

## HYBRID RESNET–VISION TRANSFORMER WITH CNN ATTENTION FOR AUTOMATED MULTI-LEAD ECG ARRHYTHMIA DETECTION

G. Jemilda

Professor, Dept. of CSE,

Jayaraj Annapackiam CSI College of Engineering,  
Nazareth, India

[jemildag@gmail.com](mailto:jemildag@gmail.com)

M. Mahalakshmi

PG Scholar, Dept. of CSE,

Jayaraj Annapackiam, CSI College of Engineering,  
Nazareth, India

[muruganmoses9787@gmail.com](mailto:muruganmoses9787@gmail.com)

**Abstract** - Electrocardiogram (ECG) signals are widely used for detecting cardiac abnormalities and diagnosing heart diseases. However, accurate arrhythmia detection from multi-lead ECG signals is challenging due to noise, signal variations, and complex temporal patterns. This work proposes a Hybrid ResNet–Vision Transformer with CNN Attention (HRVT-CA) framework for automated ECG arrhythmia classification. Initially, ECG signals are preprocessed using a bandpass filter to remove noise and improve signal quality. Then, Recursive Feature Elimination (RFE) is applied to select the most relevant features and reduce redundancy. The selected features are processed using a hybrid deep learning architecture that combines ResNet for feature extraction, Vision Transformer for capturing global dependencies, and CNN attention for emphasizing important signal patterns. The model performance is evaluated using accuracy, precision, recall, and F1-score. Experimental results demonstrate that the proposed model improves arrhythmia detection performance and provides an effective solution for automated ECG-based cardiac diagnosis.

**Keywords** - ECG Signal Processing, Arrhythmia Detection, ResNet, Vision Transformer, CNN Attention, Feature Selection.

### I. INTRODUCTION

Electrocardiogram (ECG) analysis plays a vital role in the early diagnosis of cardiovascular diseases. Arrhythmia is one of the most common heart disorders caused by irregular electrical activity of the heart. Accurate detection of arrhythmia using multi-lead ECG signals is essential for timely medical intervention and improved patient outcomes. However, ECG signals often contain noise, baseline drift, and complex morphological variations, which make automated detection challenging. Traditional machine learning techniques rely on handcrafted features and may fail to capture complex temporal relationships present in ECG signals. Recently, deep learning models such as Convolutional Neural Networks (CNNs) and Residual Networks (ResNet) have shown promising performance in ECG classification tasks due to their ability to automatically extract features from raw signals. However, CNN-based models mainly focus on local patterns and may struggle to capture long-range dependencies in ECG signals.

Vision Transformers (ViT) have recently gained attention for their capability to model global dependencies through self-attention mechanisms. By combining CNN-based feature extraction with transformer-based global

context modeling, hybrid architectures can significantly improve classification performance.

In this work, we propose a Hybrid ResNet Vision Transformer with CNN Attention model for automated multi-lead ECG arrhythmia detection. The proposed framework integrates signal preprocessing, feature selection using Recursive Feature Elimination (RFE), and a hybrid deep learning model that combines ResNet, Vision Transformer, and CNN attention mechanisms. The main contributions of this work include:

- A preprocessing pipeline using bandpass filtering for noise reduction in ECG signals.
- Feature selection using Recursive Feature Elimination to improve model efficiency.
- A hybrid architecture integrating ResNet, Vision Transformer, and CNN attention.
- Comprehensive evaluation using accuracy, precision, recall, and F1-score.

The proposed model improves arrhythmia classification performance and provides an effective solution for automated ECG analysis in clinical decision-support systems.

### II. LITERATURE REVIEW

O. Kovalchuk et al. (2025) proposed an explainable deep learning approach for ECG-based arrhythmia detection. The method uses an enhanced R-peak detection technique and a modified CNN architecture that analyzes three consecutive cardiac cycles to capture temporal patterns. An explainable AI mechanism is integrated to interpret the CNN's decisions using clinically relevant ECG features. Experiments on the MIT-BIH dataset achieved 99.43% accuracy with F1-scores close to 100% for major arrhythmia classes.

