sequential classification.

3.4.3 Proposed CvT-GRU hybrid Model

The proposed hybrid transformer-based DL model that integrates CvT for feature abstraction with GRU for temporal modelling, effectively capturing both spatial and sequential dependencies in ECG signals. The architecture begins with a patch embedding module, where a 1D convolutional layer processes the input time-series signal into overlapping local patches. This operation enables local feature extraction and sequence compression, reducing the temporal resolution while increasing the representational strength. The output is normalized using Layer Normalization, which stabilizes the learning process and accelerates convergence. The Transformer Encoder Block is then applied to the embedded sequence. This block includes multi-head self-attention to capture global dependencies, followed by a position-wise feedforward network for deep representation learning. Each sub-layer is enveloped in residual connections and normalization to preserve gradient flow and prevent degradation in deeper layers. Dropout is employed throughout the block to mitigate overfitting and improve generalization. The GRU layer then processes the output of the transformer block, summarizing sequential dependencies across time. GRU selectively retains or forgets past information through gating mechanisms. Together, this CvT–GRU hybrid architecture leverages convolution for local spatial pattern detection and transformers for global contextual modelling creating a robust, end-to-end pipeline for identifying SA from ECG signals.

The classification head serves as the final stage where the processed features are expressed as an output prediction. It begins with a fully connected dense layer comprising 64 units and a ReLU activation function, enabling the framework in learning complex non-linear relationships within

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the extracted feature set. To mitigate overfitting, L2 regularization is applied, which penalizes large weights. The regularization applied is as shown in Equation (13).

$$L_{reg} = \lambda \sum_{i=1}^{n} w_i^2 \tag{13}$$

where λ is the regularization strength and w_i, the weight parameters. A dropout layer is further added that randomly deactivates neurons during training to improve model generalization. A dense layer with a single unit and a sigmoid activation function forms the output layer that performs the binary classification. The sigmoid function is defined as in Equation (14).

$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}} \tag{14}$$

where z represents the linear output from the preceding layer and $\hat{y} \in [0,1]$ is the prediction probability of the positive class. Figure 14 represents the CvT-GRU hybrid model architecture. The algorithm of the hybrid model is as given below.

