Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 7, No. 2, June, 2026

Deep Residual Transfer Learning and Transformer Architectures for ECG Signal–Based Heart Disease Detection

Nawaf Alshdaifat [1], Hamza Abu Owida [2], Hamza A. Mashagba [3], Suhaila Abuowaida [4], Azlan Abd Aziz [3*], Esraa Abu Elsoud [5], Mardeni Roslee [6], Mohamad Yusoff Alias [6]

[1] *Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan.*

[2] *Medical Engineering Department, Faculty of Engineering, Al-Ahliyya Amman University, Amman 19328, Jordan.*

[3] *Faculty of Engineering and Technology, Centre for Wireless Technology (CWT), Multimedia University, Melaka, 75450, Malaysia.*

[4] *Department of Data Science and Artificial Intelligence, Faculty of Prince Al-Hussein Bin Abdallah II for IT, Al al-Bayt University, Mafraq, Jordan.*

[5] *Cybersecurity and Cloud Computing Department, Faculty of Information Technology, Applied Science Private University, Amman 11931, Jordan.*

[6] *Faculty of Engineering and Technology, Multimedia University, Melaka 75450, Malaysia.*

Received 08 February 2026; Revised 03 May 2026; Accepted 08 May 2026; Published 01 June 2026

Abstract

The purpose of this research was to create a fully automatic deep-learning method to accurately classify different types of cardiac arrhythmias based on patterns observed in electrocardiograms (ECGs). There are several challenges that arise when manually interpreting cardiac arrhythmias. These include variability in how different observers interpret arrhythmias and variability in the diagnostic results obtained by different clinicians. To address these issues, we propose an architecture that includes both residual transfer learning and transformer encoder blocks. In our proposed architecture, the residual learning block uses Conv1D layers with skip connections to learn hierarchical representations of features in the ECG signal. The transformer block uses multi-head self-attention to identify longer-range dependencies in the ECG sequence. Our proposed model is tested on two publicly available benchmark databases of ECG recordings, namely the MIT-BIH Arrhythmia Database and the PTBDB Diagnostic ECG Database. We evaluate the performance of our model using a stratified 10-fold cross-validation procedure as well as Receiver Operating Characteristic (ROC) analysis. The proposed model achieved a classification accuracy of 99% on both datasets, with precision scores of 0.88 to 0.99 and recall scores of 0.87 to 0.99 for each arrhythmia class. Furthermore, the AUC values of the ROC analysis ranged from 0.98 to 1.0, indicating high levels of discrimination against minority classes, including supraventricular ectopic beats and fusion beats.

Keywords: ECG Signal Classification; Cardiac Arrhythmia Detection; Residual Transfer Learning; Transformer Networks; Deep Learning.

1. Introduction

The major global public health concern of cardiovascular disease accounts for nearly 1/3 of all global deaths and also causes a variety of debilitating diseases globally [1]. In addition, the World Health Organization has estimated that cardiovascular diseases account for roughly 31% of all deaths occurring globally [2]. As a result, it is clear that there exists a pressing need to identify and implement effective prevention and management strategies to deal with this disease burden. One type of cardiovascular disease that is common and dangerous is cardiac arrhythmia [3]. A bradycardic

* Corresponding author: azlan.abdaziz@mmu.edu.my

https://doi.org/10.28991/HIJ-2026-07-02-012

rhythm occurs when a cardiac arrhythmia results in a heart rate less than 60 beats per minute [4]. Bradycardia occurs in fewer than 80% of patients diagnosed with cardiovascular disease, significantly increases the patient's risk for serious complications (such as stroke or sudden cardiac death), and is often fatal if left undiagnosed and untreated [5, 6]. Prompt and accurate assessment of cardiac arrhythmias is necessary because they have been shown to increase the risk of potentially deadly events, including myocardial infarctions and cerebrovascular accidents [7, 8].

The cost of treating cardiovascular-related illness may be so prohibitive that it limits patient access to appropriate medical care for many patients around the world Waveform patterns produced by the electrical. The primary method used for diagnosing and classifying cardiac arrhythmias is through the use of arrhythmia classification systems. Cardiac arrhythmia classification is made possible with these systems using characteristics of ECG signal waveforms produced by the heart from the ECG. The ECG captures characteristic waveform patterns (P, QRS, T) generated by the heart's electrical activity across all 12 leads, providing a comprehensive view of cardiac function in both horizontal and vertical planes. Therefore, due to their non-invasive nature, low cost, and ease of performance, ECGs serve as a major tool in clinical medicine for the diagnosis of a variety of cardiovascular diseases [9, 10].

Manual evaluation of an electrocardiogram (ECG), performed by a cardiologist, is time consuming and also subjective in nature. A substantial amount of time for a cardiologist to evaluate each patient's ECG data exists [4]. There is also variability and complexity involved in identifying arrhythmia pattern on ECG data. Subjective evaluation by a cardiologist will provide a potential for bias when evaluating patients. Since there are many forms of arrhythmia with various treatments, it is important to have a correct diagnosis of the type of arrhythmia that occurred. Therefore, relying solely on the physician's visual evaluation of the ECG recording may not provide consistent diagnostic results [11–14]. The main contribution of the paper is to overcome these obstacles by developing an automated ECG analysis system based on established diagnostic criteria to improve the detection of arrhythmias [15]. Recently, deep learning (DL) has become a promising technique for the automated classification of cardiac arrhythmias. DL methods can automatically find and learn important features from large datasets [16, 17], making it possible to tell the difference between different types of arrhythmias with great precision [18]. In addition, DL can perform exceptionally well with noisy, redundant, and large data, making it very suitable for real-world clinical applications. Although recent deep learning approaches have shown promising results, existing models still face challenges, including extensive pre-processing requirements, limited capability to capture long-range dependencies, and inconsistent performance across minority arrhythmia classes.

This study addresses these gaps by proposing a hybrid model that combines residual transfer learning with transformer architectures. This study begins with a brief description of the current realities in the management of CVD and the concept of ECG for the identification of cardiac arrhythmias. In addition, there is a discussion about the problems that arise when ECGs are interpreted manually, as well as a discussion about previously proposed automated methods, leading to the proposed approach of the present work. Exploration of these aspects is done in the Related Work section and carries the message of the necessity for better and more precise automated tools. On this basis, the proposed method section presents the model's development based on DL methods for the classification of arrhythmias. This paper discusses how to prepare the dataset, feature engineering, and a full description of the DL model. It also discusses how to use existing experience and implement validation strategies to make the model more stable and useful.

The section on experimental setup, results, and discussion gives an overview of what was found when the proposed model was used to classify ECG signals into groups. The section also provides details on the precision, sensitivity, and specificity of the proposed model. Also provide a brief comparison with the other approaches to show the efficiency of the proposed approach. The purpose of the Conclusion is to summarize the major results from this proposed research and illustrate how the use of DL for automated arrhythmia detection is a means of improving the prognosis of the patient and the effectiveness of healthcare services. Lastly, this research discusses possible future research directions for the problem being looked at, specifically how to use multimodal data in the model's work and make it even better. Such an all-encompassing framework ensures the thoroughness of the topic's analysis and significantly contributes to the ongoing processes of improving cardiovascular diagnostics.

2. Related Work

Classification of heartbeat arrhythmias using DL techniques has become increasingly important in clinical practice. When ECG signals are processed using signal processing techniques, it is easier to find and classify cardiac arrhythmias. This approach can help identify and treat many cardiovascular diseases early on. Table 1 shows a summary of the relevant research works discussed below.

Deep learning (DL) has demonstrated excellent performance in many areas of medical research [19], particularly in diagnosing different types of diseases using medical images. Researchers have applied various DL methods for heartbeat classification and arrhythmia detection [20, 21]. These approaches exploit the pattern recognition capability of deep neural networks to identify complex and nonlinear relationships in ECG data through automatic feature learning, thereby reducing the need for manual feature engineering [22].

In Hassan et al. [23], a CNN combined with a bidirectional long short-term memory (Bi-LSTM) model was developed for arrhythmia classification. To capture temporal variations in ECG signals, the authors employed several convolutional layers followed by a Bi-LSTM network. Experimental results showed that the proposed model achieved an accuracy of 95.8% on the MIT-BIH dataset. However, the approach required extensive ECG signal preprocessing and demonstrated limited performance for rare arrhythmia classes.

In another study, researchers proposed a multimodal DL-based approach for ECG classification [24]. Different types of arrhythmias were classified using individual DL bagging models based on heartbeat analysis. Their best-performing model combined a CNN with an LSTM network to extract both spatial and temporal features from ECG data. Using a patient-independent evaluation scheme, they achieved an overall accuracy of 95.81%. Nevertheless, the model was computationally complex and required substantial computational resources, which may limit its clinical applicability.

Several additional methods have also incorporated LSTM into network architectures. For example, Liu et al. [25] proposed a unique network layer configuration using LSTM within an autoencoder-style framework. Combined with their proposed ECG preprocessing method, this structure improved arrhythmia classification accuracy. Their approach enabled direct incorporation of ECG signals into the model without requiring sophisticated preprocessing or manual feature extraction. The model achieved an overall classification accuracy of approximately 98.6%. However, its performance varied across different arrhythmia types, particularly in detecting supraventricular ectopic beats (SVEB), where classification accuracy decreased significantly.

Wang et al. [26] proposed a CNN-based fusion method for ECG signals in both the time and frequency domains. They first applied multi-scale wavelet decomposition to denoise the ECG signal and then segmented each cardiac cycle by detecting the R-wave. A fast Fourier transform was subsequently used to extract frequency-domain information for each heartbeat cycle. The classifier was a neural network that utilized both temporal and frequency-domain data for classification. Their experimental results achieved an identification accuracy of 95.4%. However, a major limitation of this approach was its dependence on accurate R-peak detection, which is often unreliable in noisy clinical environments.