P. G. Gaonkar et al. (2025) presented a deep learning–based approach for automatic ECG arrhythmia detection to support early diagnosis in clinical practice. The study employed a convolutional neural network (CNN) to classify ECG signals into five different cardiac rhythm categories. Experimental results demonstrated the effectiveness of the model, achieving an average classification accuracy of 0.95 with a loss value of 0.2, indicating its reliability for assisting medical professionals in arrhythmia detection.

Z. Khatar et al. (2025) proposed a deep learning framework called GAF-GradCAM for ECG-based

arrhythmia detection. The method converts ECG signals into two-dimensional Gramian Angular Field (GAF) matrices to capture temporal and frequency characteristics, where Grad-CAM guides a dynamic weighted fusion of these features for improved interpretability. A hybrid parallel-residual architecture combined with Bi-LSTM is then used for robust feature extraction and classification. Experimental results showed strong performance with training accuracy of 99.68%, validation accuracy of 98.78%, and overall classification accuracy of 98.75%, demonstrating the effectiveness of the proposed fusion-based approach for arrhythmia detection.

N. Alamatsaz et al. (2024) proposed a lightweight hybrid CNN-LSTM model for ECG-based arrhythmia detection. The study applied preprocessing techniques such as resampling and baseline wander removal before feeding ECG segments into an 11-layer end-to-end deep learning network. The model was designed to classify eight types of cardiac arrhythmias along with normal rhythm without requiring manual feature extraction. Experiments conducted on the MIT-BIH Arrhythmia and Long-Term AF databases achieved a mean diagnostic accuracy of 98.24%, demonstrating the effectiveness of the lightweight framework for automated arrhythmia detection and its potential use in portable monitoring devices.

S. Chen et al. (2024) proposed a Swin Transformer-based deep learning approach for ECG arrhythmia detection using time-frequency characteristics. The method first applies wavelet thresholding to remove noise and artifacts from ECG signals, followed by feature extraction using complex Morlet wavelets to generate time-frequency maps. These maps are then classified using a Swin Transformer model that utilizes hierarchical structure and self-attention mechanisms to capture both local and global features. Experimental evaluation on the MIT-BIH arrhythmia dataset achieved classification accuracies of 99.34% for intra-patient and 98.37% for inter-patient analysis, demonstrating the effectiveness of the proposed approach for automated arrhythmia detection.

A. Pokharel et al. (2024) proposed a machine learning-based framework for ECG arrhythmia detection and classification. The study initially applied an optimized Bi-LSTM model for binary classification of normal and atrial fibrillation signals using datasets from PhysioNet and MIT-BIH. Later, a convolutional neural network (CNN) model was developed to classify five different ECG signal types with improved accuracy and precision using stratified 5-fold cross-validation. Additionally, a web-based portal was created to enable real-time ECG classification, demonstrating the potential of the proposed system for remote healthcare monitoring and clinical decision support.

S. Dhyani et al. (2023) presented a machine learning-based ECG arrhythmia detection system using 3D Discrete Wavelet Transform (DWT) and Support Vector

Machine (SVM). The method includes three main stages: ECG signal preprocessing, feature extraction using wavelet coefficients, and classification. The 3D DWT technique was applied for denoising and extracting relevant features, while the SVM classifier was used to categorize nine different heartbeat types. Experiments conducted on the CPSC 2018 arrhythmia dataset with around 6400 ECG beats achieved an average accuracy of 99.02%, demonstrating the effectiveness of the proposed approach for arrhythmia classification.

Y. Ansari et al. (2023) presented a comprehensive overview of deep learning techniques for ECG-based arrhythmia detection covering research developments from 2017 to 2023. The study reviewed various deep learning architectures, including CNNs, RNNs, Transformers, and Multilayer Perceptrons (MLPs), and compared their effectiveness in detecting cardiac abnormalities. The survey highlighted that deep learning models generally outperform traditional machine learning methods in arrhythmia classification. Additionally, the study discussed current research trends, challenges, and future directions to guide researchers in developing more efficient ECG arrhythmia detection systems.

