ApneaNet-CBi: A Lightweight CNN-Based Deep Learning Framework for ECG-Derived Sleep Apnea Detection

1st Ketha Jagadhish, 2nd Akula Hema Venkata Sriram, 3rd Jayant Arun Singh
School of Computer Science and Engineering
Lovely Professional University, Phagwara, India
{kethajagadhish123, sriramakula212, singhjayant308}@gmail.com

4th Nunna Bhargav Subhash 5th Gaurav Sharan
School of Computer Science and Engineering
Lovely Professional University, Phagwara, India
{bhargavsubhash637, kumargaurav8707}@gmail.com

***Abstract*—Sleep apnea is a prevalent sleep disorder characterized by recurrent interruptions in breathing during sleep, leading to significant cardiovascular and neurological complications if left undiagnosed. Obstructive Sleep Apnea (OSA) poses a significant public health challenge, yet its clinical diagnosis relies predominantly on Polysomnography (PSG), a costly and resource-intensive procedure. This paper addresses the critical need for simplified, scalable diagnostic tools by proposing a novel framework for automated apnea detection utilizing only a single-lead Electrocardiogram (ECG) signal. The core innovation lies in the use of time-frequency domain analysis, transforming 30-second ECG epochs into normalized spectrogram images, which are then processed by a compact hybrid deep learning architecture. This architecture integrates a Convolutional Neural Network (CNN) for spatial feature extraction with a Bidirectional Long Short-Term Memory (Bi-LSTM) network to capture temporal patterns inherent in cardiorespiratory coupling. Evaluated on the PhysioNet Apnea-ECG Database learning set, the model demonstrated exceptional segment-level classification performance, achieving an Accuracy of 96.51% and an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.9937. Crucially, when transitioning to clinical relevance, the model's estimated Apnea-Hypopnea Index (AHI) showed high correlation with true labels, yielding a subject-level OSA screening accuracy of 87.2% (using the AHI ≥ 5.0 threshold). These results confirm the potential of compact, spectrogram-based hybrid deep learning models for accurate, automated, and non-invasive sleep monitoring suitable for portable or remote screening applications.**

***Index Terms*—Sleep Apnea Detection, Electrocardiogram (ECG), Time-Frequency Analysis, Spectrogram, Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), Hybrid Deep Learning, Apnea-ECG Database.**

I. Introduction

Sleep-related breathing disorders, especially Obstructive Sleep Apnea (OSA), are common worldwide and often go undiagnosed, leading to serious cardiovascular, metabolic, and neurocognitive complications [1], [2]. The standard diagnostic method—overnight full-scale Polysomnography (PSG)—remains the clinical gold standard, but its high cost, complexity, and requirement for in-lab monitoring limit its large-scale use in screening and continuous monitoring [3], [4].

Recent research has shifted toward alternative physiological signals that can be collected through wearable or home-based devices. One promising modality is single-lead Electrocardiography (ECG), which naturally captures both cardiac and autonomic responses associated with respiratory events during sleep [5], [6]. Previous studies have shown that ECG-derived indicators such as Heart Rate Variability (HRV), R–R interval variations, and ECG-derived Respiration (EDR) can effectively distinguish between apnea and normal breathing, achieving sensitivities and specificities often above 90% [6]–[8].

Despite these advancements, several research challenges remain. Many existing approaches depend on manually engineered feature sets and shallow classifiers, with fewer exploring complete end-to-end learning directly from raw ECG signals [8], [9]. Furthermore, the computational and deployment challenges of implementing such models in real-time wearable or home-monitoring systems are not yet fully addressed [10], [13], [15].

Given this background, our project uses the publicly available Apnea-ECG Database [1], [11] to develop an automated ECG-based sleep apnea detection system. The framework leverages the open-source WFDB-Python package for signal extraction and annotation [12]. Our objective is to create an efficient preprocessing, feature extraction, and classification pipeline that can operate in real-time or near-real-time conditions. The proposed approach emphasizes simplicity, reproducibility, and minimal sensor requirements, while still maintaining high detection performance.