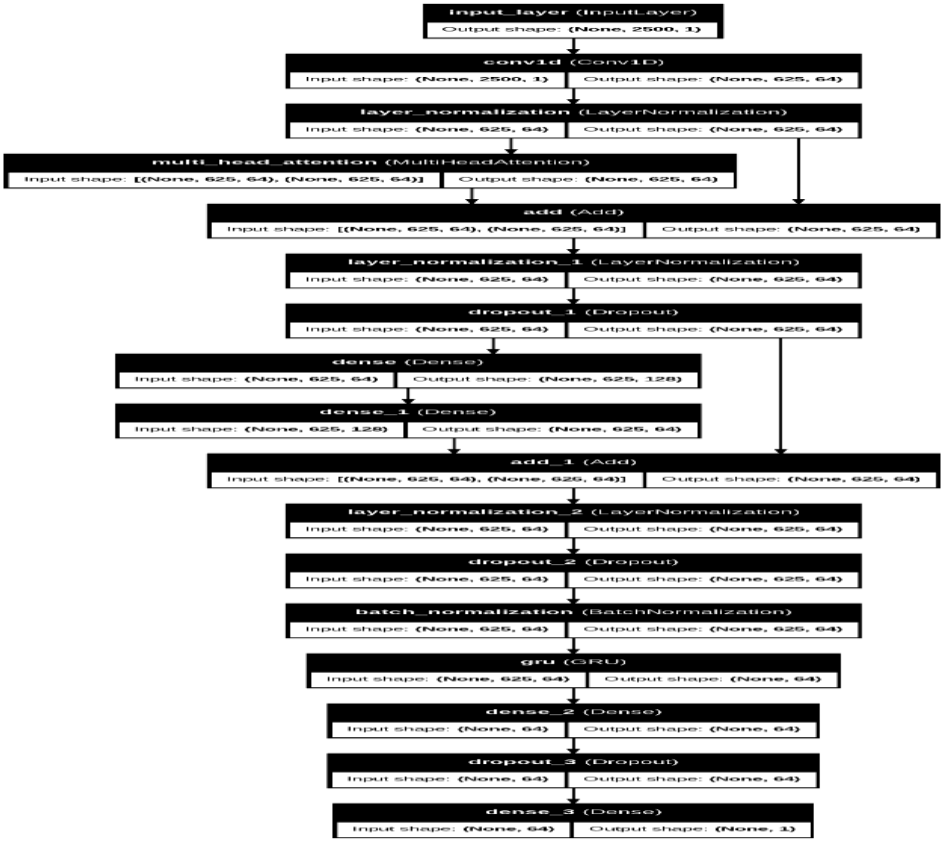

Fig.14. Model Architecture

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Algorithm: CvT-GRU Hybrid model for SA detection

Input:

- *X: ECG signal segments in time-series format*
- *Y: Label vector corresponding to X*
- *Class label* $\in \{0,1\}$*, for 0=normal and 1= Sleep apnea*

Output:

- *Predicted class label* $Y \in \{0,1\}$

Begin:

❖ ***Data collection***

- *Load Sleep Apnea Detection ECG dataset*
- *Extract ECG signal Matrix X and label vector Y*

❖ ***Pre-processing***

- *Normalization: Apply Min-max Normalization:*

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- *Data Split: Split into 70:15:15 (Training: Testing: Validation) ratio*

❖ ***CvT Feature Extraction***

- *Initial Convolutional Path Embedding:*

$$Z_0 = Conv\ 2D(X\_train)$$

- *Query, Key, Value generation using Depthwise Convolutions*

$$Q = DWConv_q(Z_0);\ \ K = DWConv_k(Z_0);\ \ \mathrm{V} = DWConv_v(Z_0)$$

- *Multi-Head Self Attention:*

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

$$MHSA(Q,K,V) = Concat(head_1, \dots, head_h)W^O$$

- *Feedforward Network:*

$$FFN(x) = W_2 \cdot GELU(W_1 \cdot x)$$

- *Layer Normalization and Residual Connection:*

$$z = LayerNorm(x + FFN(x))$$

❖ ***GRU-based Temporal Modelling***

- *Update Gate:*

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1})$$

- *Reset Gate:*

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1})$$

- *Candidate State:*

$$\tilde{h}_t = \tanh\big(W \cdot x_t + U \cdot (r_t \odot h_{t-1})\big)$$

- *Final Hidden state:*

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

❖ ***Classification Head***

- *Dense layer1:128 neurons, ReLU activation*
- *Dropout rate=0.1*

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- *Dense layer 2: 64 neurons, ReLU activation*

❖ ***Output layer***
- *1-unit Dense Layer, Sigmoid activation Function*

❖ ***Model Compilation and Training***
- *Compile model with loss = Binary crossentropy, learning rate = 0.0001, optimizer = ADAM, Epochs =50*
- *Train model: model.fit (X_train, y_train)*

❖ ***Evaluation and Model Saving***
- *Evaluate model: model. evaluate (X_test, Y_test)*
- *Tune hyperparameters*
- *Save the model*

End

3.5 Simulation setup

The proposed CvT-GRU hybrid model was implemented using a high-performance computational environment. The system configuration included an Intel Core i7 processor, an NVIDIA GeForce GTX 1080Ti GPU and 32 GB of RAM that collectively ensured efficient handling of the intensive training and evaluation processes involved in Sleep apnea detection. The Keras API, built on TensorFlow and python functioned as the programming language in the model development. The choice of framework was due to its exceptional features that provide flexibility, scalability and model customization capabilities. Google Colaboratory (Colab) was used for model training and testing, taking advantage of its free access to powerful GPUs and cloud-based execution environment that improved the study's accessibility and reproducibility. Contrary to the trainable model weights, these hyperparameters are manually selected prior to the training process and directly influence the model's rate of convergence, generalisation capability and ultimate classification performance. The hyperparameters and training settings employed in the study is summarized in Table 1.