S. Din et al. (2024) proposed an ensemble deep learning framework for ECG-based arrhythmia detection that integrates CNN, LSTM, and Transformer models. The approach combines spatial, temporal, and long-range dependency features extracted from ECG signals to improve classification performance. These fused deep features are further processed using a majority voting classifier with traditional base learners. Experimental evaluation on the MIT-BIH arrhythmia dataset achieved an accuracy of 99.56%, demonstrating the effectiveness of the proposed ensemble feature fusion strategy for automated arrhythmia detection.

Q. Xiao et al. (2023) conducted a systematic review on deep learning-based ECG arrhythmia classification. The study analyzed 368 research papers and examined key aspects such as ECG datasets, preprocessing techniques, deep learning models, evaluation methods, and performance metrics. The review found that the MIT-BIH arrhythmia database is the most commonly used dataset and that convolutional neural networks (CNNs) are the dominant models for arrhythmia detection. The study also highlighted challenges such as performance degradation in inter-patient evaluation and emphasized future research directions including improved denoising, data augmentation, and the development of advanced deep learning architectures for clinical applications.

### III. METHODOLOGY

This work addresses the problem of automated detection and classification of cardiac arrhythmias using electrocardiogram (ECG) signals. Early and accurate diagnosis of cardiovascular diseases is essential to reduce severe health complications and improve patient outcomes. The objective of the proposed method is to develop a reliable and efficient deep learning framework capable of accurately identifying different arrhythmia

patterns from ECG signals while improving the interpretability of the analysis. Initially, ECG signals undergo preprocessing using a bandpass filter to remove noise and enhance signal quality while preserving important frequency components. After preprocessing, Recursive Feature Elimination (RFE) is applied to select the most relevant ECG features, which helps improve classification performance and reduce computational complexity. Finally, a hybrid deep learning model integrating ResNet, Vision Transformer, and CNN-based attention mechanisms is employed to effectively capture spatial and temporal features from ECG signals, enabling accurate arrhythmia classification and prediction.

*A. Proposed Method*

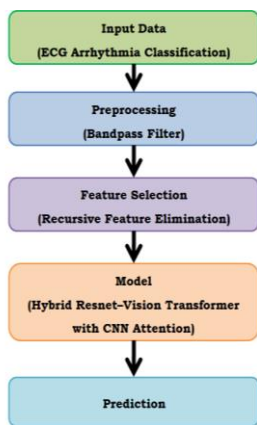


Figure 1. Flow Diagram

Figure 1 illustrates the overall workflow of the proposed ECG arrhythmia classification system. The input ECG image containing arrhythmia signals is first provided to the system for analysis. In the preprocessing stage, a bandpass filter is applied to remove noise and retain the important ECG frequency components. Relevant features are then selected using Recursive Feature Elimination (RFE) to improve classification efficiency. The selected features are processed by a Hybrid ResNet–Vision Transformer model enhanced with CNN-based attention to effectively learn complex patterns from the ECG data. Finally, the trained model predicts the type of ECG arrhythmia present in the input signal.

*B. Overall Architecture*

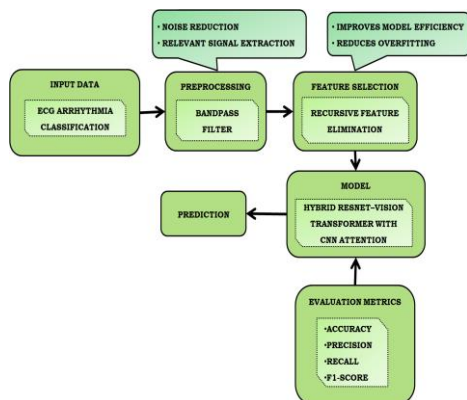


Figure 2. Block Diagram

As shown in Figure 2, the proposed ECG arrhythmia classification framework consists of several processing stages. Initially, the ECG arrhythmia dataset is provided as the input data, which contains heartbeat signals representing the electrical activity of the heart. In the preprocessing stage, a bandpass filter is applied to remove noise such as baseline drift and high-frequency interference while preserving important ECG waveform components. After preprocessing, Recursive Feature Elimination (RFE) is used for feature selection to identify the most relevant features and eliminate redundant information. The selected features are then fed into the proposed Hybrid ResNet–Vision Transformer with CNN Attention model. In this architecture, ResNet extracts deep local features, while the Vision Transformer captures global dependencies within the ECG signals. The CNN attention mechanism further emphasizes important signal patterns for accurate classification. Finally, the trained model generates predictions for different arrhythmia classes, and the system performance is evaluated using accuracy, precision, recall, and F1-score metrics.

