**Sleep Apnea Detection using ECG Derived Respiration: A Comparative Study of Deep Learning Architectures**

# Abstract

Millions of people around the world suffer from sleep apnea, a serious sleep disorder marked by frequent breathing pauses while they sleep. It can lead to serious health problems like cardiovascular disease and cognitive decline. The PhysioNet Apnea-ECG Database's ECG-derived respiration signals are used in this study's thorough comparison of three deep learning architectures for automated sleep apnea detection. CNN LSTM, CNN Transformer-LSTM, and CNN BiLSTM Transformer hybrid models were implemented and assessed; the corresponding accuracies were 90.55%, 89.70%, and 90.67%. Our method integrates conventional heart rate variability features with sophisticated signal processing methods, such as Wavelet-EMD decomposition with Hjorth parameters. With a sensitivity of 96.30%, specificity of 88.0%, and an AUC of 0.9724, the CNN BiLSTM Transformer model showed excellent performance and could be used in clinical settings to diagnose sleep disorders.

1. **Keywords:** Sleep Apnea, ECG, Deep Learning, CNN, LSTM, Transformer, Signal Processing

# Introduction

About 936 million adults worldwide suffer from sleep apnea, of which obstructive sleep apnea (OSA) is the most common type. Recurrent episodes of full or partial upper airway obstruction during sleep are the condition's hallmark, causing disturbed sleep patterns and low oxygen saturation. The polysomnography (PSG) used in traditional diagnosis is costly, time-consuming, and necessitates overnight monitoring in specialized sleep laboratories.

ECG-derived respiration (EDR) is a promising new way to find sleep apnea that uses the effects of breathing events on the heart. When someone is having an apneic episode, the lack of respiratory modulation in ECG signals, along with responses from the autonomic nervous system, creates patterns that can be seen and studied using advanced signal processing and machine learning methods.Recent progress in deep learning has shown that it works very well for biomedical signal analysis, especially for automated feature extraction and pattern recognition. Using ECG signals from the PhysioNet Apnea-ECG Database, this study looks at how well three hybrid deep learning architectures can find sleep apnea.

## Research Contributions

The main things we have to offer are:   
A full comparison of the CNN LSTM, CNN Transformer-LSTM, and CNN BiLSTM Transformer models in terms of their hybrid architecture   
Advanced Feature Engineering: Combining Wavelet-EMD decomposition with Hjorth parameters and features that show how heart rate changes over time   
Importance in the clinic: Getting a high sensitivity level of 96.30% is very important for medical screening uses.   
Reproducible Framework: An open-source implementation with clear steps for preprocessing and evaluation

# Related Work

## Traditional Approaches

Early methods for finding sleep apnea used hand-crafted features taken from ECG signals, such as heart rate variability (HRV) parameters, frequency domain features, and morphological characteristics. These approaches achieved moderate success but were limited by the complexity of manual feature engineering and the inability to capture complex temporal patterns.

## Deep Learning in Sleep Analysis

In recent years, using deep learning to find sleep disorders has become very popular. Convolutional Neural Networks (CNNs) are good at automatically extracting features from raw signals, and Recurrent Neural Networks (RNNs) are good at modeling how data changes over time in a sequence. The addition of attention mechanisms to Transformer architectures has made it even easier to find long-range dependencies in physiological signals.

.

## ECG Based Sleep Apnea Detection

Recent research has shown that it is possible to use different machine learning methods to find sleep apnea using ECG data. But most of the work that has been done so far focuses on single-architecture solutions and doesn't fully compare hybrid models that combine the best parts of different deep learning paradigms.

.

# Methodology

## Dataset

We used the PhysioNet Apnea-ECG Database, which has 70 overnight ECG recordings from people who might have sleep apnea.