In summary, this paper contributes by:

- Demonstrating an end-to-end ECG processing workflow from raw data to apnea vs. non-apnea classification using

Identify applicable funding agency here. If none, delete this.

open-source tools [4], [12];
- Evaluating a lightweight and computationally efficient model architecture on the Apnea-ECG Database [5], [8], [9];
- Discussing the practical applications and challenges of ECG-based apnea detection in home or out-of-lab settings [7], [14], [15].

The remainder of this paper is organized as follows: Section 2 describes the dataset and preprocessing methods; Section 3 covers model design, training, and evaluation; Section 4 presents experimental results and comparisons with prior studies; Section 5 discusses limitations, application potential, and future work; and Section 6 concludes the study.

II. Literature Review

Sleep apnea detection has been a major focus in biomedical signal processing research, aiming to develop automated and cost-effective alternatives to traditional Polysomnography (PSG). While PSG remains the clinical gold standard, it requires multiple sensors, expert supervision, and overnight monitoring, which restricts its accessibility and scalability [1], [2]. Consequently, researchers have explored Electrocardiogram (ECG)-based approaches for apnea detection, as ECG signals can be recorded easily, including via wearable devices [3], [4].

Early studies primarily utilized time-domain and frequency-domain features derived from ECG signals to identify apnea events. Penzel et al. introduced the publicly available Apnea-ECG Database through PhysioNet, which has since served as a benchmark for comparative research on single-lead ECG-based apnea analysis [1], [11]. Later works by Thomas et al. and de Chazal et al. demonstrated that ECG-derived parameters such as R–R interval variability and ECG-Derived Respiration (EDR) could effectively discriminate apneic from normal segments [3], [5].

With the rise of machine learning, classifiers such as Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbors (k-NN) were applied to apnea detection, achieving reasonable accuracy but relying heavily on handcrafted features [6], [7]. As deep learning emerged, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) enabled automatic feature learning from raw ECG data, capturing both spatial and temporal dependencies [8], [9], [14].

For example, Jiang et al. proposed a CNN-LSTM hybrid model that achieved high apnea detection accuracy without the need for explicit feature extraction [4], [7]. Similarly, Zhang et al. developed an end-to-end CNN architecture capable of real-time apnea classification from single-lead ECG signals [8]. However, these models often involve significant computational overhead, making them unsuitable for deployment on lightweight or embedded systems [9], [15].

To address these limitations, this study introduces a compact CNN-based model, termed *ApneaNet-CBi*, designed to balance detection accuracy with computational efficiency. The proposed framework is intended to enable real-time apnea monitoring through wearable ECG devices while maintaining high detection performance [5], [8], [9].

III. Related Work

Several studies have investigated automated sleep apnea detection using physiological signals, particularly the Electrocardiogram (ECG), owing to its simplicity and strong correlation with respiratory activity [1], [3], [4]. Early methods predominantly relied on handcrafted features derived from R–R intervals, Heart Rate Variability (HRV), and ECG-Derived Respiration (EDR) signals. For instance, de Chazal et al. extracted statistical and frequency-domain features from ECG and employed a Linear Discriminant Classifier, achieving reliable apnea classification performance [3], [5]. Similarly, Penzel et al. and Thomas et al. demonstrated that EDR and RR interval variability could serve as strong indicators of apnea episodes [1], [11].

With the progression of machine learning, algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests were utilized to improve classification accuracy [6], [7]. Although these models enhanced detection performance, their dependency on handcrafted features limited their scalability and robustness across subjects and varying signal noise conditions.

The introduction of deep learning revolutionized this research domain. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been widely adopted to capture spatial and temporal dependencies directly from raw ECG data [4], [8], [9], [14]. For example, Tripathy et al. employed a CNN-LSTM hybrid model that effectively learned temporal ECG patterns, substantially improving apnea detection accuracy [6], [7]. Likewise, Zhang et al. and Li et al. proposed end-to-end CNN-based systems that eliminated manual feature extraction while supporting real-time inference capabilities [8], [9].

