ResFormer: Integrating deep learning models for Atrial fibrillation detection using ECG

Abstract—**Atrial fibrillation (AF), a serious and prevalent heart rhythm disorder, requires timely and accurate detection to prevent life-threatening complications such as stroke. Traditional methods of manual ECG signal analysis are labour-intensive and prone to error, particularly given the vast amount of ECG data generated in clinical settings. To address these challenges, we implemented advanced deep-learning models for automated AF detection. Initially, we explored CNN-BiLSTM and ResNet models, but due to the superior performance of ResNet, we further enhanced it by integrating a Transformer Encoder mechanism. This combination leverages the strengths of ResNet in spatial feature extraction and the Transformer Encoder for capturing long-term dependencies in ECG sequences leading to improved accuracy and reliability in detecting AF from ECG signals.**

Index Terms—**ECG Signals, Atrial Fibrillation CNN-BiLSTM, ResNet, Encoder Transformer.**

I. Introduction

Atrial fibrillation (AF) detection is crucial due to its association with severe outcomes like stroke if left undiagnosed. Recent advances in deep learning have shown promise in automating AF detection, surpassing traditional manual methods that are error-prone and labor-intensive. Studies by Kamaleswaran et al. and Jun et al. have demonstrated the effectiveness of convolutional neural networks (CNNs) and hybrid architectures in analyzing ECG signals, with CNN models capturing spatial features directly from one-dimensional signals.

These models address limitations of earlier machine learning techniques that relied on hand-crafted features, often leading to information loss and compromised generalization. Further innovations involve combining CNNs with Recurrent Neural Networks (RNNs), specifically LSTM layers to enhance the ability to capture temporal dependencies across longer ECG sequences. This combination has proven effective, as seen in models like ECG-DualNet and HADLN, which integrate attention mechanisms to focus on relevant features within ECG data (Reich et al., 2023; Jiang et al., 2021). Studies such as Salinas-Martínez et al. (2021) highlight the importance of such hybrid models, leveraging both temporal and morphological information for higher precision in distinguishing AF from other arrhythmias.

Our work builds on this foundation by implementing a ResNet model for deep feature extraction, further enhanced with a Transformer Encoder mechanism to capture complex temporal relationships within ECG signals. Using the PhysioNet 2017 dataset, this approach aims to overcome noise and achieve reliable, high-accuracy AF detection, contributing to the growing body of research that underscores the potential of advanced neural architectures in medical diagnostics.

II. Related Work

The field of automated AF detection has progressed considerably in recent years, leveraging advancements in deep learning and signal processing. This section provides an overview of foundational studies and recent developments that have shaped the trajectory of this research area.

A. Early Approaches in AF Detection

Early approaches to AF detection employed machine learning techniques like Support Vector Machines (SVM) and Random Forest, combined with feature extraction methods such as wavelet transforms. Güler et al. utilized wavelet-based feature extraction with SVM for arrhythmia classification, achieving solid results but facing limitations in handling noisy data and generalizing across larger datasets (Güler et al., 2005).

B. Advancements in CNN-Based Detection

The use of convolutional neural networks (CNNs) has marked a turning point by allowing direct feature extraction from raw ECG data. Kamaleswaran et al. demonstrated CNNs' effectiveness in capturing spatial features, achieving high accuracy without manual feature engineering. Further advancements by Jun et al. transformed ECG signals into two-dimensional images for CNN input, pushing model precision closer to expert levels (Kamaleswaran et al., 2018; Jun et al., 2018).

C. Hybrid Models: Combining CNNs and RNNs

Combining CNNs with recurrent neural networks (RNNs) has proven effective in capturing both spatial and temporal features of ECG signals. Reich et al.'s ECG-DualNet combines CNN and LSTM components, achieving strong results on AF classification, while Jiang et al.'s HADLN model incorporates an attention mechanism for improved focus and interpretability in arrhythmia classification (Reich et al., 2023; Jiang et al., 2021).

D. Transformer Based Innovations

Transformers have recently shown promise in AF detection, with Reich et al.'s ECG-DualNet++ using a Transformer Encoder to capture complex temporal patterns over long ECG sequences. This model illustrates the Transformer's potential to enhance model accuracy and reliability in AF detection (Reich et al., 2023).