Table.1. Hyperparameter Specifications

Hyper parameters	Values
Epochs	50
Batch Size	8
Activation function	ReLU, Softmax
Optimizer	ADAM

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Dropout	0.1
Number of Attention Heads	4
Loss function	Binary crossentropy
Learning Rate	0.0001

RESULTS AND DISCUSSIONS

A set of standard evaluation metrics has been employed for the performance analysis of the proposed framework as illustrated in Equation (15) to Equation (18). These measures are mathematically computed using the core elements of the confusion matrix: True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). Accuracy indicates the overall correctness, while recall and precision highlight the model's ability to detect and classify SA events without many misses or false alarms.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

$$Precision = \frac{TP}{TP+FP} \tag{16}$$

$$Recall = \frac{TP}{TP+FN} \tag{17}$$

$$F1 - score = 2 \times \frac{precision \times Recall}{Precision+Recall} \tag{18}$$

The accuracy plot offers a clear indication of how well the model is learning to make accurate predictions over epochs, both on training and validation sets. A steadily increasing validation accuracy indicates improved generalization. Conversely, the loss plot reflects how well the model minimizes the prediction error, helping to detect issues like underfitting or overfitting. A large gap between training and validation loss typically signals overfitting, whereas high loss on both suggests underfitting. Together, these plots guide model tuning and ensure stable, effective learning.

The accuracy plot of the CvT-GRU hybrid model illustrated in Figure 15 demonstrates rapid and stable convergence. A steep rise in training accuracy is observed within the initial epochs, after which it consistently plateaus around 98.7%, indicating effective learning without overfitting. The validation accuracy also stabilizes early, maintaining a consistent value of approximately 97.8% throughout the 50 epochs. The marginal separation between training and validation

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

accuracy curves indicates remarkable generalization ability, reflecting its robustness in accurately classifying unseen data.

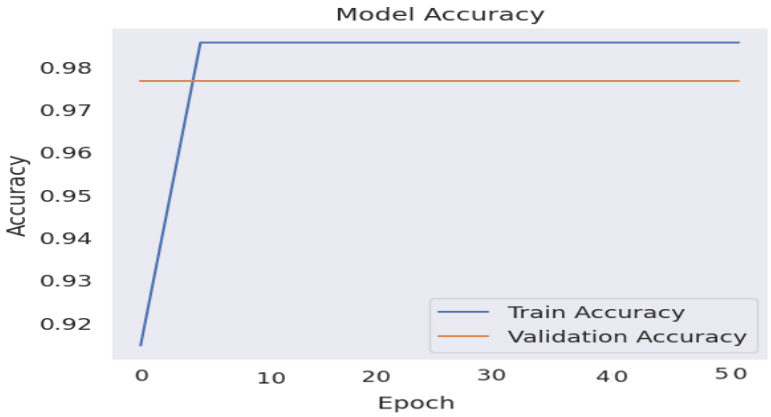

Fig.15. Accuracy Plot of the CvT-GRU Hybrid Model

Figure 16 illustrates the loss plot of the CvT-GRU hybrid model across 50 epochs. Initially, the training loss drops sharply, indicating rapid convergence in the early epochs. After this steep decline, the training loss stabilizes around 0.09 with minor fluctuations, suggesting consistent optimization without overfitting. In contrast, the validation loss begins around 0.125 and decreases slightly before plateauing near 0.13 for the majority of the training period. The relatively stable validation loss across epochs signifies that the model maintains generalization performance without significant variance or signs of underfitting. The narrow margin between training and validation loss further confirms the optimal balance between learning ability and generalization.