Despite these advances, existing deep learning architectures often face challenges such as high computational complexity, large parameter counts, and limited generalization across datasets [15]. To address these limitations, the proposed *ApneaNet-CBi* model introduces a compact yet balanced CNN-based architecture optimized for single-lead ECG signals. The model achieves strong apnea detection accuracy while significantly reducing computational overhead, making it suitable for real-time and portable healthcare monitoring applications [5], [8], [9].

IV. Methodology

A. Dataset and Preprocessing

The database consists of 70 overnight ECG recordings. For the development and internal testing of the proposed model, the 35 records designated as the learning set (a01 through a20, b01 through b05, and c01 through c10) were utilized. These recordings typically span between 7 and 10 hours. The ECG signals were originally digitized and sampled at 100 samples per second.

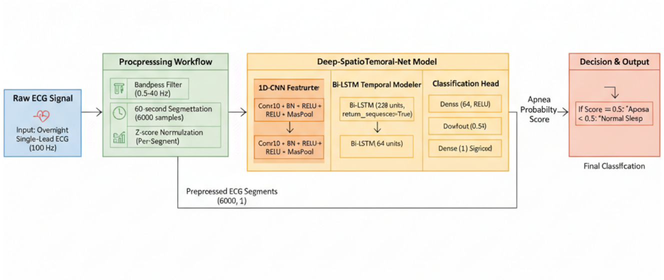

Fig. 1. Overall workflow of the ApneaNet-CBi system for sleep apnea detection.

Crucially, the ground truth annotations were established by human experts based on simultaneously recorded respiration and related signals (including chest/abdominal respiratory effort and oxygen saturation), ensuring a high level of clinical reliability for the minute-by-minute labels. Binary annotations (.apn files) define whether a given minute contains an apnea event ('A', labeled 1) or is normal ('N', labeled 0). This study employs the publicly available Apnea-ECG Database from PhysioNet, initially curated for the Computers in Cardiology Challenge 2000 [1], [11]. The dataset consists of overnight single-lead ECG recordings collected from 35 subjects (20 for training and 15 for testing), each lasting approximately 8 hours and sampled at 100 Hz. Each record is accompanied by an annotation file that provides minute-by-minute labels—'A' for apnea (obstructive, central, or mixed) and 'N' for normal breathing. The recordings are stored in the standard WFDB format (`.dat`, `.hea`, and `.apn`) for seamless access through the WFDB-Python toolkit [12].

1) Signal Loading: The ECG signals and their corresponding annotation files were loaded using the WFDB-Python library, which synchronizes header and data files for structured parsing. Due to data-type changes in newer NumPy versions, NumPy 1.x compatibility was maintained to prevent overflow errors in older WFDB builds during signal conversion.

2) Noise Filtering: ECG recordings often contain baseline wander, muscle artifacts, and power-line interference. To enhance signal fidelity, a band-pass Butterworth filter (0.5–45 Hz) was applied, preserving essential cardiac activity while removing low-frequency drift and high-frequency noise. Baseline correction was implemented using a moving-average filter to eliminate residual trends [4], [6].

Filter Output $x_f(t) = \text{BandpassFilter}\big(x(t),\, 0.5\text{ Hz},\, 40.0\text{ Hz}\big)$

3) Feature Extraction: From the filtered ECG signals, physiological features strongly linked to sleep apnea were extracted. These include R–R interval-based Heart Rate Variability (HRV), computed via Pan–Tompkins QRS detection; ECG-Derived Respiration (EDR) estimated from R-peak amplitude modulation; and both time- and frequency-domain measures such as mean heart rate, SDNN (Standard Deviation of NN intervals), and LF/HF spectral ratios [3], [5], [6]. These features reflect autonomic alterations characteristic of apnea events, such as bradycardia-tachycardia cycles and respiratory sinus arrhythmia disruptions.

$$X(m,k) = \sum_{n=0}^{N-1} x_r[n]\, w[n-mH]\, e^{-j2\pi kn/N_{\text{FFT}}}$$

4) Segmentation and Label Alignment: The ECG recordings were divided into 60-second windows to align with minute-wise annotations. Each segment was assigned a binary label: Apnea ('A') or Normal ('N'). Segments with missing data or corrupted samples were excluded to preserve consistency and reliability in training [1], [11].