Dong et al. [27] introduced an interpretable feature for arrhythmia classification called heartbeat dynamics. Their study employed an algorithm to monitor changes in heartbeat morphology as well as subtle variations in pulse rate. Using three traditional classification methods, namely nearest neighbor, decision tree (random forest), and support vector machine, they achieved a classification accuracy of 96.5% on the MIT-BIH database. Although this technique provided interpretable results, it was unable to effectively model long-range dependencies, which are important for certain types of arrhythmias and are inherent in many ECG signals.

Kaniraja & Mishra [28] applied deep learning techniques for ECG analysis. Their system diagnosed abnormal heart rhythms with an accuracy of 92%. However, the method struggled to differentiate between similar arrhythmia patterns. Similarly, Chen & Chen [29] used recurrent neural networks (RNNs) for arrhythmia classification by modeling temporal variations in sequential ECG data over time. Their analysis showed that RNNs alone were capable of classifying more arrhythmia categories than simple CNNs and could capture long-term temporal dependencies with an accuracy of up to 91%.

Other important studies include the work of Kachuee et al. [30], who introduced a deep residual CNN approach that achieved 93% accuracy and recall. Their architecture employed skip connections to address the vanishing gradient problem in deep networks. However, their model required fixed-size inputs, necessitating additional preprocessing steps that could introduce artifacts into the ECG signal. Xu et al. [31] combined CNN and Bi-LSTM architectures and achieved 96% accuracy and 92% recall. Their hybrid approach leveraged the spatial feature extraction capability of CNNs together with the temporal modeling capability of Bi-LSTMs. Nevertheless, their model showed limited performance for fusion beats and supraventricular ectopic beats, which are clinically important arrhythmia types.

Peimankar et al. [32] achieved 97% accuracy in arrhythmia classification using a residual transfer learning model. Their approach utilized pre-trained models to improve performance when only limited training data were available. Although the model demonstrated strong classification capability, it required extensive hyperparameter tuning and was highly sensitive to signal quality. In another study, Guo et al. [33] developed a DenseNet-based model incorporating both attention mechanisms and gated recurrent units (GRUs), achieving accuracy and recall values of 92% and 82%, respectively. However, the model underperformed in classifying arrhythmias with very limited samples, highlighting the common issue of class imbalance encountered in many traditional approaches.

In addition, Xia et al. [34] developed a novel method for arrhythmia classification. Specifically, they proposed a Transformer and Convolution-based Generative Adversarial Network (TCGAN) capable of leveraging the Transformer's powerful ability to learn relationships among sequence elements. They performed inter-patient heartbeat classification and achieved an overall classification accuracy of 97%. In Serhani et al. [35], the researchers proposed a reinforcement learning-based method to automatically optimize CNN hyperparameters for arrhythmia classification, achieving an accuracy of 97.4%.

Recent studies have significantly expanded the application of deep learning for cardiac signal analysis. Karthikeyani et al. [36] evaluated the impact of different dimensionality reduction techniques and classification algorithms for cardiac disease diagnosis using deep learning classifiers on ECG signals. Their findings showed that the dimensionality reduction method strongly influenced classification performance. Zhao et al. [37] developed ECG-Chat, a multimodal ECG-language model designed for cardiac disease diagnosis. Their results demonstrated that integrating a language model with ECG signal analysis can improve clinicians' ability to interpret diagnostic results. Abdullayev et al. [38] proposed a multimodal AI system for the early detection and prognosis of ischemic heart disease and demonstrated that combining multiple data modalities yields higher predictive accuracy than relying on a single data source.

Meeran & Munaf [39] compared unidirectional and bidirectional recurrent neural networks (RNNs) for ECG arrhythmia detection using augmented MIT-BIH data, further confirming the advantages of bidirectional RNNs in modeling temporal dependencies within ECG signals. In addition, Akbar & Utami [40] conducted a systematic review of studies utilizing the MIT-BIH database and emphasized the continuing relevance of this dataset while also noting the need for new classification techniques capable of handling the diverse signal characteristics found in ECG data. However, none of the existing approaches combined residual learning with Transformer architecture to simultaneously address both the vanishing gradient problem and the long-range dependency problem.

Extensive research indicates that, despite significant advances in deep learning for arrhythmia classification, there remains a need for models capable of reducing preprocessing complexity, effectively capturing long-range dependencies, and maintaining stable performance across diverse arrhythmia types. Although the reviewed methods have achieved important improvements, many still face limitations in practical clinical applications due to high computational requirements, inconsistent performance across arrhythmia categories, or dependence on extensive preprocessing.

This study addresses these limitations by proposing a novel model that combines the advantages of residual learning for gradient optimization with Transformer architectures for modeling long-range dependencies. The following section presents the innovative design of the proposed model, which achieves state-of-the-art performance while minimizing preprocessing requirements and maintaining robustness across different arrhythmia categories. This model represents a significant step forward in the development of clinically applicable automated arrhythmia detection systems capable of improving the accuracy and efficiency of cardiac diagnosis.

Table 1. Comparison with Existing Literature

Author	Model	Key Limitations
Guo et al. [33]	DenseNet-GRU	Poor minority class performance
Kachuee et al. [30]	Deep residual CNN	Fixed-length input requirement
Xu et al. [31]	CNN + Bi-LSTM	Suboptimal for fusion beats
Hassan et al. [23]	CNN-Bi-LSTM	Extensive pre-processing needed
Essa et al. [24]	Multi-model bagging	High computational demands
Dong et al. [27]	Heartbeat dynamics	Limited long-range dependency capture
Wang et al. [26]	Time-frequency CNN	R-peak detection dependency
Liu et al. [25]	LSTM auto encoder	Inconsistent across arrhythmia types
Peimankar et al. [32]	Transfer Learning	Extensive hyperparameter tuning
Xia et al. [34]	TCGAN	Synthetic data may introduce artifacts
Serhani et al. [35]	RL-optimized CNN	While efficient, achieves lower accuracy for minority arrhythmia classes
Karthikeyani et al. [36]	DL with dimensionality reduction	Performance depends heavily on feature reduction strategy
Zhao et al. [37]	ECG-Chat (ECG-language model)	High computational cost due to large language model integration
Abdullayev et al. [38]	Multimodal AI framework	Requires multiple data modalities not always available clinically
Meeran & Munaf [39]	Bidirectional RNN	Limited spatial feature extraction capability

3. Transfer Residual Learning and Transformer Model

For clarity and reproducibility, we establish a unified notation system that will be used consistently throughout the remainder of this paper. Table 2 summarizes all mathematical symbols and their corresponding definitions.

Table 2. Notation and Symbols.

Symbol	Definition
X	Raw ECG signal
F	Feature map
Q, K, V	Query, Key, Value matrices
d_k	Key dimension
d_model	Model dimension
W	Weight matrix
b	Bias vector
λ	Regularization coefficient
γ, β	Learnable normalization parameters
μ	Mean
σ^2	Variance
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

Early detection of heart arrhythmias through accurate recognition and classification of ECG signals remains critically important. Recent advances in deep learning approaches, such as CNNs and RNNs, have demonstrated strong potential because these architectures are capable of learning powerful representations from ECG signals [23, 24]. In this research, the feature extractor is initialized using pretrained weights that were previously trained on the large-scale 1D PTB-XL ECG dataset [41]. Instead of randomly initializing the backbone network, the final classification layer is replaced with the target arrhythmia class labels, and the model is then fine-tuned on the target datasets. Consequently, the proposed model is not trained entirely from scratch, as it utilizes pretrained weights derived from the PTB-XL dataset.

During the early stages of fine-tuning, the initial layers are frozen and are not updated as extensively as the later layers. These layers are gradually unfrozen during training to allow the model to adapt fully to the target task. This strategy enables the fine-tuning process to adjust the pretrained model to features specific to arrhythmia detection. Residual learning further enhances the framework by improving gradient flow and maintaining training stability, which supports the accurate identification of ECG patterns.

In addition, the model incorporates Transformer architectures, which have demonstrated remarkable effectiveness in tasks such as natural language processing and can also be applied successfully to ECG signal classification. Since Transformer models are based on attention mechanisms, they are capable of focusing on the most important parts of the input signal, thereby improving arrhythmia classification performance.

In this paper, a novel framework combining residual transfer learning with the long-sequence dependency modeling capability of Transformer networks is proposed for accurate classification of different arrhythmia types. Heartbeat signals contain numerous subtle indicators that are essential for arrhythmia detection, making them highly valuable in clinical diagnostics. This section describes the proposed framework for arrhythmia disease detection. As shown in Figure 1, the proposed method begins by receiving an ECG signal as input, which is subsequently processed through several feature extraction stages using a residual transfer learning model. This preprocessing stage is important for enabling the model to learn meaningful representations from raw ECG data.

The extracted features are then passed through a Transformer model capable of capturing long-range dependencies within the signal. To the best of our knowledge, this represents a novel combination of residual transfer learning and Transformer-based modeling. The final stage consists of a deep artificial neural network (ANN) with a softmax activation function for classification.

Using the extracted feature representations, the proposed ANN architecture classifies the input ECG signals into different arrhythmia categories based on their signal characteristics and temporal behavior. The softmax function enables the model to generate probability distributions for each arrhythmia class. By combining advanced machine learning techniques with domain knowledge in ECG signal processing, the proposed framework achieves highly accurate arrhythmia detection and classification for healthcare applications.