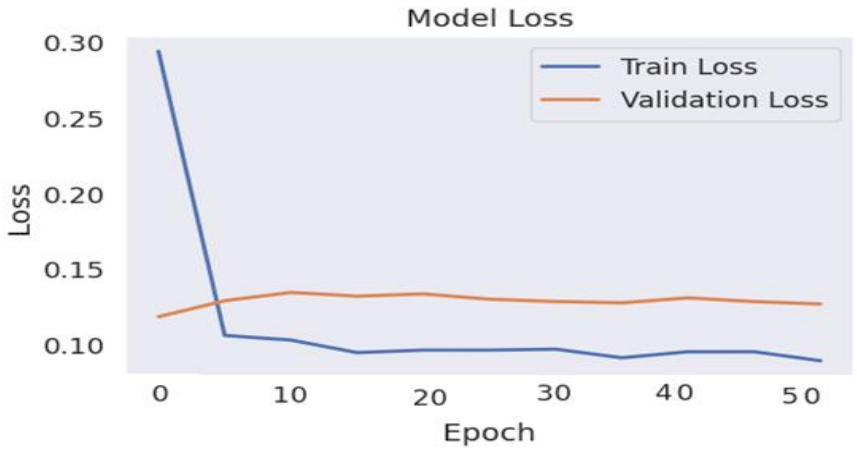

Fig.16. Loss Plot of the CvT-GRU Hybrid Model

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

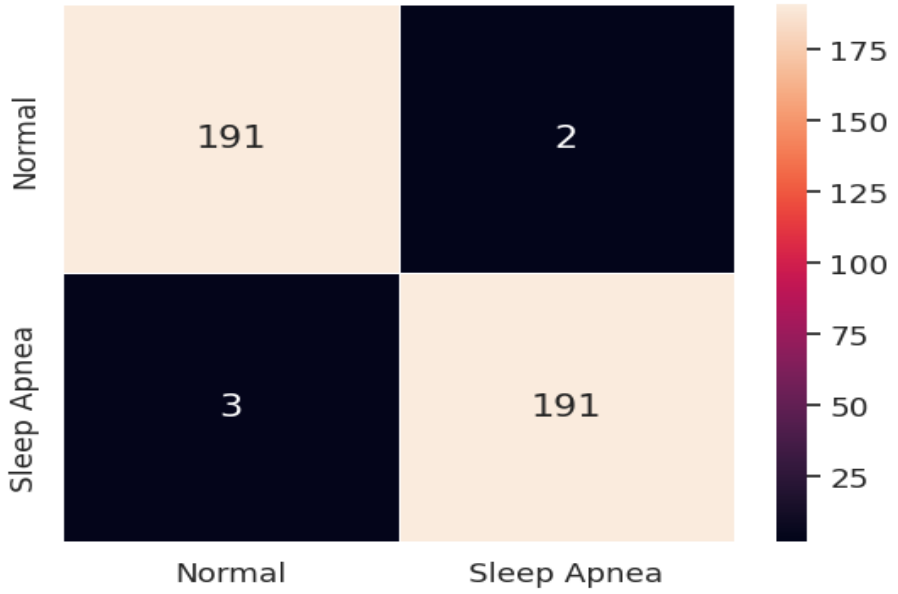

Fig.17. Confusion Matrix of the CvT-GRU Hybrid Model

Figure 17 represents the confusion matrix that demonstrates a clear visualization of the binary classification performance of the proposed model in distinguishing between Normal and Sleep Apnea cases. Out of 193 actual Normal instances, 191 were correctly classified, with only 2 misclassified as Sleep Apnea. Similarly, out of 194 Sleep Apnea instances, the model accurately identified 191, with just 3 misclassified as Normal. This minimal misclassification reflects high sensitivity and specificity. The balanced distribution of TPs and TNs, alongside the very low FP and FN counts, strongly supports the model's robustness and near-perfect classification capability.

	precision	recall	f1-score
Normal	0.98	0.99	0.99
Sleep Apnea	0.99	0.98	0.99
accuracy			0.99
macro avg	0.99	0.99	0.99
weighted avg	0.99	0.99	0.99

Fig.18. Classification Report

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The classification report illustrated in Figure 18 demonstrates remarkable performance across all evaluation metrics. For the Normal class, a precision of 0.98, a recall of 0.99 and an F1-score of 0.99 were achieved underscoring highly reliable detection of non-apnea cases. Similarly, for the Sleep Apnea class, the precision reached 0.99, recall was 0.98 and F1-score remained at 0.99, confirming the model's ability to accurately detect apnea events with minimal misclassification. The overall accuracy stands at 0.99, while both macro and weighted averages for precision, recall and F1-score are consistently 0.99, reflecting a balanced and robust classification performance across classes.