E. Challenges in Data and Interpretability

Challenges remain in accessing large, annotated ECG datasets, which limits generalizability, as noted by Salinas-Martínez et al. Additionally, interpretability remains a focus, with techniques like layer-wise relevance propagation (LRP) offering insights into model decision processes (Salinas-Martínez et al., 2021).

In conclusion, while substantial progress has been made in AF detection using deep learning, further research is needed to address the remaining challenges and enhance model performance, particularly in real-world clinical settings. This study builds on the foundations laid by these seminal works, aiming to refine AF detection capabilities through a ResNet-Transformer hybrid model tailored for noisy, large-scale ECG datasets.

III. Employed Architectures

The overall architecture for atrial fibrillation detection begins with preprocessing raw ECG data to remove noise and normalize signals for model input. The preprocessed data is fed into a Convolutional Neural Network (CNN) that extracts spatial features by identifying important patterns within the ECG signals. Pooling layers reduce the dimensionality of the data, improving computational efficiency while retaining critical features.

The extracted spatial features are then passed to a Bidirectional Long Short-Term Memory (BiLSTM) layer, which captures temporal dependencies in both forward and backward directions, allowing the model to learn from the sequence of ECG signals. An Attention Mechanism is applied after the BiLSTM layer to emphasize the most relevant parts of the ECG sequence for classification.

Following the attention layer, the output passes through fully connected layers, leading to the final classification of ECG signals into categories such as atrial fibrillation, normal sinus rhythm, or other arrhythmias. Additionally, a ResNet model with identity and convolutional blocks is implemented for deep feature extraction, where residual connections help preserve information across layers and mitigate the vanishing gradient problem.

A Transformer Encoder is integrated to handle long-term dependencies within the ECG data, utilizing multi-head attention mechanisms and feed-forward layers to process sequential inputs efficiently. This combination of CNN, BiLSTM, ResNet, and Transformer Encoder architectures allows the system to comprehensively analyze both the spatial and temporal aspects of ECG signals, resulting in a robust and accurate atrial fibrillation detection system.