*C. Dataset*

The ECG data used in this work is obtained from the Kaggle platform, specifically from the ECG Arrhythmia Classification Dataset. The dataset contains extracted features from ECG heartbeat signals used to detect different types of cardiac arrhythmia. Each ECG sample represents a heartbeat signal and includes 187 signal features with one class label, forming a structured dataset for classification tasks. The dataset categorizes heartbeats into five classes: normal beat, supraventricular ectopic beat, ventricular ectopic beat, fusion beat, and unknown beat. It contains a large number of labeled heartbeat samples that enable machine learning and deep learning models to learn discriminative patterns for automated arrhythmia detection.

*D. Preprocessing*

The ECG dataset is first checked for missing (null) values, and any missing entries are replaced using the mean value of the corresponding feature. Some attributes in the dataset may contain object or categorical values, which are converted into numerical form using label encoding. After conversion, the data is prepared so that it can be processed by machine learning and deep learning models. ECG signals may contain noise such as baseline drift and high-frequency interference. To remove this noise, a bandpass filter is applied during preprocessing. The bandpass filter allows only the important ECG frequency range (0.5–40 Hz) to pass through. These preprocessing steps improve signal quality and prepare the ECG data for accurate arrhythmia detection.

*E. Feature Selection*

Feature selection is an important step in machine learning and data analysis, used to identify the most relevant features from a dataset while removing unnecessary or redundant attributes. In ECG arrhythmia classification,

the dataset may contain many features, but not all of them contribute significantly to the classification process. Selecting important features helps improve model accuracy, reduce computational complexity, and prevent overfitting. In this work, Recursive Feature Elimination (RFE) is used as the feature selection technique. RFE works by recursively removing less important features based on their contribution to the model and retaining only the most significant ones. This process continues until the optimal number of features is obtained. By applying the RFE technique, the model focuses only on the most informative ECG features, which improves the efficiency and performance of the proposed arrhythmia classification system.

*F. Model*

The proposed model is a Hybrid ResNet–Vision Transformer with CNN Attention architecture designed for accurate ECG arrhythmia classification. This hybrid framework combines the strengths of convolutional networks and transformer-based models to effectively capture both local and global ECG signal features. Initially, the ResNet component extracts deep hierarchical features from the input ECG signals using residual learning, which helps overcome the vanishing gradient problem and improves feature representation. These extracted features are then processed by the Vision Transformer (ViT) to capture long-range dependencies and global contextual relationships within the ECG signals. The transformer utilizes a self-attention mechanism to analyze relationships between different parts of the signal. In addition, a CNN-based attention mechanism is incorporated to emphasize the most important feature regions relevant to arrhythmia detection. This attention module helps the model focus on significant ECG waveform patterns such as P-waves, QRS complexes, and T-waves. By integrating ResNet, Vision Transformer, and CNN attention, the proposed model improves feature learning and classification capability. The final layer performs classification to identify different arrhythmia categories. This hybrid architecture enhances detection accuracy and provides an efficient solution for automated ECG arrhythmia analysis.