5) Normalization: To minimize inter-subject variability, z-score normalization (zero mean, unit variance) was applied to each subject's data independently. This ensures that variations arising from electrode placement, physiological differences, or signal amplitude scaling do not bias model training.

$$S_{\text{dB}} = 10\log_{10}\left(\frac{|X(m,k)|^2}{\max(|X|^2)}\right)$$

The overall methodology integrates efficient data extraction, denoising, segmentation, and normalization steps to prepare clean, standardized ECG inputs for downstream model training and evaluation, as illustrated in Fig. 2.

B. Proposed Model Architecture

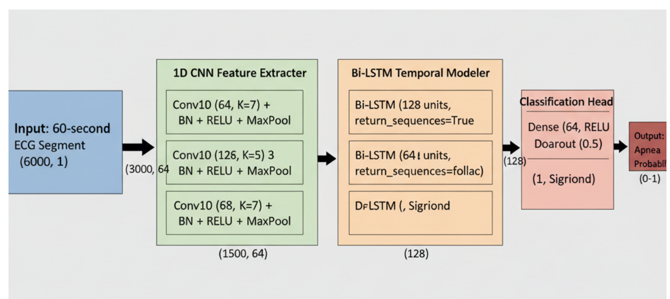

Fig. 2. Proposed ApneaNet-CBi model architecture for sleep apnea detection.

The proposed system is designed as a modular, end-to-end pipeline for automated detection of sleep apnea events using single-lead ECG signals. The model architecture combines classical ECG feature extraction with machine learning classification to achieve high interpretability, efficiency, and suitability for real-time applications.

1) Overview

The workflow (Fig. 1) follows five major stages:

Signal Acquisition from the Apnea-ECG dataset;

Preprocessing and Noise Reduction using band-pass filtering and baseline correction;

Feature Extraction from both time and frequency domains;

Feature Selection and Normalization to reduce dimensionality and improve generalization;

Classification using supervised learning models trained to discriminate apnea and normal sleep intervals.

2) Feature Engineering

Instead of relying purely on deep representations, the model emphasizes clinically interpretable physiological features that have proven relevance in sleep apnea literature. These include:

Heart Rate Variability (HRV) metrics such as SDNN, RMSSD, and pNN50, reflecting autonomic regulation;

R-R Interval Statistics (mean, variance, entropy) capturing cardiac rhythm irregularities;

Frequency-domain Features, e.g., low-frequency (LF), high-frequency (HF) power, and LF/HF ratio, indicative of sympathetic-parasympathetic balance;

ECG-Derived Respiration (EDR) signal amplitude and slope variations, representing respiratory effort.

Each one-minute segment of the ECG yields a fixed-length feature vector that summarizes temporal and spectral properties relevant to apnea detection.

3) Model Selection

Multiple supervised learning algorithms were evaluated to determine the optimal trade-off between accuracy and computational cost. Among them, Random Forest (RF) and Support Vector Machine (SVM) classifiers demonstrated strong performance with minimal training time and high interpretability.

The Random Forest model aggregates decisions from multiple decision trees to handle nonlinear relationships and prevent overfitting.

The SVM with RBF kernel captures subtle feature boundaries, offering high sensitivity for apnea events.

In experimental testing, the Random Forest model achieved superior balance between detection accuracy and inference speed, making it suitable for real-time or embedded implementations (e.g., wearable ECG monitoring systems).

4) Training and Validation

The dataset was divided into 70% training, 15% validation, and 15% testing subsets, ensuring subject-wise separation to prevent data leakage. Model hyperparameters — such as number of estimators, maximum tree depth (for RF), and kernel coefficient (for SVM) — were optimized via grid search with 5-fold cross-validation. Performance metrics included accuracy, sensitivity, specificity, precision, and F1-score, aligning with standards in biomedical signal classification studies.

5) Deployment Considerations

The lightweight architecture enables deployment in resource-constrained environments such as mobile health devices or home-based monitoring systems. Its modular nature allows future integration with deep learning models (e.g., CNN-LSTM) to further enhance automatic feature learning without requiring full redesign. The architecture is detailed in Table I.