The ROC curve for the CvT-GRU hybrid model illustrated in Figure 19 demonstrates exceptional classification performance, with the curve following the top-left boundary of the graph. This results from a high TP rate and a very low FP rate across thresholds. The steep initial rise of the curve reflects the model's remarkable sensitivity, while the near-perfect alignment above the diagonal baseline of random chance confirms its high discriminative power. Such a pattern is indicative of an Area Under the Curve (AUC) value close to 1.0, supporting the reliability in distinguishing between normal and sleep apnea cases with excellent confidence.

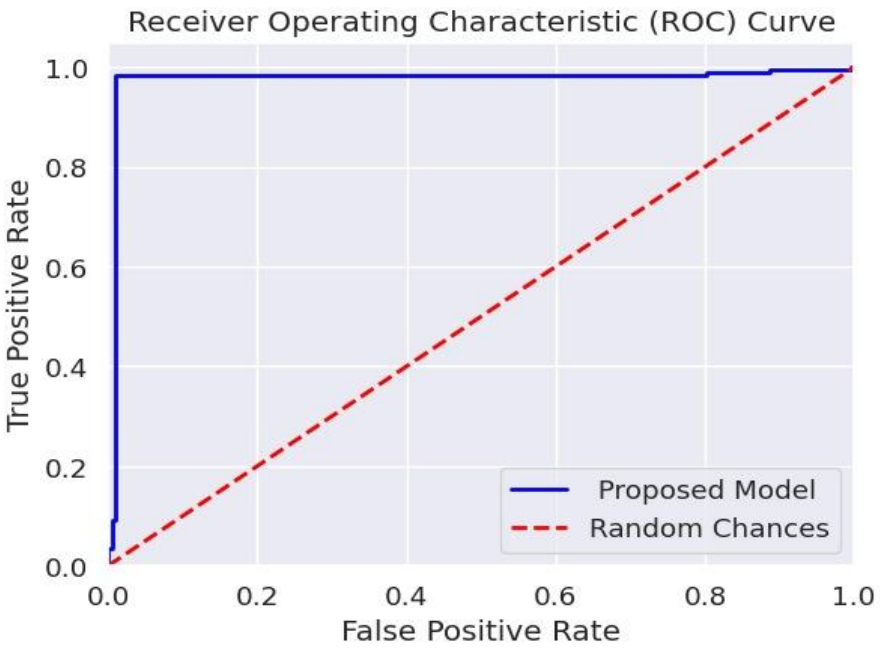

Fig.19. ROC Curve of the Proposed Model

To comprehensibly analyze the proposed model, a comparative study was done with the CvT-GRU model against the existing model for SA detection using various algorithms. Table 2 represents accuracy comparison of proposed model with existing methods.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table.2. Accuracy Comparison

Author [Ref]	Methodology Used	Accuracy (%)
Kolhar et al. [11]	Dual-branch CNN	94
Liu et al. [13]	EfficientNet + XGBoost	92.80
Jiménez-García et al. [14]	CNN + RNN + Grad-CAM	84
Bhongade & Gandhi [15]	WIVIDOSA-Net	90.09
Thompson et al. [16]	1DCNN-500	96.99
Arslan [17]	Hybrid DL+ML	95.74
Kang et al. [18]	MLP	73
Wang et al. [19]	OSAnet CNN	93.20
Lin et al. [20]	RAPIDEST	80
Molnár et al. [21]	Supervised ML	83
Urtnasan et al. [22]	5-layer CNN	82.20
Paul et al. [23]	ANN	91.83
Hemrajani et al. [24]	MobileNetV1 + GRU	90.29
Javeed et al. [25]	XGBoost + BiLSTM	97
Proposed Model	**CvT-GRU**	**98.70**