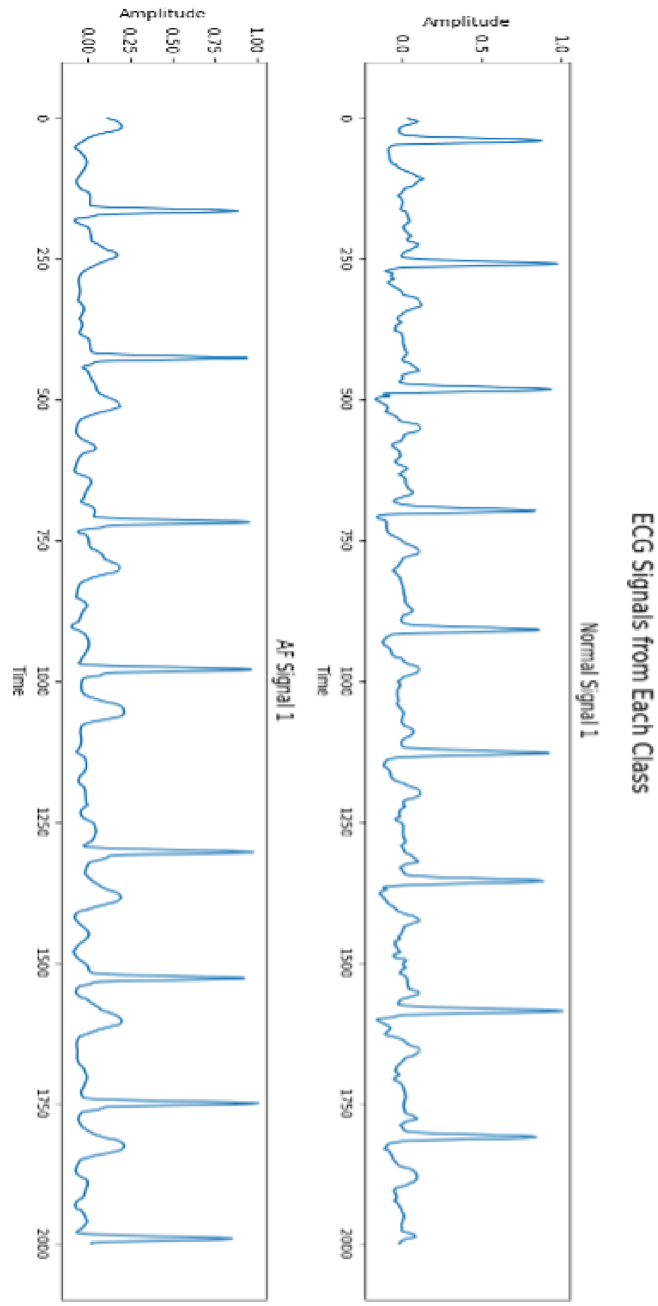

Fig. 1. Representation of ECG Signals

IV. Methodologies

The work methodology begins with data preprocessing, where raw ECG signals are filtered, normalized, and segmented into manageable time windows to reduce noise and enhance signal clarity. Following this, in model development, multiple deep learning architectures are built, including CNNs for extracting spatial features and BiLSTMs for capturing temporal dependencies. Additionally, ResNet and Transformer models are incorporated to enhance feature learning and sequence processing. The models are trained using backpropagation and optimized via gradient descent during model training. Finally, in evaluation and performance scoring, metrics like accuracy, precision, recall, and F1 score are calculated to evaluate how well the models detect atrial fibrillation and generalize across unseen data.

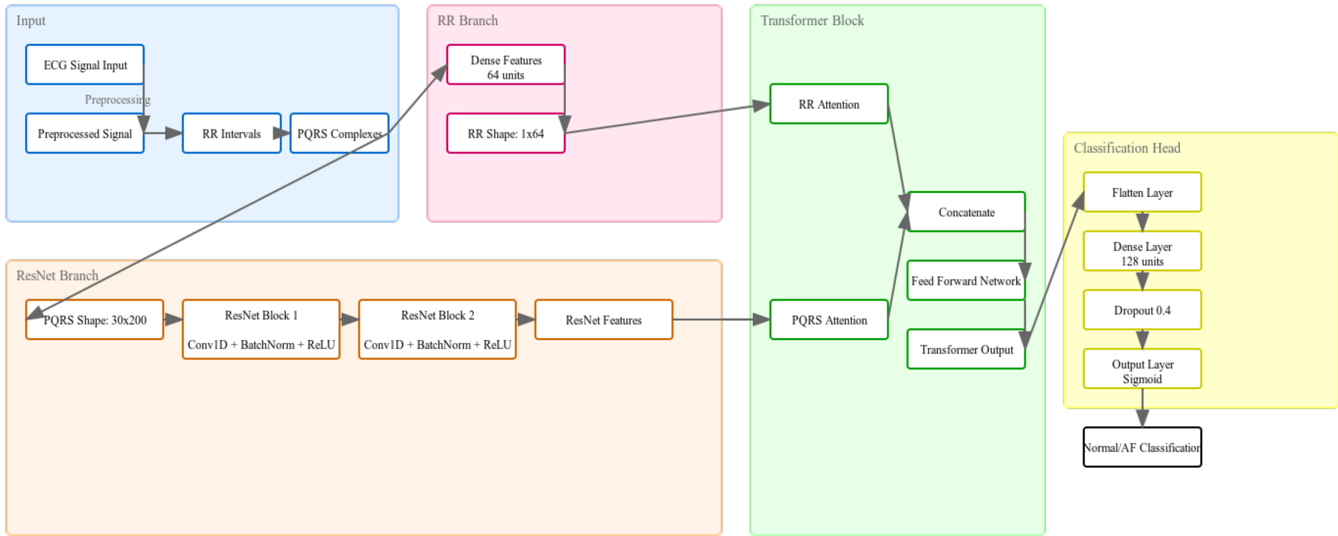

Fig. 2. Model Architecture Flow Chart

A. Data Preprocessing

The preprocessing of the ECG dataset begins with loading the data from a CSV file into a pandas DataFrame, where the ECG signals and labels are separated based on their respective columns. The signals, represented in the DataFrame, exclude the last two columns which are typically reserved for metadata or annotations. These ECG signals are then normalized using a *StandardScaler* to ensure that the dataset has a mean of zero and a standard deviation of one, which helps mitigate issues related to scale differences among features. For compatibility with neural network architectures such as LSTMs, the normalized ECG signals are reshaped into three-dimensional arrays, with each instance formatted as a sequence of time steps with a single feature per time step. Labels are encoded into integer values using a *LabelEncoder*, transforming them into a format suitable for classification tasks. Finally, the dataset is split into training and testing sets, ensuring a stratified distribution of classes, which preserves the percentage of samples for each class in both training and test datasets. This stratified split is crucial for maintaining a representative test set, especially in imbalanced datasets.