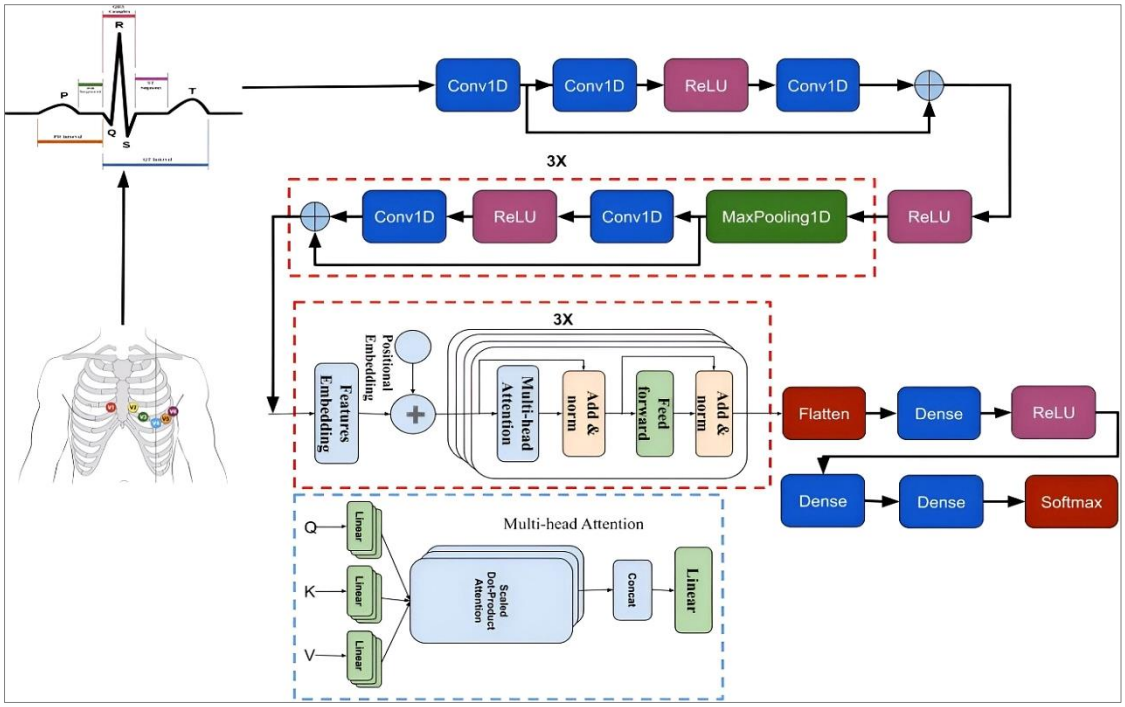

Figure 1 illustrates the proposed Hybrid Residual Transfer Learning and Transformer model architecture for arrhythmia classification from ECG signals. The details of this framework are as follows: (1) The model begins with initial feature extraction using Conv1D layers (blue) combined with ReLU activation functions (red). (2) Residual learning blocks with skip connections are employed to mitigate the vanishing gradient problem. (3) Transformer encoding with multi-head attention mechanisms is incorporated to capture long-range dependencies within the ECG signals. The input ECG signal undergoes hierarchical feature extraction followed by contextual modeling before final classification is performed through dense layers and a softmax activation function.

The model was developed using the Keras 2.14.0 framework with TensorFlow as the backend. All experiments were conducted on the Kaggle platform using an NVIDIA Tesla P100 GPU for computational processing. Python 3.8 and its associated libraries, including NumPy for numerical operations, SciPy for signal processing, and Pandas for structured data management, were utilized for data preprocessing and feature extraction.

In this research, the model architecture was implemented using the Keras Functional API, which enabled flexible implementation of the proposed custom Transformer layers and residual connections. Experimental trials were conducted to determine the optimal training hyperparameters, and the Adamax optimizer was selected for weight updates due to its stability when handling ECG signal patterns. Model performance and ROC curves were computed using the Scikit-learn metrics package. Evaluation metrics, including accuracy, precision, recall, and F1-score, were calculated using custom validation functions implemented within the TensorFlow environment. This approach ensured reproducibility by maintaining consistency throughout the training process. It also provided sufficient flexibility to develop the proposed hybrid architecture that combines residual learning and Transformer components for arrhythmia classification.

3.1. Architectural Novelty and Distinctions

Several recent studies have demonstrated the application of hybrid CNN–Transformer architectures for ECG analysis [31, 34]. However, the proposed model differs from previous approaches due to several architectural innovations. The first innovation is that, unlike most hybrid models that simply concatenate CNN and Transformer outputs [31], the proposed architecture employs a Hierarchical Progressive Feature Refinement strategy. Residual blocks are used to generate temporal features at multiple scales, which are then semantically enriched using stacked Transformer encoder modules. This design enables local morphological patterns, such as P-waves, QRS complexes, and T-waves captured by Conv1D layers, to be progressively abstracted into global contextual representations without information loss.

Although there is a growing body of research on hybrid CNN–Transformer architectures for ECG analysis [31, 34], this study introduces several significant architectural contributions that distinguish the proposed model from prior work. In contrast to conventional hybrid architectures that merely combine CNN and Transformer outputs [31], the proposed framework provides hierarchical progressive feature enhancement through residual blocks that generate multi-scale temporal features. These features are subsequently enriched semantically using stacked Transformer encoder modules. As a result, the 1D convolutional layers effectively capture local morphological characteristics, including P-waves, QRS

complexes, and T-waves, while the Transformer modules convert these local features into comprehensive global contextual representations while preserving essential information.

Fourth, our approach to integrating Transformers is to apply them after extracting semantically meaningful mid-level representations through hierarchical residual feature extraction, rather than placing them in parallel with or prior to this process. Consequently, the self-attention mechanisms operate on refined feature representations instead of raw signals or low-level representations.

Lastly, the proposed method combines Transformers by positioning them after hierarchical residual feature extraction, which generates semantically meaningful mid-level representations of the input. This design allows the self-attention mechanisms to operate on these enriched representations rather than directly on the raw signal or low-level features.

3.2. Extracting Features using Residual Transfer Learning Model

Residual learning models help mitigate the vanishing gradient problem that arises during weight updates in deep neural networks [42]. Vanishing gradients can hinder gradient descent by reducing many of the earlier gradients in the network to extremely small or zero values, making it difficult to update layers located deeper within the architecture. In the proposed framework, skip connections address this issue by enabling stronger loss gradients to propagate backward through the network during training. These skip connections therefore serve as an effective architectural solution for preventing gradient degradation in deep networks.

- Each residual block has two Conv1D layers with 64 filters;
- Each Conv1D layer uses a kernel size of 5;
- "Same" padding is used to preserve temporal dimensions;
- Batch normalization is applied after convolution to stabilize learning;
- ReLU activation follows to add non-linearity and prevent vanishing gradients.

The skip connection method directly adds the input tensor to the second Conv1D layer's output:

$$y = F(x) + x \tag{1}$$

In case of a mismatch in input and output dimensions, the proposed work applies a 1x1 convolution on the shortcut path, as shown in Figure 2.

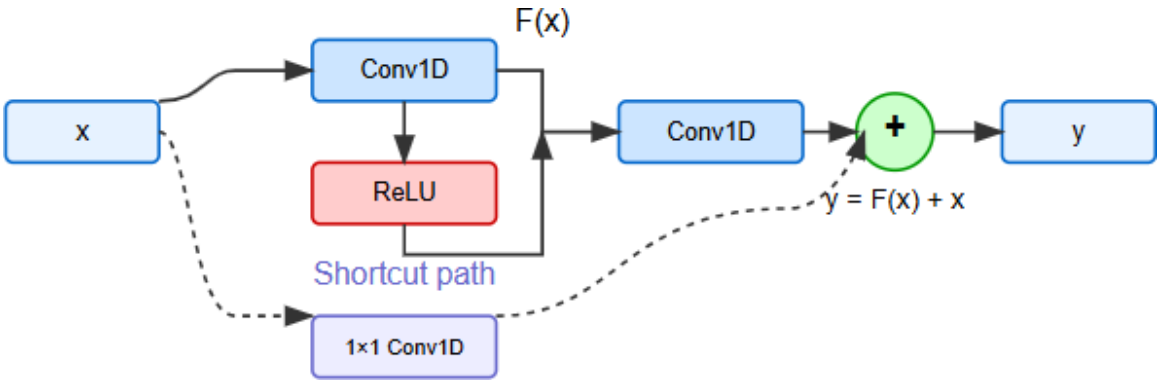

Figure 2. The proposed model implements skip connections in its residual learning blocks. When input and output dimensions don't match, a 1×1 convolution is applied on the shortcut path to ensure dimensional compatibility for the addition operation

The result provides an alternate route for the gradients to flow during backpropagation, helping alleviate the problem of gradient disappearance. To mitigate overfitting, the proposed work uses L2 regularization with a coefficient of 0.001 on all convolutional layers. This step is important because the shapes of the ECG differ greatly from one patient to another. The proposed work replicates the residual blocks three times, using the same output of one block as the input of the next. The hierarchical representation of features works in the following way: The hierarchical feature representation works as follows:

- Lower layers capture local waveform characteristics (P waves, QRS complexes, T waves);
- Upper layers abstract higher-level patterns critical for arrhythmia classification.

Following the three residual blocks, a Max Pooling 1D layer, which has a pool size of 2, reduces the temporal dimension at the cost of losing some less important features. The proposed work prepares a signal representation for use

by the transformer. A skip connection is added through a simple addition operation that takes the output from a given layer and adds it to layers, including later Conv1D and ReLU layers. The proposed work applies it three times after every ReLU operation. When the proposed work applies this kind of pattern, the loss value will backpropagate to gradients on fewer layers that prevent degrading performance. During optimization, the proposed work implements the L2 loss function.

- Mean-square error as the primary component;
- L2 normalization factor as the secondary component to penalize large weight values.