The data set has:   
Taking notes Time: 8 to 10 hours for each subject   
100 Hz is the sampling rate.   
Notes: Labels for apnea and normal every minute   
Total Segments: 5,289 segments that are one minute long 47.9% apnea, 52.1% normal

## Signal Preprocessing

There are several steps in our preprocessing pipeline that are meant to improve signal quality and find useful features:

### Advanced ECG Preprocessing

The preprocessing class does:   
Wavelet-EMD Decomposition: Daubechies-4 wavelet decomposition comes first, and then Empirical Mode Breaking down   
Hjorth Parameters: Getting Activity, Mobility, and Complexity features out   
Heart Rate Variability: Traditional time-domain HRV features like SDNN, RMSSD, and pNN50

### Feature Extraction

Our process for extracting features includes:   
Wavelet-EMD 45 dimensions of features: Using 4-level wavelet decomposition and then EMD on each sub-band HRV for multi-resolution analysis

10 dimensions of features: Parameters for heart rate variability in the time domain

Total Feature Vector: a 55-dimensional feature representation for each segment

## Deep Learning Architectures

## CNN LSTM Model

## The CNN LSTM architecture takes in two inputs:

## Feature Path: Engineered features go through dense layers Signal Path: CNN layers process the raw ECG, and then LSTM layers model the time.Details about the architecture: CNN 3 convolutional layers with 32, 64, and 128 filters and 2 LSTM layers with 100 and 50 units Dense: 4 layers of classification with dropout regularization

## Parameters: 195,273 762.79 KB

### CNN Transformer-LSTM Model

This architecture includes attention mechanisms:   
Mechanism of Attention: Dual-layer attention to improve feature focus   
Parts of a Transformer: Normalizing layers and adding residual connections   
Adding LSTM: Modeling time after processing attention

**Details about the architecture:**

Three CNN convolutional layers with batch normalization   
Attention: Two layers of attention and two layers of LSTM with dropout   
Parameters: 228,809 893.79 KB

### CNN BiLSTM Transformer Model

The most advanced architecture combines:

Bidirectional LSTM: Better modeling of time in both directions   
Multi-Head Attention: A transformer block with five attention heads   
Cross-Modal Fusion: Combining feature and signal representations based on attention

**Details about the architecture:**

CNN 3 convolutional layers with batch normalization   
BiLSTM has three bidirectional layers with 100, 50, and 25 units each.

The transformer has a better transformer block with multi-head attention.   
parameters: 524,257 2.00 MB

## Training Configuration

We trained all of the models with:   
Optimizer: Adam with a learning rate of 0.001   
Loss Function: Cross-entropy in binary   
Batch Size: 16 to 32 (depends on the architecture)   
Epochs: 30 to 100 with early stopping   
20% of the data is for validation.   
Class Weights: Making the weights even for an uneven dataset

## Evaluation Metrics

We used a wide range of evaluation metrics that are useful in medical settings:

Correctness: Overall performance of the classification   
Sensitivity Remember: the ability to find apnea episodes   
Specificity: Being able to correctly identify normal segments   
AUC stands for "area under the ROC curve."   
Accuracy: the value of a positive prediction

# Results

## Performance Comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Accuracy | AUC | Sensitivity | Specificity | Precision |
| CNN LSTM | 90.55% | 0.9724 | 94.00% | 88.00% | 87.00% |
| CNN Transformer-LSTM | 89.70% | 0.9700 | 94.48% | 85.30% | 86.00% |
| CNN BiLSTM Transformer | 90.67% | 0.9724 | 96.30% | 88.00% | 88.00% |

## Detailed Analysis

### CNN LSTM Performance

The CNN LSTM model was 90.55% accurate and performed the same on all metrics. The model showed:   
Balanced Performance: A good balance between sensitivity and specificity   
Effective Training: Convergence in 30 epochs   
Computational Efficiency: The model with the fewest parameters that was tested

### CNN Transformer-LSTM Performance

With the CNN Transformer-LSTM model, the accuracy was 89.70%.   
High Sensitivity: 94.48% of apneas were found   
Benefits of Attention: Better focus on features thanks to attention mechanisms Training Stability: Stable training with an early stop at epoch 32

### CNN BiLSTM Transformer Performance

The CNN BiLSTM Transformer model had the best overall performance, with an accuracy rate of 90.67%.   
Better Sensitivity: 96.30% of apneas were found   
Strong Architecture: Processing in both directions and paying attention to different modes Clinical Suitability: Medical screening needs the highest sensitivity possible.