The comparative evaluation of sleep apnea detection illustrated in Figure 20 reveals a diverse landscape of DL and ML approaches. Among CNN-based models, the dual-branch CNN achieved a notable accuracy of 94%, although its susceptibility to overfitting and limited dataset diversity affected generalizability. EfficientNet combined with XGBoost performed well with 92.80% accuracy but required complex preprocessing steps and lacked interpretability. Hybrid architectures like CNN + RNN enhanced temporal analysis but showed moderate performance (84%) and overreliance on specific input signal patterns.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

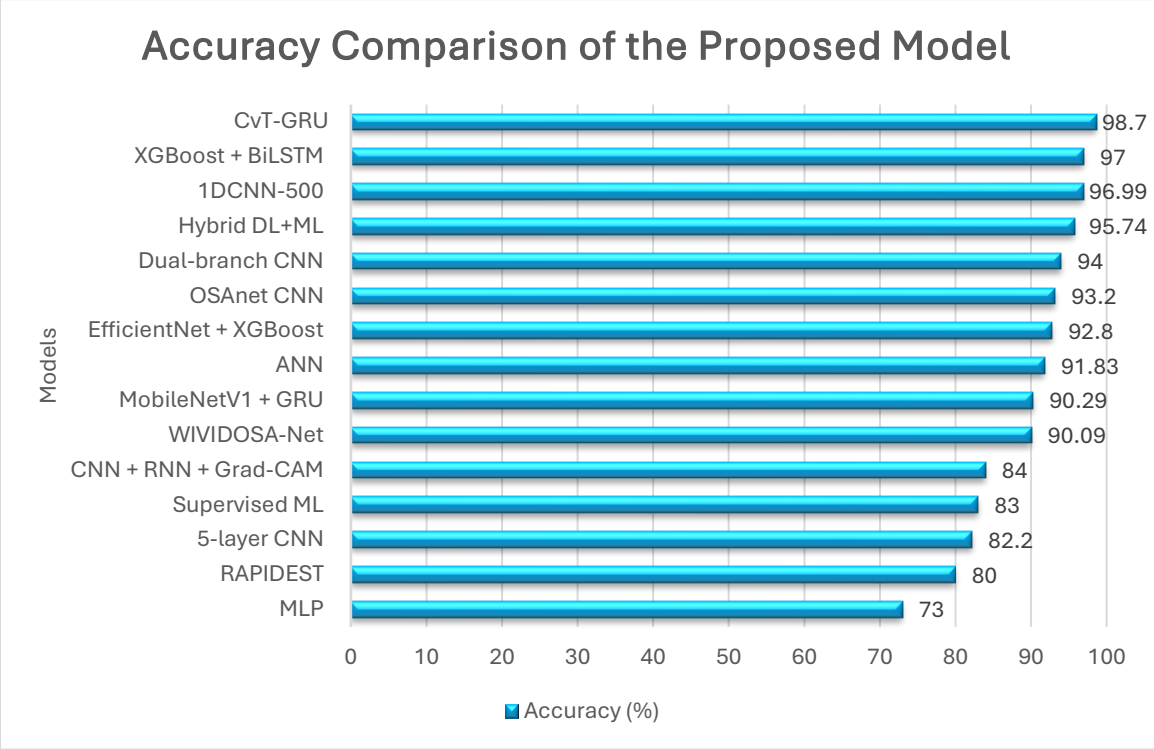

Fig.20. Accuracy comparison

Models such as WIVIDOSA-Net (90.09%) and MobileNetV1 with GRU (90.29%) utilized spectrogram or lightweight architectures for ECG analysis but often neglected raw temporal dynamics or were sensitive to signal noise. Other ML-based approaches like RAPIDEST and MLP models yielded lower accuracies, reflecting limitations in capturing nonlinear physiological variations or demographic stratification challenges. Although the 1DCNN-500 and hybrid XGBoost with BiLSTM models achieved high accuracies (96.99% and 97%), their applicability was limited by small datasets or lack of longitudinal validation. In contrast, the proposed CvT-GRU model demonstrates superior robustness and accuracy in detecting sleep apnea by effectively capturing both spatial and sequential patterns from ECG signals. Leveraging the hybrid strengths of convolutional vision transformers and gated recurrent units, the model achieves a higher classification accuracy of 98.70%, outperforming others. The consistent performance without excessive preprocessing or feature engineering highlightst the study's practical applicability and scalability in real-world clinical scenarios.