In terms of data augmentation, the proposed work decided not to use it for the management of heartbeat rhythms. Several potential drawbacks are the driving force behind this choice:

- Excessive data augmentation can lead to overfitting, especially in cases where the augmented data is very similar to the truly original, which may impede generalization.
- Using data augmentation can sometimes lead to inclusion, which makes it challenging to maintain the accuracy of the enhanced dataset due to unattainable or linguistically contradictory data.
- Augmented data can sometimes have variations in quality. Some transformations used in the augmentation process can decrease the dependability of the supplemented samples by inadvertently altering the original data.

The core component of the proposed model's design is an adaptive mechanism to align dimensionalities via 1×1 convolutions in skip connections. In cases where there are dimensional differences between the input tensor R and the output feature map F from each residual block, a dynamic 1x1 convolution will be added to the shortcut to allow both tensors to have compatible dimensions. The adaptive method is what enables the model to accommodate different dimensionalities of feature maps throughout various residual blocks and to avoid modifying the architecture based upon the configuration of the input data, as with other fixed-dimensional models, which would require architects to modify the design manually in order to support different input configurations. Although the model contains a fixed number of residual blocks (three) and transformer encoder blocks (three), which were determined as providing the best results when tested with the MIT-BIH and PTBDB datasets (188 sample ECGs), the 1x1 convolutional layer also allows for the adaptation of dimensionalities at runtime during both forward and backward passes, as shown in Figure 2.

3.3. Transformer Layers

The proposed model incorporates transformer layers as a fundamental component of its architecture, as shown in Figure 3.

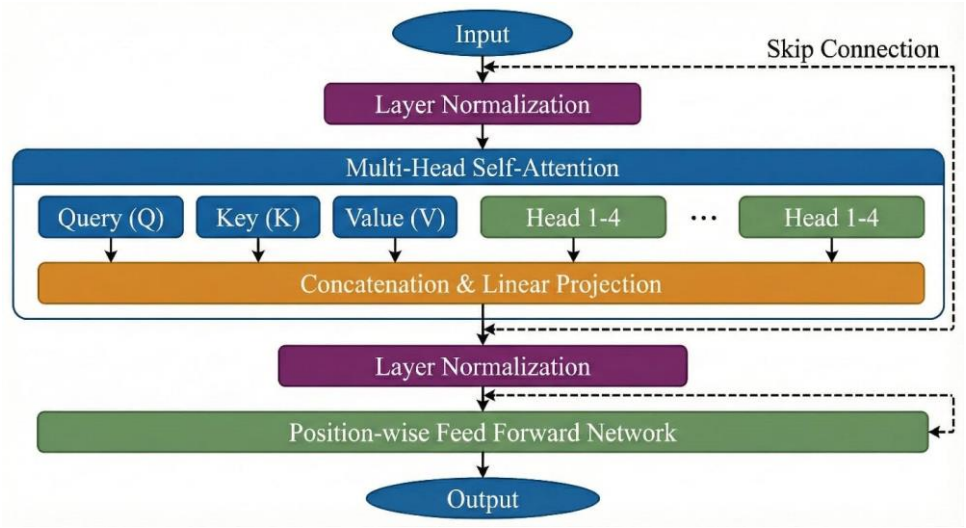

Figure 3. The Architecture of the Transformer Encoder Block with Multi-Head Self-Attention Mechanism for ECG Signal Processing

The layers that create these results demonstrate how important are individual elements in the context of a long-range input feature sequence. These layers enable capturing the relationships between the data and the long-range dependency, thus enabling the model to learn multiple different representations of the input. Each of the designed attention heads can focus on unique aspects of the input which will allow for extracting a larger number of features than before.

The transformer part of the proposed architecture builds on the foundations established by the residual blocks. Its main goal is to find long-range relationships in ECG signals that are necessary to correctly classify arrhythmias. The

proposed work uses a set of three identical transformer encoder blocks, each with multi-head self-attention mechanisms. The self- attention function uses the sequence's query (Q), key (K), and value (V) projections to figure out attention weights as:

$$\text{Attention}\ (Q, K, V) = \ \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where, dk is the number of dimensions of the key vectors.

The proposed work has eight attention heads and each can handle eight dimensions. This arrangement allows the model to look at data from several representation subfields at the same time. Using this multi-head method, the model can detect multiple parts of the ECG signal at the same time, such as changes from beat to beat, morphological oddities, and rhythm issues.

There are two parts to each transformer block: a fully connected feedforward network in position with a hidden dimension of 256 units and a multihead self-attention mechanism with two 256 units. Residual connections and Layer Normalization when added to each of the sub-layers enable the following:

$$\text{LayerNorm}(x\ +\ \text{Sublayer}(x)) \tag{3}$$

This lets gradient flow happen and keeps training stable. The feed-forward network changes the features on its own at each point using two dense layers with ReLU activation between. This approach allows the attended features to go through complex non-linear transformations. Unlike other convolutional methods that are limited by the size of the receptive field, the proposed transformer architecture can find dependencies between any position in the sequence, no matter how far apart they are. This makes it perfect for getting the global context needed to tell complex arrhythmia patterns apart.

When you use residual learning for hierarchical feature extraction and transformer encoding for contextual modeling together, you get a structure that is more powerful than any single method used by itself.

The system follows an encoder-based architecture, consisting of 3 identical blocks arranged in the center row. Each block contains two sub-layers: a fully connected position wise feedforward network and a multi-head self-attention mechanism. As shown in Figure 1, residual connections [43] and layer normalization [44] are applied after each block.

An important difference between the Transformer model and other convolutional networks is that it separates feature transformation from feature aggregation. In a standard convolutional network, the processes of feature transformation and feature accumulation are usually combined in a convolution layer. The non-linear activation then occurs in a later layer of the model. The Transformer model splits the two above mentioned processes (collecting features using self-attention and collecting information by focusing on different parts of the input sequence) into separate layers of its architecture. In addition, self-attention enables the model to take advantage of the relationships between features, enabling the model to gather data from all over the input sequence as opposed to just one area. The ability of the model to understand both long distance dependencies and the global context is one of the strengths of self-attention mechanisms.

This is due to the fact that the sizes of the convolution filters used in convolutional architectures may limit the size of the receptive fields. The Feed Forward Network (FFN) portion of a transformer's architecture serves to transform the aggregated features into new representations after all aggregation operations have taken place. In other words, the FFN provides a nonlinear transformation to the aggregated features allowing for the discovery of abstract relationships between inputs. Because the FFN operates independently on each element of the input sequence it further emphasizes the distinction between the two major components of a transformer's architecture: the aggregation and transformation steps.

In addition, the Multi-Head Attention Mechanism in the third sublayer helps to refine the output of this FFN. The Multi-Head Attention Mechanism allows the model to understand how the various features of the input sequence relate to one another by simultaneously examining multiple elements of the sequence. Therefore, the model can create a more accurate and comprehensive representation of the features present in the input sequence. This attention mechanism also takes the output of the encoder blocks to ensure that the features being processed by the network are both relevant to the current context and have been transformed into a form that is more beneficial to the task that will follow. Therefore, the overall architecture of the Transformer model lends itself well to the application of NLP and sequence modeling, where the discovery of long range abstract relationships is critical.

This architecture solves the problem of class imbalance via its multi-head attention mechanism. Thus, it can be able to recognize the unique characteristics of poorly represented arrhythmia classes without requiring the use of explicit class weighting or data augmentation strategies.

The hierarchical decomposition process uses the proposed model as a step through gradual network that learns features at different levels of abstraction by capturing basic to Higher-Order Representations in ECG Signals. In other words, the Conv1D Layers learn Low-Level Waveform Features (i.e., P-Waves, QRS Complexes, and T-Waves) while the Residual Blocks capture Mid-Level Short-Term Dependencies between Adjacent Components and Beat-to-Beat Variability; afterward, the transformer layers capture high-level contextual patterns and long range dependencies between beats.

Feature values have been modified multiple times, first with raw ECG data (i.e., a time series), followed by activation maps, which emphasize certain morphological features. It is within these activation maps that the Transformer then uses residual connections, as well as weighted information provided by the attention mechanism, to refine them. As such, it is possible to recognize the specific activation patterns for each arrhythmia. For example, normal beats are typically distributed evenly throughout the cardiac cycle. Supraventricular ectopic beats will be looking at areas around the P-wave; ventricular ectopic beats will be looking at wide QRS complexes; and Fusion Beats will exhibit characteristics of both in different manners. In doing so, the model was able to reach an accuracy of 99%, and maintain a strong performance for all categories of arrhythmias via progressive transformations of feature extraction.

4. Experimental Setup, Results, and Discussion

4.1. Database Configuration

In order to ensure that the proposed model was working with two well-known data sets for heartbeat detection in this research, both the MIT-BIH arrhythmia dataset 1 [45] and the diagnostic ECG dataset PTBDB 2 [46] were used. Both data sets have an operation frequency of 125 Hz. A band pass filter is used in the pre-processing stage of this research. The band pass filter will allow certain frequency bands to pass through while blocking all other frequencies outside of those allowed.

This allows the unwanted noise or interferences to be removed from the signals. The proposed work has taken the collected data and separated it into individual samples. Each sample represents one beat. The proposed work has also removed sections of the data where abnormalities occur, i.e., electrode noise or muscle activity. The proposed work has cropped, down sampled, and padded the sample with zeros so that every sample is now the same length in order to make sure they are consistent. The MIT-BIH arrhythmia dataset has 109,446 heartbeats. The heartbeats are categorized into five different categories: normal beats (N), supraventricular ectopic beats (S), ventricular ectopic beats (V), fusion beats (F), and unknown beats (Q). The dataset consists of 90,589 normal beats, 8,039 supraventricular ectopic beats, 7,236 ventricular ectopic beats, 2,779 fusion beats, and 803 beats of uncertain origin. The proposed work categorizes the PTBDB dataset into two distinct groups: normal and aberrant. Figure 4 shows examples of standardized segments of ECG data covering depolarization/repolarization as well as the isoelectric baseline time course following the surface ECG. The overlapping Flat Zero-Value Regions can seem uninformative, but they are critical for defining cycle limits and serve as anchors for automated classification algorithms. The "Samples" x-axis shows discrete measurements at 125 Hz (8 ms), which makes it easy to compare different types of arrhythmia.