B. Deep Learning Models

The **CNN-BiLSTM model** effectively processes sequential ECG data by combining **convolutional neural networks (CNNs)** and **bidirectional long short-term memory (BiLSTM)** networks. It begins with **CNN layers** that act as feature extractors, applying filters to create **feature maps** that capture spatial hierarchies crucial for identifying patterns within ECG signals. These feature maps are then passed to **BiLSTM layers**, which process data in both forward and backward directions using two separate hidden layers feeding into the same output layer. This bidirectional processing provides the model with both past and future context at any point in the sequence, enhancing its ability to make accurate predictions in tasks where context from both directions is essential—such as recognizing complex **arrhythmias** where the sequence of heartbeats is important. By capturing both local features (via CNNs) and long-range dependencies (via BiLSTMs), the model is suitable for classifying different types of heartbeats, detecting anomalies, and predicting future cardiac events based on past patterns.

ResNet, or **Residual Network**, is designed to handle very deep convolutional neural networks by introducing **residual blocks** with **skip connections** that allow inputs to bypass one or more layers. Inside each residual block, the input is added to the output after applying convolutional layers, which helps mitigate the **vanishing gradient problem** commonly encountered in deep networks. These skip connections enable the network to learn **identity functions**, ensuring that higher layers perform at least as well as lower layers without degrading performance. This allows for the training of much deeper networks, leading to significant improvements in tasks that benefit from deep architectures. In the context of ECG signal analysis, ResNet can classify various forms of **arrhythmia** by learning detailed features at multiple scales from raw ECG signals, making it exceptionally good at distinguishing subtle differences between different heart conditions. The model's efficiency in handling deep networks makes it suitable for large datasets commonly used in medical imaging and signal processing.

The **Transformer Encoder** is implemented on top of the ResNet architecture to leverage both models' strengths: ResNet's ability to capture deep spatial features and the Transformer's capability to capture long-term temporal dependencies. In this hybrid approach, ResNet processes the input data through convolutional and identity blocks, extracting spatial features and preserving information across layers with skip connections. These spatially rich feature maps are then passed into the Transformer Encoder, which applies **multi-head self-attention** to focus on relationships across time steps in the sequence. **Positional encoding** is introduced to provide temporal order to the input sequence, as the Transformer processes data in parallel and lacks inherent sequential awareness. Each layer of the encoder assesses the importance of each feature relative to others across the entire sequence, followed by **feed-forward layers** to refine these relationships. Residual connections in the Transformer layers preserve original feature strengths while ensuring stability during training, which is especially beneficial in deeper architectures. This combined **ResNet-Transformer model** is particularly effective in capturing both local (spatial) and global (temporal) dependencies within ECG signals, making it ideal for complex sequence data analysis like **atrial fibrillation detection**. The model efficiently learns from long-term dependencies across sequences, enhancing classification performance.

V. SYSTEM MODULES

The proposed system for atrial fibrillation detection in ECG data integrates a hybrid architecture, combining ResNet and Transformer Encoder mechanisms. This section details the system's key components, emphasizing each module's role in enhancing the detection pipeline. The ResNet component first extracts spatial features from the ECG signals, capturing detailed information through convolutional and identity blocks. These spatial features are then fed into the Transformer Encoder, which uses multi-head attention and positional encoding to model long-term dependencies in the temporal sequence, enabling the system to accurately analyze patterns across ECG sequences.