CONCLUSION

SA is a serious sleep disorder marked by repeated respiratory cessations during sleep, potentially giving rise to fatigue, cardiovascular issues and cognitive impairment. A hybrid DL model integrating Convolutional Vision Transformer (CvT) with GRU was proposed in the study for the accurate SA detection from ECG signals. The CvT component efficiently captures spatial

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

hierarchies and feature dependencies in the input signal representations, while the GRU layer models temporal dynamics and sequential patterns crucial for detecting apnea-related fluctuations. Trained and evaluated on the ECG Data for Sleep Apnea Detection, the framework achieved an outstanding classification accuracy of 98.7% with minimal overfitting. The classification metrics yielded an F1-score of 0.99 for both classes. The results affirm the effectiveness of the CvT-GRU hybrid model in distinguishing SA events with high precision compared to the existing methods. Future research may explore the integration of multi-channel physiological signals to enhance prediction robustness, the deployment of the model on edge devices for real-time monitoring and adaptation to pediatric or elderly populations with distinct apnea patterns.

References

[1] Andrisani, G., & Andrisani, G. (2023). Sleep apnea pathophysiology. Sleep and Breathing, 27(6), 2111-2122.

[2] Pase, M. P., Harrison, S., Misialek, J. R., Kline, C. E., Cavuoto, M., Baril, A. A., ... & Himali, J. J. (2023). Sleep architecture, obstructive sleep apnea, and cognitive function in adults. JAMA network open, 6(7), e2325152-e2325152.

[3] Regn, L. C. D. D., Davis, L. C. A. H., Smith, L. C. W. D., Blasser, M. C. J., & Ford, L. C. C. M. (2023). Central sleep apnea in adults: diagnosis and treatment. Federal Practitioner, 40(3), 78.

[4] Li, Y. E., & Ren, J. (2022). Association between obstructive sleep apnea and cardiovascular diseases: OSA and CVD risk. Acta Biochimica et Biophysica Sinica, 54(7), 882.

[5] Rosa, D., Amigoni, C., Rimoldi, E., Ripa, P., Ligorio, A., Fracchiolla, M., ... & Perger, E. (2022, May). Obstructive sleep apnea and adherence to continuous positive airway pressure (CPAP) treatment: let's talk about partners!. In Healthcare (Vol. 10, No. 5, p. 943). MDPI.

[6] Gambino, F., Zammuto, M. M., Virzì, A., Conti, G., & Bonsignore, M. R. (2022). Treatment options in obstructive sleep apnea. Internal and emergency medicine, 17(4), 971-978.

[7] Finnsson, E., Arnardóttir, E., Cheng, W. J., Alex, R. M., Sigmarsdóttir, Þ. B., Helgason, S., ... & Sands, S. A. (2023). Sleep apnea endotypes: from the physiological laboratory to scalable polysomnographic measures. Frontiers in Sleep, 2, 1188052.

[8] Teplitzky, T. B., Zauher, A. J., & Isaiah, A. (2023). Alternatives to polysomnography for the diagnosis of pediatric obstructive sleep apnea. Diagnostics, 13(11), 1956.

[9] Cho, T., Yan, E., & Chung, F. (2024). The STOP-Bang questionnaire: a narrative review on its utilization in different populations and settings. Sleep Medicine Reviews, 78, 102007.

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[10] Brennan, H. L., & Kirby, S. D. (2023). The role of artificial intelligence in the treatment of obstructive sleep apnea. Journal of Otolaryngology-Head & Neck Surgery, 52(1), s40463-023.

[11] Kolhar, M., Alfridan, M. M., & Siraj, R. A. (2025). AI-Driven Detection of Obstructive Sleep Apnea Using Dual-Branch CNN and Machine Learning Models. Biomedicines, 13(5), 1090.

[12] Huang, J., Zhuang, J., Zheng, H., Yao, L., Chen, Q., Wang, J., & Fan, C. (2024). A machine learning prediction model of adult obstructive sleep apnea based on systematically evaluated common clinical biochemical indicators. Nature and Science of Sleep, 413-428.