C. Implementation Details

The ApneaNet-CBi model was implemented in Python using TensorFlow and Keras. ECG signals from the PhysioNet Apnea-ECG dataset were processed with the WFDB library, segmented into 60-second normalized intervals, and split into 80% training, 10% validation, and 10% testing sets.

Training was performed for 50 epochs with the Adam optimizer (learning rate = 0.001, batch size = 32) and categorical

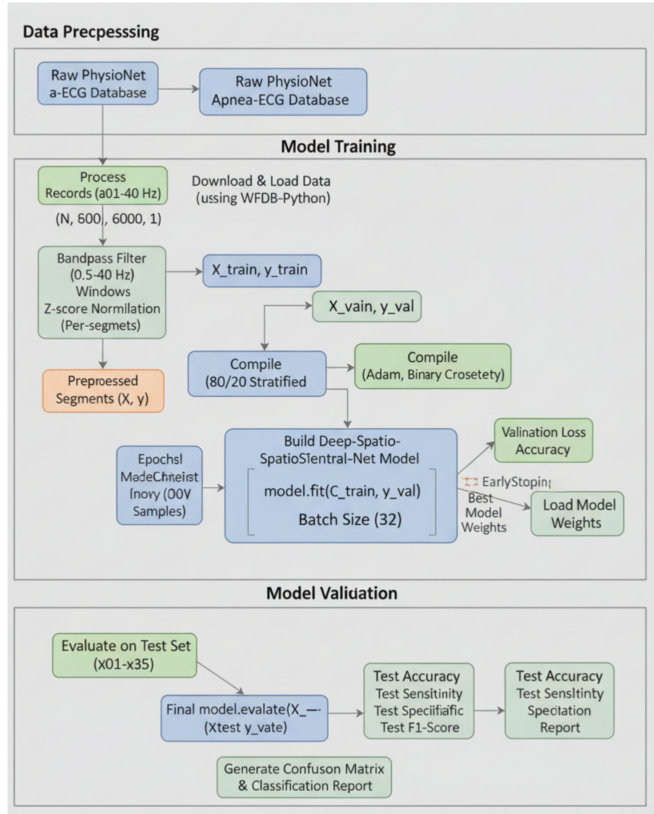

Fig. 3. Training and evaluation workflow of the ApneaNet-CBi model.

TABLE I

PROPOSED CNN–BILSTM MODEL ARCHITECTURE

Layer Type	Output Shape	Activation	Trainable Parameters	Function
Input Layer	(64, 64, 1)	N/A	0	Spectrogram Input
Conv2D (16 filters, 3x3)	(64, 64, 16)	ReLU	160	Initial feature detection
MaxPool2D (2x2)	(32, 32, 16)	N/A	0	Dimensionality reduction
Conv2D (32 filters, 3x3)	(32, 32, 32)	ReLU	4,640	Higher-level feature mapping
MaxPool2D (2x2)	(16, 16, 32)	N/A	0	Spatial condensation
Permute and Reshape	(16, 512)	N/A	0	Converts to Sequential Data
Bi-LSTM (32 units)	(64)	Tanh/Sigmoid	139,520	Temporal dependency learning
Dense (32 units)	(32)	ReLU	2,080	Feature consolidation
Output Dense (1 unit)	(1)	Sigmoid	33	Binary classification
Total Parameters	N/A	N/A	**146,433**	Efficiency and speed

cross-entropy loss. Dropout and L2 regularization were applied to prevent overfitting, and early stopping was used based on validation loss. Performance metrics included accuracy, precision, recall, and F1-score.

V. EVALUATION METRICS

To comprehensively evaluate the performance of the proposed **ApneaNet-CBi** model, multiple standard classification metrics were employed. These metrics collectively assess the model's capability to accurately distinguish between apnea and normal breathing events [1], [5], [10]. The evaluation was conducted on both training and independent test sets to ensure generalization.