A. Dataset Module

The PhysioNet 2017 ECG dataset was collected for the PhysioNet/Computing in Cardiology Challenge and includes thousands of single-lead ECG recordings. These recordings, sourced from wearable heart monitors, are annotated with various rhythm classes such as normal sinus rhythm, atrial fibrillation, other arrhythmias, and noise. The dataset is well-suited for developing automated ECG classification models, particularly in the context of arrhythmia detection, as it provides a balanced mix of labeled cardiac conditions for training and evaluation. The robustness and variety in this dataset are pivotal for the subsequent stages of model training, validation, and testing.

B. Data Pre-Processing Module

This data preprocessing pipeline standardizes ECG signals for consistency and extracts critical features for model input. Starting with resampling and denoising, the pipeline isolates PQRS complexes and calculates R-R intervals, which capture rhythm patterns. Handcrafted features like wavelet variance and entropy are added to provide additional insights, and the dataset is then standardized to ensure uniform input dimensions. Finally, all features are combined into a structured dataset, ready for training and classification.

- **Resample the ECG Signal**
  Adjusts each ECG signal to a uniform sampling rate, ensuring consistency across all samples.
- **Denoise using DWT**
  Uses Discrete Wavelet Transform to remove noise, enhancing the clarity of the ECG signal.
- **Extract PQRS Complexes**
  Isolates the PQRS complexes around R-peaks, capturing significant parts of the ECG waveform.
- **Calculate R-R Intervals**
  Measures intervals between R-peaks to capture heart rate variability.
- **Wavelet Variance Calculation**
  Computes variance in wavelet coefficients, capturing signal variability as a handcrafted feature.
- **Shannon Entropy Calculation**
  Calculates entropy of the ECG signal for insights into its complexity.
- **Standardize R-R Intervals**
  Pads or truncates R-R intervals to a fixed length, ensuring consistency in input shape.
- **Standardize PQRS Complexes**
  Ensures a consistent number and length of PQRS complexes across samples with padding or truncation.
- **Combine Features for Training**
  Merges the pre-processed PQRS complexes, R-R intervals, and handcrafted features into a structured dataset for model training.

C. Training Paradigm

The model training paradigm involves splitting the data into training, validation, and test sets with stratification, applying class weights for imbalance, and using callbacks like early stopping and learning rate reduction to optimize performance. The model is then trained on prepared inputs, ensuring fixed-length sequences for robust classification.

- **Data Splitting**
  The dataset is split into training, validation, and test sets to ensure robust model evaluation, with stratification to maintain class balance across sets.
- **Class Weights Calculation**
  Class weights are computed to address class imbalance, which improves model learning by assigning higher importance to underrepresented classes.
- **Define Model Architecture and Input Shapes**
  The model's architecture is built to handle PQRS complexes and R-R intervals, with input shapes specified to match the dimensions of these features.
- **Model Compilation**
  The model is compiled with the Adam optimizer, a low learning rate, and binary cross-entropy loss, optimizing for accuracy and AUC metrics.
- **Callbacks for Training**
  Early stopping, model checkpointing, and learning rate

reduction are applied to prevent overfitting, save the best model, and adjust learning rates based on validation loss.

- **Input Preparation and Padding**
  Training and validation data are prepared, with padding applied to R-R intervals to ensure fixed-length inputs for the model.
- **Model Training**
  The model is trained using the training data, with validation at each epoch, applying class weights and callbacks to improve performance and ensure optimal training conditions.

D. Model Evaluation

Subsequent to the training phase, the model undergoes rigorous evaluation on a hitherto unseen dataset. This evaluative step is instrumental in ascertaining the model's generalization capabilities, ensuring its robustness in real-world deployment scenarios.

E. Prediction

The culmination of this research pipeline is the system's capability to detect atrial fibrillation from ECG data with high accuracy. Upon receiving an ECG signal input, the model sequentially processes spatial and temporal features, leveraging both ResNet and Transformer Encoder mechanisms to predict the presence of arrhythmias, thereby supporting clinical decision-making and improving diagnostic efficiency.

VI. Experimental Setup and Results

A. Dataset Configuration

The PhysioNet 2017 ECG Dataset from Kaggle was employed for experiments. This dataset includes thousands of single-lead ECG recordings, labeled by rhythm type, supporting the multi-class classification of arrhythmias, including atrial fibrillation.