[13] Liu, M. H., Chien, S. Y., Wu, Y. L., Sun, T. H., Huang, C. S., Hsu, K. C., & Hang, L. W. (2024). EfficientNet-based machine learning architecture for sleep apnea identification in clinical single-lead ECG signal data sets. BioMedical Engineering OnLine, 23(1), 57.

[14] Jiménez-García, J., García, M., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Vaquerizo-Villar, F., Álvarez, D., ... & Hornero, R. (2024). An explainable deep-learning architecture for pediatric sleep apnea identification from overnight airflow and oximetry signals. Biomedical Signal Processing and Control, 87, 105490.

[15] Bhongade, A., & Gandhi, T. K. (2025). WIVIDOSA-Net: Wigner–Ville distribution based obstructive sleep apnea detection using single lead ECG signal. Biomedical Engineering Advances, 9, 100159.

[16] Thompson, S., Reilly, D., Fergus, P., & Chalmers, C. (2023). Detection of obstructive sleep apnoea using features extracted from segmented time-series ECG signals with a one dimensional convolutional neural network. IEEE Access, 12, 1076-1091.

[17] Arslan, R. S. (2023). Sleep disorder and apnea events detection framework with high performance using two-tier learning model design. PeerJ Computer Science, 9, e1554.

[18] Kang, C., An, S., Kim, H. J., Devi, M., Cho, A., Hwang, S., & Lee, H. W. (2023). Age-integrated artificial intelligence framework for sleep stage classification and obstructive sleep apnea screening. Frontiers in Neuroscience, 17, 1059186.

[19] Wang, B., Tang, X., Ai, H., Li, Y., Xu, W., Wang, X., & Han, D. (2022). Obstructive sleep apnea detection based on sleep sounds via deep learning. Nature and Science of Sleep, 2033-2045.

[20] Lin, X. X., Lin, P., Yeh, E. H., Liu, G. R., Lien, W. C., & Fang, Y. (2022). RAPIDEST: A framework for obstructive sleep apnea detection. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 387-397.

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[21] Molnár, V., Kunos, L., Tamás, L., & Lakner, Z. (2023). Evaluation of the applicability of artificial intelligence for the prediction of obstructive sleep apnoea. Applied Sciences, 13(7), 4231.

[22] Urtnasan, E., Kim, Y., Yang, J. W., Kim, S. H., Koh, S. B., & Hwang, S. (2023). AI-based Prediction Model for Incident of Obstructive Sleep Apnea Using ECG Signals: Utilization of MrOS. Digital Health Research, 1(1).

[23] Paul, T., Hassan, O., Alaboud, K., Islam, H., Rana, M. K. Z., Islam, S. K., & Mosa, A. S. (2022). ECG and SpO2 signal-based real-time sleep apnea detection using feed-forward artificial neural network. AMIA Summits on Translational Science Proceedings, 2022, 379.

[24] Hemrajani, P., Dhaka, V. S., Rani, G., Shukla, P., & Bavirisetti, D. P. (2023). Efficient deep learning based hybrid model to detect obstructive sleep apnea. Sensors, 23(10), 4692.

[25] Javeed, A., Berglund, J. S., Dallora, A. L., Saleem, M. A., & Anderberg, P. (2023). Predictive power of XGBoost_BiLSTM model: a machine-learning approach for accurate sleep apnea detection using electronic health data. International Journal of Computational Intelligence Systems, 16(1), 188.

[26] https://www.kaggle.com/datasets/ucimachinelearning/ecg-data-for-sleep-apnea-detection

[27] Yu, S., Xie, L., & Huang, Q. (2023). Inception convolutional vision transformers for plant disease identification. Internet of Things, 21, 100650.

[28] Manlises, C. O., Chen, J. W., & Huang, C. C. (2024). A gated recurrent unit model based on ultrasound images of dynamic tongue movement for determining the severity of obstructive sleep apnea. Ultrasonics, 141, 107320.

Received: August 07, 2025