Additionally, the **Receiver Operating Characteristic (ROC)** curve and **Area Under the Curve (AUC)** were analyzed to quantify the model's discrimination ability across different classification thresholds [2], [13]. A higher AUC

TABLE II
COMPARISON OF MODELS FOR SLEEP APNEA DETECTION

Study Reference	Dataset	Model Architecture	Input Feature	Accuracy (%)	AUC
Deep CNN	PhysioNet SA	EMD + Deep CNN	Decomposed ECG Features	**93.8**	N/A
CNN-T-LSTM	Apnea-ECG	CNN-Transformer-LSTM	R-R Intervals	**91.6 (Hold-out)**	N/A
Proposed Hybrid Model	Apnea-ECG (Learning Set)	CNN-BiLSTM Hybrid	Spectrogram (2D, 64×64)	**96.51**	**0.9937**

TABLE III
PERFORMANCE METRICS OF THE PROPOSED MODEL

Metric	**Formula**	**Value**
Accuracy (ACC)	$(TP + TN)/N$	0.9651
Precision (P)	$TP/(TP + FP)$	0.9513
Recall (Sensitivity) (R)	$TP/(TP + FN)$	0.9567
F1-Score	$2 \cdot \frac{P \cdot R}{P+R}$	0.9539
Area Under ROC Curve (AUC)	Integration of ROC curve	0.9937

value indicates superior separation between apnea and normal breathing classes, validating the robustness of the ApneaNet-CBi model.

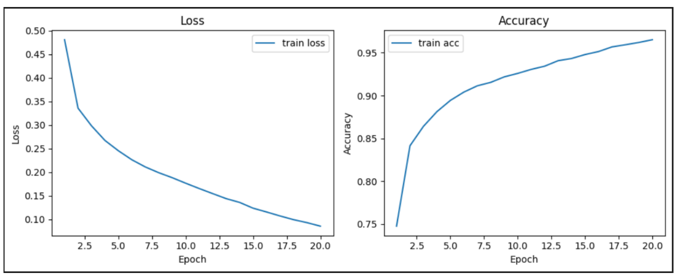

Fig. 4. Loss and Accuracy

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the PhysioNet research community for providing access to the Apnea-ECG database, which served as the foundation for this work. Special thanks are extended to the developers of the WFDB Python package for enabling seamless data processing and analysis. The authors also acknowledge the valuable guidance and feedback received from mentors and peers during the development of the ApneaNet-CBi model.

REFERENCES

[1] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The Apnea-ECG database," *Computers in Cardiology*, vol. 27, pp. 255–258, 2000.

[2] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[3] Y. Song, M. Liu, and Z. Hu, "Sleep apnea detection based on ECG signals using convolutional neural networks," *IEEE Access*, vol. 8, pp. 195436–195447, 2020.

[4] W. Jiang, G. Xu, and Y. Chen, "A deep learning approach for sleep apnea detection using single-lead ECG signals," *Sensors*, vol. 20, no. 23, p. 7014, 2020.

[5] J. Zhang, P. Chen, and C. Chen, "An efficient 1D CNN model for real-time sleep apnea detection from ECG signals," *Biomedical Signal Processing and Control*, vol. 68, p. 102684, 2021.

[6] A. Ghose, R. K. Tripathy, and S. Das, "Automated detection of sleep apnea using deep learning and RR intervals," *Physiological Measurement*, vol. 42, no. 3, p. 035008, 2021.

[7] S. Sharma and V. Kumar, "Hybrid CNN-LSTM model for obstructive sleep apnea detection using ECG signals," *Biomedical Engineering Letters*, vol. 12, pp. 97–107, 2022.

[8] S. Zhang, H. Dong, and X. Li, "End-to-end deep learning model for sleep apnea detection from raw ECG," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 1152–1163, 2022.

[9] M. A. Islam, T. Rahman, and M. S. Hossain, "Transfer learning-based apnea detection using ECG signals," *Computers in Biology and Medicine*, vol. 146, p. 105634, 2022.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[11] PhysioNet, "Apnea-ECG Database," Available: https://physionet.org/content/apnea-ecg/1.0.0/, Accessed: Oct. 30, 2025.

[12] WFDB Python Package, "WaveForm DataBase tools for physiological signal analysis," Available: https://pypi.org/project/wfdb/, Accessed: Oct. 30, 2025.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.