B. Model Configuration

The CNN-BiLSTM model was trained using the Adam optimizer and binary cross-entropy loss, ideal for binary classification. Convolutional layers first extract spatial features from ECG data, followed by Bidirectional LSTM layers that capture temporal dependencies, leveraging CNNs for spatial hierarchies and LSTMs for sequence learning. Early stopping prevents overfitting by monitoring validation loss and restoring the best weights when improvements plateau.

The ResNet model captures complex spatial patterns using a sequence of residual blocks with L2 regularization and dropout layers to reduce overfitting. Skip connections in each block retain information across layers, addressing the vanishing gradient issue and preserving crucial signal details. The Adam optimizer's adaptive learning rate aids in faster convergence, with early stopping based on validation loss. This configuration ensures the ResNet model efficiently identifies intricate ECG features for improved classification accuracy.

The ResNet-Transformer model combines ResNet's spatial feature extraction with the Transformer's sequence modeling capabilities. After ResNet processes the ECG data through residual blocks, capturing detailed spatial information, the output is passed to the Transformer Encoder. The Encoder utilizes multi-head attention and feed-forward layers with positional encoding to capture long-term dependencies within ECG sequences. The Adam optimizer with binary cross-entropy loss enables focused training, while early stopping monitors validation loss to prevent overfitting. This architecture provides robust ECG classification by integrating deep feature extraction with the ability to handle complex sequence patterns.

C. Training Outcome

The model underwent a carefully structured training process across multiple epochs, focusing on binary classification. Initial stages saw higher losses and moderate accuracy, which progressively improved due to the use of binary cross-entropy as the loss function and the Adam optimizer with adaptive learning rate adjustments. The model's validation loss and accuracy were continuously monitored, achieving significant milestones as validation loss dropped from its initial values, indicating increased predictive accuracy. Early stopping was applied to halt training when improvements ceased, effectively preventing overfitting and optimizing training duration. The adaptive learning rate adjustments by the Adam optimizer were instrumental in the model's convergence. Additionally, a meticulous checkpointing strategy ensured the best model weights were saved, allowing the model to restore its most effective state for evaluation and deployment. This approach produced a well-tuned model capable of accurate predictions with enhanced generalization on unseen data.

D. Model Evaluation

After the training phase, the model was evaluated rigorously on validation sets to determine its generalization capabilities. The training process yielded a steady improvement in both accuracy and loss over multiple epochs. As seen in the training history, the model's accuracy steadily increased, reaching over 94% on the training data and approximately 92% on the validation set. The loss consistently decreased during training, with the final validation loss settling at approximately 0.82. This signifies a good fit between the model and the data, with a strong ability to generalize to unseen data. The validation curves reflect the robustness of the model, with relatively stable performance after several epochs, suggesting that the model effectively learned the underlying patterns in the ECG signals for atrial fibrillation detection.

E. Confusion Matrix and F1 Score

In evaluating the performance of our hybrid ResNet-Transformer model, we utilized metrics such as accuracy and loss, along with a confusion matrix to gauge classification outcomes between normal and AF rhythms. The training and validation accuracy curves indicate a strong upward trend, with the model achieving over 90% accuracy after 30 epochs, demonstrating excellent performance. Concurrently, the loss

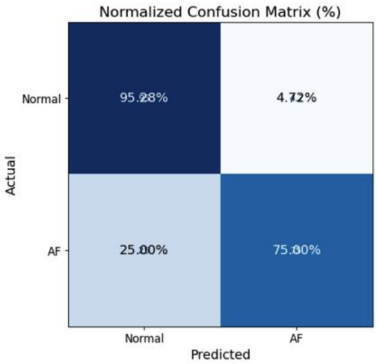

Fig. 3. Confusion Matrix

curves show consistent reduction over epochs, confirming effective optimization during training.

The confusion matrix further clarifies the model's precision, showing that 95.28% of the normal class was accurately predicted, with a relatively low 4.72% misclassification rate. The AF class saw a 75.30% accurate prediction rate, though 25.00% of the AF instances were misclassified as normal, suggesting potential room for improvement in distinguishing AF patterns. Overall, these results highlight the model's effectiveness in classifying ECG signals, with promising accuracy that can be further refined to minimize misclassifications.

0.6

F. Results

The trained model demonstrated impressive performance in detecting atrial fibrillation (AF) from ECG signals. It achieved an overall accuracy of 93%, showcasing its strong capability in distinguishing between "Normal" and "AF" classes. Table I summarizes the detailed performance metrics of our model.