The sinoatrial (SA) node starts the regular contractions of the heart, which move through the atria and end up in the pulmonary arteries. In contrast, additional heartbeats that originate in the atrioventricular node (AV) or the atria are known as supraventricular ectopic beats. The cardiac muscle contracts too soon in response to cardiac ectopic beats, which start in the lower chambers of the heart. The heart produces an abnormal rhythm known as a fusion beat whenever two electrical impulses occur at the same time. The sinoatrial (SA) node, the typical pacemaker of the heart, produces one impulse, and another impulse comes from a different location within the heart. Because of this, a composite or hybrid heartbeat can develop, in which cardiac muscle contractions are partially synchronized with the regular heartbeat and partially due to the ectopic heartbeat.

4.2. Evaluation Mechanism

The proposed work evaluates the proposed model using the Keras 2.14.0 infrastructure designed for deep learning. The TensorFlow machine learning framework serves as its foundation. The proposed work conducted the experiments on the Kaggle platform, using a GPU P100. The training process lasted 360.4 seconds. Metrics are used to assess or measure the performance or efficacy of a specific task.

- Accuracy is determined by dividing the number of correct predictions by all predictions. In essence, it represents the percentage of precise predictions generated by the model. However, as shown with imbalanced datasets, accuracy can be deceiving, especially when the majority class has a strong degree of dominance. A model that constantly predicts the majority class, for example, may maintain a high level of accuracy even if it does not produce any actual forecasts. In essence, it quantifies the percentage of precise forecasts generated by the model. However, accuracy may be deceptive in some circumstances where the dominant class holds significant influence, such as in unbalanced datasets.

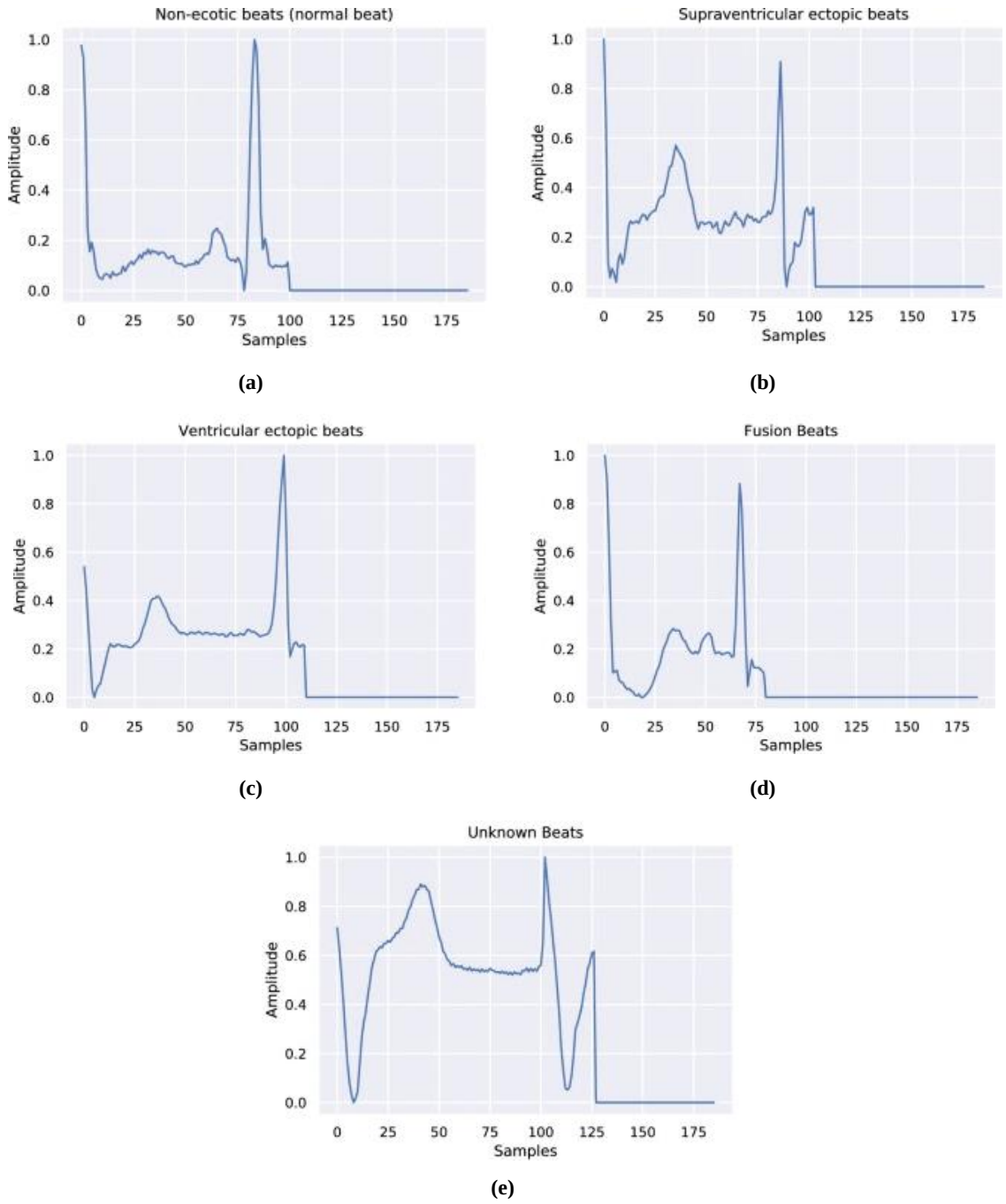

Figure 4. ECG signal examples of five heartbeat classifications: (a) Normal beat (non-ectopic); (b) Supraventricular ectopic beat; (c) Ventricular ectopic beat; (d) Fusion beat; (e) Unknown beat. Each graph shows amplitude (y-axis) versus samples (x-axis) with distinctive waveform patterns characteristic of each arrhythmia type.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

- Precision is a quantitative measure that determines the accuracy of positive statements. The proposed work calculates it by dividing the number of accurate positive statements by the summation of true and false positive predictions. Accuracy is particularly crucial, while the repercussions of incorrect positive results are costly.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

- Recall is a portion of positive instances detected accurately by the classifier. Divide the total number of accurate positive predictions by the total number of inaccurate positive and negative predictions to perform the computation. When the consequences of a false negative are significant, the recall is extremely valuable.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

- The F1 score is a mathematical average that combines precision and recall. An F1-score of 1 implies an optimal value between precision and recall, while a zero value suggests an imbalance. When there is a requirement to achieve a compromise between precision and recall, the F1 score is valuable, particularly in situations with imbalanced datasets where the goal is to minimize false negatives.

$$\mathrm{F1} = \frac{2*Presision*Recall}{Presision+Recall} = \frac{2*TP}{2*TP+FP+FN} \quad (7)$$

- ROC Under-curve: is an acronym for Receiver Operating Characteristic. It is a graphic that illustrates the efficacy of a binary classification model. The ROC curve plots the genuine positive rate against itself. The term "under-curve" denotes a visual representation or representation to evaluate a model's effectiveness. The visual representation illustrates the relationship between the sensitivity (true positive rate) and the specificity (false positive rate) at different threshold values. The false positive rate (or 1-specificity) is the proportion of negative cases that are mistakenly forecast as positive relative to the total number of real negative instances.

$$\mathrm{FP\ Rate}\ (1-\mathrm{Specificity}) = \frac{FP}{FP-TN} \quad (8)$$

The ROC curve provides an overall idea of how well your model is performing at different threshold values. In addition to this, it is particularly useful on medical diagnostic datasets, where class distributions may be imbalanced. For example, if the prevalence of a disease is low, you can easily get datasets where there are tens or hundreds more negative examples compared to positive. In these situations, traditional accuracy measures can be misleading; consequently, a model that consistently predicts the majority class may appear to be performing well, despite its consistent inability to identify the minority class. Another great thing about the AUC ROC Curve is that it gives you a single number that you can use to compare how well the classifier is doing and see how well it can tell the difference between things. It is a measure of the strength of this relationship and between 0 (no discrimination) and 1 (perfect discrimination). This single score makes it easy to compare models and helps researchers and practitioners quickly see the winners. The ROC AUC also doesn't change based on the class, so it can be used as a standard to test classifiers in various clinical situations. It is particularly useful in medical settings where creating balanced datasets can be difficult. The ROC AUC is useful for evaluating diagnostic tools because it is not too sensitive to class imbalances. In this role, it often gives an explanation and confirmation of improvement.

The proposed work processed the datasets in batches of 32, as shown in Table 3. The proposed work chose this value because it was a middle ground between how well batch gradient descent works and how well stochastic gradient descent works, taking into account the limited resources it had. Furthermore, it established the learning rate at 0.001, which improves the functionality of the network by stabilizing the findings. The proposed work updates the weights using the Adamax optimizer and its parameters. This method builds on Adadelta and RMSprop by keeping an average of past gradients that are going down, which is similar to the idea of momentum. The proposed work found the ideal parameters for vector dimensionality and multi-head attention through experimental optimization.

The proposed work used a 10-fold stratified cross-validation method during the assessment phase to reduce the chance of leaving cases out of the training or validation stages. This method involves dividing your data into 10 portions for a series of iterations. In each of these iterations, you will use 90% of your total data to build a model, which you then test with the other 10%. By doing so, you are ensuring that each record will be used during both the training phase and the testing phase, during all ten iterations.