## Training Dynamics

All of the models trained steadily and with good regularization:   
Early Stopping: Used patience mechanisms to stop overfitting.   
Learning Rate Scheduling: Reducing the learning rate in an adaptive way Class Balancing: How to handle a dataset that isn't balanced well

## Feature Importance Analysis

The hybrid method of combining engineered features with raw signal processing worked well:   
Wavelet-EMD Features: Captured signal characteristics at multiple resolutions   
HRV Features: Provided temporal patterns that were physiologically meaningful   
Processing Raw Signals: Allowed CNN layers to automatically find features

# Discussion

# 7.1 Clinical Implications

The performance metrics that were reached show that ECG-based sleep apnea detection is clinically useful:   
Very Sensitive: The CNN BiLSTM Transformer model has a sensitivity of 96.30%, which makes it good for screening applications where missing apnea cases would be clinically important.   
Balanced Performance: The models have a high sensitivity and a good specificity (85% to 88%).   
Screening that doesn't cost too much: Detection based on ECG is a more accessible option than polysomnography.

## 7.2 Architectural Insights

The comparison shows that there are important architectural factors to think about:

### Benefits of Bidirectional Processing

### The CNN BiLSTM Transformer's better performance shows how important it is to model physiological signals in both directions over time. The ability to process sequences in both directions captured more complete temporal dependencies.

### Effectiveness of Attention Mechanism

Adding attention mechanisms to both CNN Transformer-LSTM and CNN BiLSTM Transformer models made them better at focusing on important signal parts. The benefits were strongest when the models were used with bidirectional processing.

**Strategy for Feature Fusion**

The dual-input architecture worked well with both engineered features and raw signal processing, and cross-modal attention did a great job of combining information from different sources.

## 7.3 Limitations and Future Work

### Limitations of dataset

The PhysioNet database is for a certain group of people.   
Note Size: Minute-by-minute labels might not show events that last only a few minutes. Missing Severity Details: Binary classification doesn't show how bad the apnea is.

### Computational Consideration

Complexity of the Model: CNN BiLSTM Transformer needs a lot of computing power.Processing in real time: Problems with putting in place continuous monitoring systems.Deployment at the Edge: Model compression is necessary for mobile health apps

### Future Direction

### What Comes Next Multi-Modal Integration: Putting ECG together with other body signals Models made just for you: Adapting to specific subjects to improve accuracy Explainable AI: Making models that can be understood to help with clinical decision-making Continuous Learning: systems that get better with more data

# Conclusion

This study compares three deep learning architectures in detail for detecting sleep apnea based on ECG. The CNN BiLSTM Transformer model had the best results, with 90.67% accuracy, 96.30% sensitivity, and 0.9724 AUC. This shows that combining bidirectional temporal modeling with attention mechanisms works well.

The main points are:

Hybrid Architectures: Using CNN, LSTM, and Transformer parts together works better than using just one architecture.

Feature Engineering: Combining Wavelet-EMD decomposition with traditional HRV features makes the model work better.

Clinical Viability: The performance metrics that were met show that the technology could be used in clinical settings to screen for sleep disorders.

Processing in both directions: BiLSTM layers make a big difference in how well temporal patterns can be recognized.

The results show that using advanced deep learning architectures to detect sleep apnea using ECG can achieve clinically relevant performance. This is a promising alternative to traditional polysomnography for screening and monitoring sleep disorders.

# References

Benjafield, A. V., et al. 2019. A study of the literature to find out how common and serious obstructive sleep apnea is around the world. The Lancet Respiratory Medicine, 7 8, 687 698.   
Goldberger, A. L., et al. 2000. PhysioBank, PhysioToolkit, and PhysioNet are parts of a new research tool for studying complex physiological signals. Circulation, 101 23, e215-e220.

Penzel, T., and others. 2002. The database for apnea-ECG. Computers in Cardiology, 29, 255–258.   
Vaswani, A., et al. 2017. You just need attention. 30, 5998–6008 of Advances in Neural Information Processing Systems.   
Hochreiter, S., and Schmidhuber, J. 1997. Long-term memory. Neural Computation, 9 8, 1735–1780.

⁂