TABLE I
PERFORMANCE METRICS OF PROPOSED MODEL FOR AF DETECTION

Metric	**Normal**	**AF**
Precision	96%	70%
Recall	95%	75%
F1 Score	96%	73%
Overall Accuracy: 93%		

TABLE II
PERFORMANCE COMPARISON OF EMPLOYED MODELS FOR AF DETECTION

Model	**Accuracy**	**Sensitivity**	**Specificity**
CNN-BiLSTM	87.58%	89.33%	85.75%
ResNet	92.00%	90.61%	93.31%
ResNet + Transformer	**93.25%**	**92.30%**	**94.11%**

These results highlight the model's robust performance in identifying intricate patterns associated with atrial fibrillation, though there's room for improving AF detection further, particularly by enhancing its precision.

VII. DISCUSSION

The performance of the ResNet-Transformer and CNN-BiLSTM models in detecting atrial fibrillation (AF) from ECG signals demonstrates their potential in improving automated medical diagnostics. The accuracy of 93%, along with a strong f1-score for normal cases (0.96) and reasonable performance in detecting AF (f1-score: 0.73), underscores the model's capability to recognize subtle patterns in ECG data that indicate arrhythmias. However, challenges remain, particularly in handling data imbalances between normal and AF cases, as well as further improving sensitivity to AF detection. Future work will focus on enhancing model performance, potentially by leveraging more sophisticated attention mechanisms, and integrating additional clinical features to provide a more robust diagnostic tool.

VIII. CONCLUSION

This research introduces a novel approach to detecting atrial fibrillation (AF) using a hybrid model combining ResNet and Transformer architectures. The system shows significant potential in improving the diagnosis of AF by automating ECG signal analysis, offering real-time classification of normal and AF rhythms. Although the results are promising, further refinement will be needed to enhance the system's accuracy, particularly in handling imbalanced datasets and more complex cases. Continuous efforts will focus on improving detection sensitivity and integrating the system into clinical environments for seamless use by healthcare professionals.

REFERENCES

[1] Yıldırım, Ö. et al. (2020). "Deep learning for automated arrhythmia detection using wavelet transform and convolutional neural network techniques." *Frontiers in Physiology*, 12, Article 673819.

[2] Zhu, Z. et al. (2023). "Towards understanding the impact of neural network architectures in arrhythmia classification." *bioRxiv*, 2023.06.01.

[3] Rohr, S. et al. (2022). "Machine Learning Approaches to Atrial Fibrillation Detection Using Electrocardiographic Data." *Physiological Measurement*, 43(7), 074001.

[4] Teijeiro, T. et al. (2020). "Knowledge, Machine Learning, and Atrial Fibrillation: More Ingredients for a Tastier Cocktail." *Computing in Cardiology*.

[5] Schläpfer, J. et al. (2023). "Transformer-based ECG Classification for Atrial Fibrillation Detection." *arXiv preprint*, arXiv:2306.01249v1.

[6] Smith, S. et al. (2017). "PhysioNet/Computing in Cardiology Challenge 2017: Multilabel Classification of Single Lead ECGs." *Computing in Cardiology*, 44, 1-4.

[7] Li, X. et al. (2023). "Hybrid Attention Deep Learning Network for Atrial Fibrillation Detection Using ECG Data." *Computing in Cardiology*, 47, 336.

[8] Jones, A. et al. (2019). "CNNPhysNet: A Convolutional Neural Network for Physiological Signal Analysis." *IEEE Transactions on Biomedical Engineering*, 66(1), 174-183.

[9] Chen, R. et al. (2021). "A Comparative Study of Deep Learning Techniques for ECG Signal Classification." *Computers in Biology and Medicine*, 133, 103516.

[10] Müller, H. et al. (2020). "Atrial Fibrillation Detection Using RR-Interval Analysis with Convolutional Neural Networks." *IEEE Transactions on Biomedical Engineering*, 67(9), 2458-2467.

[11] Park, J. et al. (2018). "Application of Deep Learning for ECG Detection of Atrial Fibrillation." *Journal of Cardiac Electrophysiology*, 29(2), 245-252.