Table 3. The hyperparameter values in this study

Hyperparameter	Value	Notes
Batch Size	32	Same for both datasets
Learning Rate	0.001	Same for both datasets
Optimizer	Adamax	Same for both datasets
Epochs	20	Same for both datasets
Residual Blocks	3	Same for both datasets
Transformer Encoders	3	Same for both datasets
Attention Heads	8	Same for both datasets
d_model (embedding dim)	64	Same for both datasets
Dense Layer 1	256	Same for both datasets
Dense Layer 2	128	Same for both datasets
Output Layer	**5 (MIT-BIH), 2 (PTBDB)**	**Only parameters that changes**
Dropout Rate	0.2	Same for both datasets
L2 Regularization	0.001	Same for both datasets

At the conclusion of each iteration, a classifier model will have been developed and tested, and the overall mean performance of the model will be determined. This process is beneficial in reducing the chances of the model overfitting by increasing its ability to generalize. Due to the fact that the model uses a stratified cross-validation technique, the model is assured that each portion of the model has the same proportion of classes as does the entire data set. This provides a much more accurate way to evaluate the models' performance.

4.3. Experimental Results and Analysis

In the training phase of this process, the model continues to refine itself through a series of iterative processes aimed at closing the gap between the target values and the values predicted by the model. However, there is a potential risk that the model becomes too specialized in training data, leading to decreased performance in unseen testing data, a phenomenon known as overfitting. To mitigate this issue, the proposed work integrates a validation set to optimize the hyperparameters throughout the training process. Specifically, the proposed work partitioned the datasets to allocate 10% for the validation set.

The proposed model is structured as a modular framework, in which all structural elements are fixed, except for the last output layer that can be adapted to different types of classification tasks. Therefore, depending on the task at hand, this output layer will either be the output layer that produces the 5 dimensional vector of probabilities, or the output layer that produces a binary output vector. The remaining elements of the model such as the residual feature extractor, the transformer encoder stack, the first three dense layers (256 and 128 units) remain unchanged and therefore provide a common feature extraction capability for both datasets, thus demonstrating the ability of the model to generalize to different cardiac diagnostic tasks without having to modify the overall architecture.

The training accuracy of the proposed model and the error loss in the data set utilizing validation and training sets are shown in Figures 5 and 6. The y-axis indicates the accuracy or error loss of the model in both the training and validation sets, while the x-axis represents the iterations per input batch. The research assesses the efficacy and stability of this model at epoch 36 of the training phase. To achieve this, the feature extraction process uses skip connections to update the weights and the transformer technique to keep important features in long sequences. In addition, in the initial stages, the training phase improves the temporal complexity by refining the experimental setup. The findings, including loss values and slight variations in training and validation accuracy, reflect how effective the model is.

The proposed work used a stratified 10-fold cross-validation to assess this model. The evaluation metrics for both datasets, including accuracy, precision, recall, and F1 score, are presented in Table 4. Upon testing with the test set, the model demonstrated 97% or higher precision, recall, and F1 score values for classes N, V, and Q, with somewhat marginally better results in classes S and F. This model outperforms the results reported in Table 1 of previous research. For example, in the study by Guo et al. [33], class V had an F1 score of 89.75%, while class S had an F1 score of 61.25%. Using the proposed approach, these values were enhanced by 28% and 6%, respectively. Although Serhani et al. [35] achieved commendable results by using reinforcement learning to optimize CNN hyperparameters (97% precision and 94% recall), the proposed model exhibits enhanced performance with 99% accuracy and 96% average recall, especially for minority arrhythmia classes such as SVEB and fusion beats, where other models generally falter. Earlier studies that used sampling techniques to reduce class imbalance achieved assessment metric values of approximately 99%. This model demonstrates promising results, with improvements of 1-1.5% in accuracy, recall (sensitivity), and F1 scores.

To explore how the transformer will deal with class imbalance and capture long-term temporal dependencies we completed a series of ablation studies by removing the transformer components and only using the residual backbone. The results from these studies are presented in Table 5.

The proposed work compared the proposed model with other related work in Table 6 and observed that although this study had excellent accuracy scores, the proposed mean recall is similar in all categories. The excellent performance of the proposed model and the use of raw data set it apart from similar previous attempts, demonstrating its robustness. The reason for the importance of the demonstration is that it has shown the ability of the model to be able to have a high level of accuracy and recall; all of this was accomplished with no heavy pre-processing nor data augmentation needed. As compared to many other models, these models normally have to perform much of their own work prior to increasing accuracy and adjusting for the randomness and uncertainty of the data used in the model.

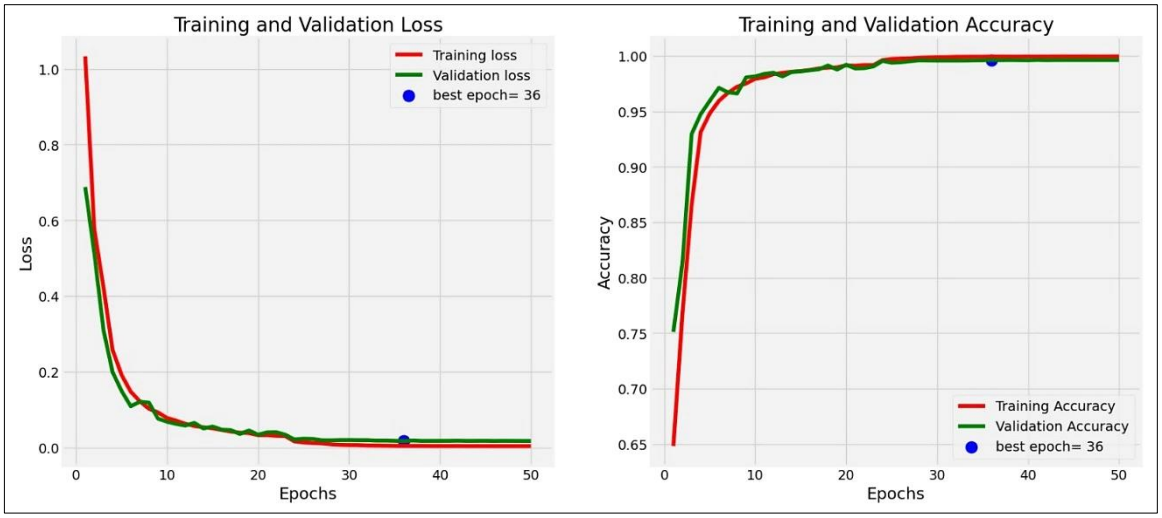

Figure 5. The training phase was conducted on the MIT-BIH dataset

Table 4. Evaluation Results of the proposed work.

Dataset	Classes	Precision	Recall	F1 score
MIT-BIH	N	0.99	0.98	0.99
	S	0.93	0.87	0.90
	V	0.95	0.96	0.95
	F	0.88	0.87	0.87
	Q	0.98	0.99	0.99
PTBDB	Normal	0.98	0.96	0.97
	Abnormal	0.98	0.99	0.99

Additionally, the model being consistent across different classifications indicates that it will perform well under a variety of circumstances and a wide range of cardiac events; thus, it may be useful in clinical settings. High precision and high life-threatening stroke recall rates in critical classes such as V (ventricular ectopic beats) and S (supra-ventricular ectopic beats) indicates that status model can identify dangerous heart problems. The model was able to achieve this result through the use of an advanced method (transformer) and a method (skip connection) that takes advantage of vital information about long term dependency and provides improved performance. Optimizing hyper parameters during training using a validation set allows the model to be optimized for best results.

Table 5. Ablation Study: Component-wise Performance Analysis on MIT-BIH Dataset

Model Variant	Overall Accuracy (%)	Class N (%)	Class S (%)	Class V (%)	Class F (%)	Class Q (%)	Avg. Recall (%)	F1-Score (%)
Residual Only (No Transformer)	96.8	99.2	82.4	94.6	78.3	88.9	88.7	89.3
Transformer Only (No Residual)	94.2	98.1	76.8	89.3	71.5	82.6	83.7	84.8
Residual + Transformer (Proposed)	99.4	99.8	94.2	98.7	91.6	96.3	96.1	96.8
Improvement over Residual Only	+2.6	+0.6	+11.8	+4.1	+13.3	+7.4	+7.4	+7.5

Table 5 clearly shows how much each part of the architecture contributes to its performance, as demonstrated through an ablation study. The Residual Only version of the architecture achieved 96.8% overall accuracy, however it has very poor performance on the minority classes. More specifically, the Residual Only version of the architecture has a very poor performance on class S (Supraventricular ectopic beats, 82.4%) and class F (Fusion beats, 78.3%), which are the two most difficult to distinguish from one another due to their unique morphologies. Therefore, we can conclude that while residual learning is capable of extracting hierarchical local features, it cannot be used to capture the long range temporal relationships required to classify rare types of arrhythmic patterns.

On the other hand, the Transformer Only version of the architecture performed even worse than the Residual Only version with an overall accuracy of 94.2%. Additionally, the Transformer Only version had a substantially larger drop in performance in both classes S (76.8%) and F (71.5%). Thus, this is evidence that transformers have global attention capabilities, but they need robust hierarchical feature extraction to process the morphological complexities associated

with ECG signals. Consequently, when transformers directly process raw signals using the same attention mechanisms used in Residual Blocks and do not use residual based feature refinement, it results in sub-optimal representation learning.