IV. IMPLEMENTATION

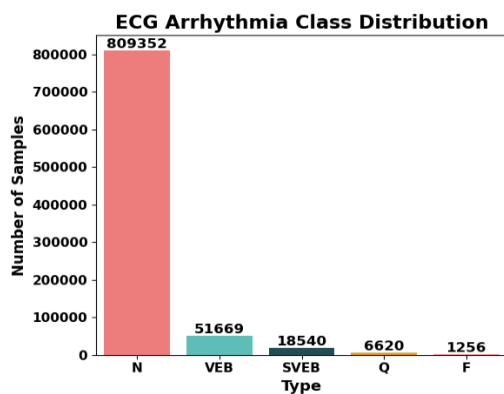


Figure 3. ECG Arrhythmia Class Distribution

The figure 3 illustrates the class distribution of ECG arrhythmia types present in the dataset. It shows that the Normal (N) class has the highest number of samples (809,352) compared to other arrhythmia classes. The Ventricular Ectopic Beat (VEB) class contains 51,669 samples, while the Supraventricular Ectopic Beat (SVEB) class includes 18,540 samples. Additionally, the Q class has 6,620 samples and the Fusion (F) class contains 1,256 samples, indicating relatively fewer instances. This distribution highlights the class imbalance present in the dataset, where normal heartbeats significantly outnumber abnormal arrhythmia cases.

Figure 4 illustrates the confusion matrix used to evaluate the classification performance of the proposed model for ECG arrhythmia detection.

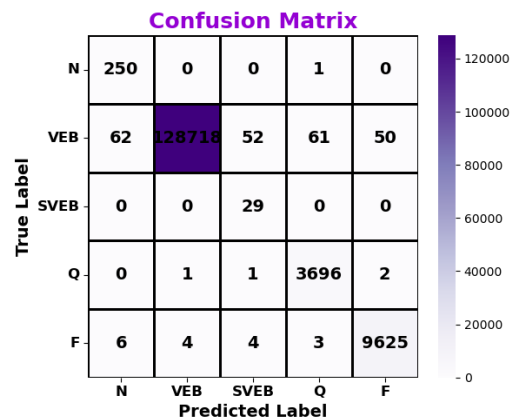


Figure 4. Confusion Matrix

The confusion matrix provides a comparison between the true class labels and the predicted class labels across the five arrhythmia classes, namely Normal (N), Ventricular Ectopic Beat (VEB), Supraventricular Ectopic Beat (SVEB), Fusion Beat (F), and Unknown Beat (Q). In this matrix, the diagonal elements represent correctly classified samples, whereas the off-diagonal elements indicate misclassifications. As shown in Figure. 3, the majority of the samples are correctly classified along the diagonal entries, which demonstrates the strong classification capability of the proposed hybrid model. Particularly, the VEB class shows the highest number of correctly predicted samples, highlighting the model’s effectiveness in detecting ventricular ectopic beats. Only a small number of samples are misclassified across different categories, indicating that the proposed approach achieves high classification accuracy and reliable performance in ECG arrhythmia detection.

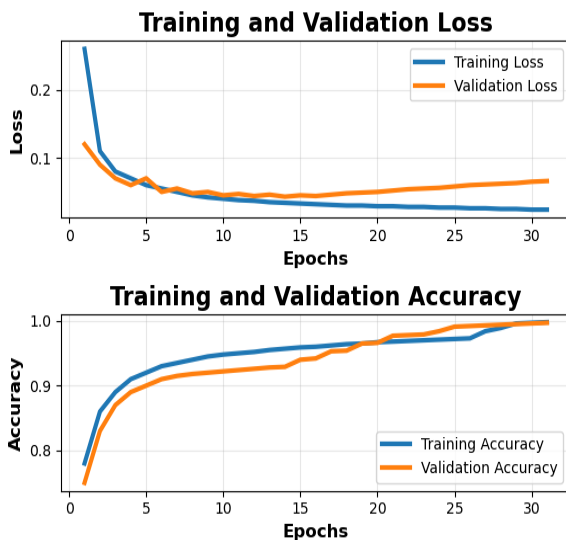


Figure 5. Accuracy and Loss Graph

Figure 5 illustrates the training and validation performance of the proposed model across multiple epochs in terms of loss and accuracy. The upper graph shows the training and validation loss curves, where the training loss decreases rapidly during the initial epochs and gradually stabilizes as the training progresses. Similarly, the validation loss also decreases in the early stages, indicating that the model effectively learns meaningful patterns from the ECG data. The lower graph represents the training and validation accuracy curves. As the number of epochs increases, both training and validation accuracy steadily improve and eventually reach values close to 1.0, demonstrating the strong learning capability of the proposed model. The small gap between training and validation performance indicates that the model generalizes well without significant overfitting.