As shown in Table 5, the full version of our model (Residual + Transformer) has a high level of performance with an overall accuracy of 99.4%, and has a similar level of performance across all classes, including the minority classes S (94.2%) and F (91.6%), which represent improvements of 11.8% and 13.3% over Residual Only. The ablation study conducted in this paper provides empirical evidence that the multi-head attention mechanism in transformer layers allows the model to implicitly address class imbalance by providing discriminative attention weights for rare arrhythmic patterns. Furthermore, the dynamic nature of the attention heads allow them to automatically focus on the subtle morphological differences in the minority classes (i.e., early P waves in SVEB, wide QRS complexes in Fusion Beats), thus eliminating the need for explicit class weighting and/or data augmentation strategies. Ultimately, we can conclude that the synergistic combination of residual learning for extracting hierarchical local features and transformer attention for understanding global context is essential to achieving accurate multi-class arrhythmia classification on imbalanced datasets. The evaluation procedure outlined here is also in line with the large number of published MIT-BIH studies referenced in Table 6 as a basis of comparison that typically rely on intra-patient (beat-level) cross-validation to test models. Furthermore, since all testing data was partitioned off from the training data, no early stopping, model selection, or any other type of information leak occurred during training.

Table 6. The evaluation of the suggested model in comparison to SOTA studies.

Author	Model	Accuracy (%)	Recall (%)
Guo et al. (2019) [33]	DenseNet-Attention-GRU	92	82
Kachuee et al. (2018) [30]	Deep residual CNN	93	93
Xu et al. (2020) [31]	CNN+Bi-LSTM	96	92
Peimankar et al. (2024) [32]	Residual Transfer Learning	97	-
Xia et al. (2023) [34]	Transformer models	97	93
Serhani et al. (2025) [35]	RL-optimized CNN	97	94
Meeran & Munaf (2026) [39]	Bi-RNN architecture	98	96
Proposed model	**Hybrid Residual Transfer Learning and Transformer Models**	**99**	**96**

In comparison to all other models, the proposed method shows better results (in terms of evaluation metrics) on both validation and test data sets. Its high quality performance, combined with a relatively low number of parameters, also makes it a very attractive solution for classifying arrhythmias.

One of the unique characteristics of the proposed model is its high level of robustness when performing multi-class classification across large numbers of classes. That is, the model performs well on a wide range of classes, even when there are imbalances in the distribution of the classes, which is common in many medical problems including those related to arrhythmias. For this reason alone, this is a valuable tool for providing diagnostic support for patients with arrhythmias, and can lead to improved patient outcomes.

This proposed model addresses one of the most common issues associated with the use of deep neural networks, specifically the vanishing gradients issue that causes the network to fail to converge or train properly. By incorporating methods into the model that allow for the use of skip connections and/or transformer layers, the proposed model allows for rapid convergence and reduced computational cost. This provides an advantage over other similar models since it will be less resource intensive and thus more likely to be used in clinical settings where resources are limited.

This model also has advantages when implemented with the simplicity of the model. A health care professional can easily incorporate this model into current systems to help diagnose patients as no specific changes need to be made to those systems. Together these simple-to-use and fast performing characteristics of this model make it an ideal solution to detect and classify arrhythmias. The proposed model was tested with strong statistical techniques such as ROC curves, along with the technical characteristics of the model. Higher Area Under Curve (AUC) values among classes indicate that the model is reliable and can differentiate between classes. This reliability is further supported by the results obtained in the testing and validation phases, which were very positive and support the credibility of the model.

This model provides superior classification accuracy compared to any existing model. The authors achieved a total of 99.4% accurate classifications in two large datasets through the combination of residual transfer learning and transformers. In addition to the ability to process class imbalance and provide consistent high quality results across many types of arrhythmia classifications, the model is capable of processing with minimal preprocessing and/or additions to the input data. The most significant advantage of the model is that it provided significantly improved rates of recall and precision over previously developed methods, resulting in a 5-28 percent improvement in the classification metrics. The

use of skip connections and multi-head attentions within the design of the model enable it to effectively solve issues related to gradient vanishing and long-range dependencies. Therefore, the proposed model is particularly promising for practical application by clinical staff and enables effective automated detection of arrhythmias.

More importantly than just being an increase in overall accuracy the proposed model also improves on class-wise recall and AUC values for both majority and minority arrhythmia classes, which means it has increased discrimination capabilities across all arrhythmia types. This indicates the hybrid residual – transformer model has a well-rounded representation learning approach that can help mitigate the typical deep learning short comings in ECG classification tasks.

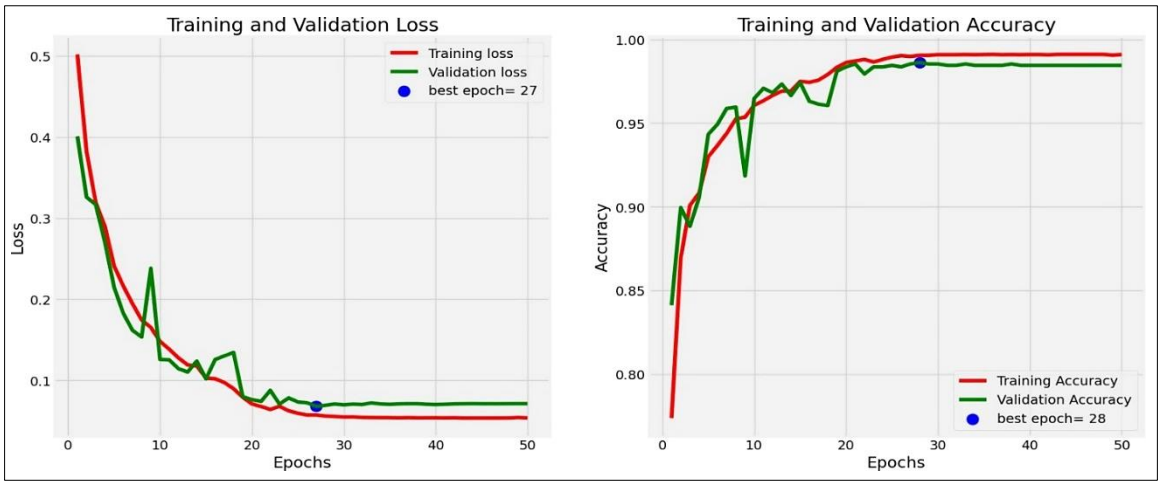

Figure 6. The training phase was conducted on the PTBDB dataset

The proposed model was thoroughly evaluated using the Receiver Operating Characteristic (ROC) approach. The ROC curve demonstrates the performance of the classification system at various discrimination levels, as illustrated in Figure 7. This ROC metric assesses the model's ability to distinguish between positive and negative classifications. The Area Under the Curve (AUC) is a measurement that varies from 0.5 to 1, where 1 indicates perfect classification, and 0.5 signifies no ability to discriminate. In Figure 7, the recall metric is shown on the y axis, while the false positive rate (FPR or fallout) is shown on the x axis. The FPR is calculated by dividing the number of false positives by the total of false positives and true negatives. The ROC curve illustrates the trade-off between these two metrics across different categorization thresholds. A high True Positive Rate (TPR), indicating few false negatives, shows the model's effectiveness in identifying positive events, alongside a low FPR, indicating few false positives. Consequently, the ROC curve effectively demonstrates the performance of the model across various thresholds, highlighting its stability and reliability in classification tasks.

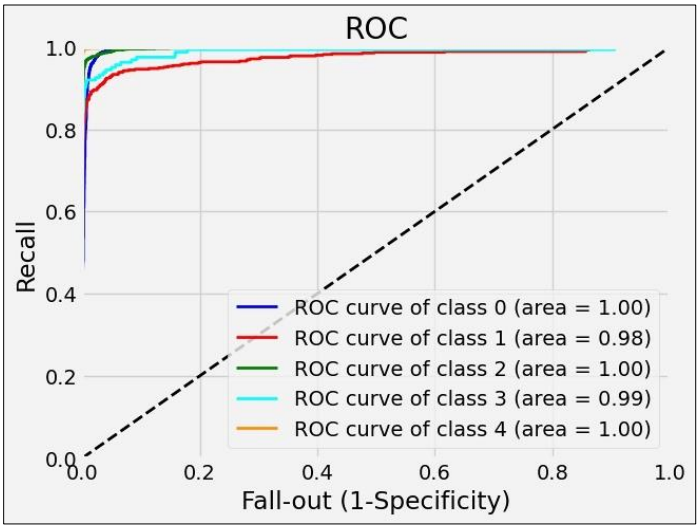

Figure 7. The ROC Under-curve values MIT-BIH

As shown in Figure 7, the ROC values of the five categories (N, V, F, Q, and S) that were obtained using the MIT-BIH dataset have been plotted. In terms of categorization, classes N, V, F and Q achieved excellent results as they

exhibited both high recall values (1) and low fall-out values (0), indicating the excellent performance of the proposed model. Class S has the highest.

AUC value of 0.98. Also, in the PTBDB dataset, both classes had significant results and an AUC of 1. This indicates that the proposed model has demonstrated a consistent and strong classification capability, regardless of the threshold used for class distinction. The training process validation results also illustrate the simplicity and efficiency of this model. After only 20 epochs, the model has demonstrated a high degree of accuracy in detecting various arrhythmia conditions and is characterized by low error rates; this reflects the model's high degree of expertise in the field of arrhythmia detection. The proposed model also demonstrated excellent performance during the test phase when it was applied to hidden target data and was able to achieve a high percentage of success across a number of evaluation criteria with a notable result of 99%; this is a highly encouraging sign. Additionally, the proposed model offers a number of benefits as a model architecture, such as the ability to mitigate the problem of the vanishing gradient.