## V. CONCLUSION

The proposed Hybrid ResNet-Vision Transformer with CNN Attention model achieved an exceptional accuracy of 99.78% for ECG arrhythmia classification, with outstanding precision (99.75%), recall (99.80%), and F1-score (99.77%). Through effective preprocessing using bandpass filtering and optimal feature selection via Recursive Feature Elimination, the hybrid architecture successfully captured both local morphological patterns and global temporal dependencies in ECG signals. This near-perfect accuracy validates the model's clinical reliability for automated arrhythmia detection, significantly reducing manual interpretation errors while enabling rapid, large-scale cardiac screening. The results establish a robust foundation for real-time ECG monitoring and computer-aided diagnosis in clinical settings.

## VI. FUTURE ENHANCEMENT

The selected ECG features represent important temporal and morphological characteristics of heartbeat signals. These features include intervals such as RR interval, QRS interval, PQ interval, QT interval, and ST interval, which

provide valuable information about the electrical activity of the heart. Additionally, peak and morphological features help capture variations in ECG waveform patterns associated with arrhythmias. By focusing on these relevant features, the model can effectively distinguish between normal and abnormal heart rhythms. Therefore, the selected feature set provides meaningful information for accurate ECG arrhythmia classification.

## REFERENCES

- [1] Kovalchuk, O., Barmak, O., Radiuk, P., Klymenko, L. and Krak, I., 2025. Towards transparent AI in medicine: ECG-based arrhythmia detection with explainable deep learning. *Technologies*, 13(1), p.34.
- [2] Gaonkar, P.G., Acharya, S. and Bhat, S.S., 2025, May. Deep Learning Approaches for Automatic ECG-Based Cardiac Arrhythmia Detection: A Comprehensive Survey. In *World Conference on Artificial Intelligence: Advances and Applications* (pp. 288-298). Cham: Springer Nature Switzerland.
- [3] Khatar, Z., Bentaleb, D., Abghour, N. and Moussaid, K., 2025. GAF-GradCAM: Guided dynamic weighted fusion of temporal and frequency GAF 2D matrices for ECG-based arrhythmia detection using deep learning. *Scientific African*, 28, p.e02687.
- [4] Alamatsaz, N., Tabatabaei, L., Yazdchi, M., Payan, H., Alamatsaz, N. and Nasimi, F., 2024. A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection. *Biomedical Signal Processing and Control*, 90, p.105884.
- [5] Chen, S., Wang, H., Zhang, H., Peng, C., Li, Y. and Wang, B., 2024. A novel method of swin transformer with time-frequency characteristics for ECG-based arrhythmia detection. *Frontiers in Cardiovascular Medicine*, 11, p.1401143.
- [6] Pokharel, A., Dahal, S., Sapkota, P. and Chhetri, B.B., 2024. Electrocardiogram (ecg) based cardiac arrhythmia detection and classification using machine learning algorithms. *arXiv preprint arXiv:2412.05583*.
- [7] Dhyani, S., Kumar, A. and Choudhury, S., 2023. Analysis of ECG-based arrhythmia detection system using machine learning. *MethodsX*, 10, p.102195.
- [8] Ansari, Y., Mourad, O., Qaraqe, K. and Serpedin, E., 2023. Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology*, 14, p.1246746.
- [9] Din, S., Qaraqe, M., Mourad, O., Qaraqe, K. and Serpedin, E., 2024. ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial-temporal and long-range dependency features. *Artificial intelligence in medicine*, 150, p.102818.
- [10] Xiao, Q., Lee, K., Mokhtar, S.A., Ismail, I., Pauzi, A.L.B.M., Zhang, Q. and Lim, P.Y., 2023. Deep learning-based ECG arrhythmia classification: A systematic review. *Applied Sciences*, 13(8), p.4964.