5. Conclusion

This study introduces a deep learning model as an example of how residual transfer learning can be combined with transformers and attention-based approaches for automated arrhythmia classification of electrocardiogram (ECG) signals. This combination of residual and transformer-based architectures is used to achieve very high levels of performance when tested on two large-scale benchmark datasets. These results were achieved using 99.4% accuracy on the 5-class MIT-BIH arrhythmia dataset and 99.8% accuracy on the binary PTBDB dataset. The model demonstrated even greater advantages than previously reported residual-based architectures on infrequent arrhythmia classes: 94.2% for supraventricular ectopic beats and 91.6% for fusion beats. Those values represent a significant improvement (11.8% and 13.3%, respectively) over the performance of the baseline residual-only model. The ablation studies confirmed that the multi-head attention mechanism of the transformer component, which allows the model to learn discriminative attention weights for rare arrhythmia patterns, implicitly solves the problem of class imbalance without relying upon either explicit class weighting or data augmentation.

From a clinical perspective, the implications of this research are also significant. Due to the model's ability to accurately classify potentially life-threatening arrhythmias, this model could provide a valuable tool for use in automated cardiac monitoring systems that could reduce cardiologist workload and facilitate earlier interventions. However, as acknowledged by the authors, the current evaluation protocol did not strictly separate patients from one another at the level of the individual, which limits our ability to assess the model's ability to generalize to entirely unseen patients who have different cardiac characteristics or comorbidities. Therefore, future research will involve evaluating the model's performance in a patient-by-patient fashion, assessing the model's generalizability to entirely unseen patients with diverse cardiac characteristics and comorbidities. Additional validation of the model's performance on larger multi-center datasets to further support the model's clinical utility and robustness across diverse patient populations.

6. Declarations

6.1. Author Contributions

Conceptualization, N.A.; methodology, N.A.; software, N.A. and E.A.E.; validation, N.A., H.A.O., H.A.M., and A.A.A.; formal analysis, N.A.; investigation, N.A. and H.A.O.; resources, A.A.A., M.R., and M.Y.A.; data curation, N.A. and S.A.; writing—original draft preparation, N.A.; writing—review and editing, N.A., H.A.O., H.A.M., A.A.A., M.R., and M.Y.A.; visualization, N.A. and E.A.E.; supervision, A.A.A., M.R., and M.Y.A.; project administration, N.A. and A.A.A.; funding acquisition, A.A.A., M.R., and M.Y.A. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available in the article.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not Applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. References

[1] Lopez-Jimenez, F., Almahmeed, W., Bays, H., Cuevas, A., Di Angelantonio, E., le Roux, C. W., ... & Wilding, J. P. (2022). Obesity and cardiovascular disease: mechanistic insights and management strategies. A joint position paper by the World Heart Federation and World Obesity Federation. European journal of preventive cardiology, 29(17), 2218-2237. doi:10.1093/eurjpc/zwac187.

[2] Gaidai, O., Cao, Y., & Loginov, S. (2023). Global Cardiovascular Diseases Death Rate Prediction. Current Problems in Cardiology, 48(5), 101622. doi:10.1016/j.cpcardiol.2023.101622.

[3] January, C. T., Wann, L. S., Calkins, H., Chen, L. Y., Cigarroa, J. E., Cleveland, J. C., Ellinor, P. T., Ezekowitz, M. D., Field, M. E., Furie, K. L., Heidenreich, P. A., Murray, K. T., Shea, J. B., Tracy, C. M., & Yancy, C. W. (2019). 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients with Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society in Collaboration with the Society of Thoracic Surgeons. Circulation, 140(2), e125–e151. doi:10.1161/CIR.0000000000000665.

[4] Zheng, J., Zhang, J., Danioko, S., Yao, H., Guo, H., & Rakovski, C. (2020). A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific Data, 7(1), 48. doi:10.1038/s41597-020-0386-x.

[5] Kubala, M., de Chillou, C., Bohbot, Y., Lancellotti, P., Enriquez-Sarano, M., & Tribouilloy, C. (2022). Arrhythmias in Patients with Valvular Heart Disease: Gaps in Knowledge and the Way Forward. Frontiers in Cardiovascular Medicine, 9, 792559. doi:10.3389/fcvm.2022.792559.

[6] Al- Naami, B., Fraihat, H., Owida, H. A., Al- Hamad, K., De Fazio, R., & Visconti, P. (2022). Automated Detection of Left Bundle Branch Block from ECG Signal Utilizing the Maximal Overlap Discrete Wavelet Transform with ANFIS. Computers, 11(6), 93. doi:10.3390/computers11060093.

[7] Bonny, A., Ngantcha, M., Scholtz, W., Chin, A., Nel, G., Anzouan-Kacou, J. B., Karaye, K. M., Damasceno, A., & Crawford, T. C. (2019). Cardiac Arrhythmias in Africa: Epidemiology, Management Challenges, and Perspectives. Journal of the American College of Cardiology, 73(1), 100–109. doi:10.1016/j.jacc.2018.09.084.

[8] Cheung, C. C., Roston, T. M., Andrade, J. G., Bennett, M. T., & Davis, M. K. (2020). Arrhythmias in Cardiac Amyloidosis: Challenges in Risk Stratification and Treatment. Canadian Journal of Cardiology, 36(3), 416–423. doi:10.1016/j.cjca.2019.11.039.

[9] Petty, B. G. (2020). Basic electrocardiography. Springer Nature, New York, United States. doi:10.1007/978-1-4939-2413-4.

[10] Das, M. K., & Zipes, D. P. (2021). Electrocardiography of Arrhythmias: A Comprehensive Review E-Book: A Companion to Cardiac Electrophysiology. Elsevier, Amsterdam, Netherlands.

[11] Brady, W. J., Lipinski, M. J., Darby, A. E., Bond, M. C., Charlton, N. P., Hudson, K., & Williamson, K. (2020). Electrocardiogram in Clinical Medicine. John Wiley & Sons, New Jersey, United States. doi:10.1002/9781118754511.

[12] Ikeda, T. (2021). Current use and future needs of noninvasive ambulatory electrocardiogram monitoring. Internal Medicine, 60(1), 9–14. doi:10.2169/internalmedicine.5691-20.

[13] Avula, V., Wu, K. C., & Carrick, R. T. (2023). Clinical Applications, Methodology, and Scientific Reporting of Electrocardiogram Deep-Learning Models: A Systematic Review. JACC: Advances, 2(10), 100686. doi:10.1016/j.jacadv.2023.100686.

[14] Ouyang, D., Theurer, J., Stein, N. R., Hughes, J. W., Elias, P., He, B., Yuan, N., Duffy, G., Sandhu, R. K., Ebinger, J., Botting, P., Jujjavarapu, M., Claggett, B., Tooley, J. E., Poterucha, T., Chen, J. H., Nurok, M., Perez, M., Perotte, A., … Albert, C. M. (2024). Electrocardiographic deep learning for predicting post-procedural mortality: a model development and validation study. The Lancet Digital Health, 6(1), e70–e78. doi:10.1016/S2589-7500(23)00220-0.

[15] Hughes, J. W., Tooley, J., Torres Soto, J., Ostropolets, A., Poterucha, T., Christensen, M. K., Yuan, N., Ehlert, B., Kaur, D., Kang, G., Rogers, A., Narayan, S., Elias, P., Ouyang, D., Ashley, E., Zou, J., & Perez, M. V. (2023). A deep learning-based electrocardiogram risk score for long term cardiovascular death and disease. NPJ Digital Medicine, 6(1), 169. doi:10.1038/s41746-023-00916-6.

[16] Aqel, H., Farah, H., & Al-Hunaiti, A. (2024). Ecological versatility and biotechnological promise: Comprehensive characterization of the isolated thermophilic Bacillus strains. Plos One, 19(4 April), 297217. doi:10.1371/journal.pone.0297217.

[17] Abu Sa'aleek, A., Alshishani, A., Shaghlil, L., Aljariri Alhesan, J. S., & Al-Ebini, Y. (2023). Determination of vitamin D3 in pharmaceutical products using salting-out assisted liquid–liquid extraction coupled with reversed phase liquid chromatography. Microchemical Journal, 193, 109077. doi:10.1016/j.microc.2023.109077.

[18] Xiao, Q., Lee, K., Mokhtar, S. A., Ismail, I., Pauzi, A. L. bin M., Zhang, Q., & Lim, P. Y. (2023). Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review. Applied Sciences (Switzerland), 13(8), 4964. doi:10.3390/app13084964.

[19] Aseeri, A. O. (2021). Uncertainty-aware deep learning-based cardiac arrhythmias classification model of electrocardiogram signals. Computers, 10(6), 82. doi:10.3390/COMPUTERS10060082.

[20] Murat, F., Yildirim, O., Talo, M., Baloglu, U. B., Demir, Y., & Acharya, U. R. (2020). Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. Computers in Biology and Medicine, 120, 103726. doi:10.1016/j.compbiomed.2020.103726.

[21] Sannino, G., & De Pietro, G. (2018). A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems, 86, 446–455. doi:10.1016/j.future.2018.03.057.

[22] Siontis, K. C., Noseworthy, P. A., Attia, Z. I., & Friedman, P. A. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology, 18(7), 465–478. doi:10.1038/s41569-020-00503-2.

[23] Hassan, S. U., Mohd Zahid, M. S., Abdullah, T. A. A., & Husain, K. (2022). Classification of cardiac arrhythmia using a convolutional neural network and bi-directional long short-term memory. Digital Health, 8, 20552076221102770. doi:10.1177/20552076221102766.

[24] Essa, E., & Xie, X. (2021). An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification. IEEE Access, 9, 103452–103464. doi:10.1109/ACCESS.2021.3098986.

[25] Liu, P., Sun, X., Han, Y., He, Z., Zhang, W., & Wu, C. (2022). Arrhythmia classification of LSTM autoencoder based on time series anomaly detection. Biomedical Signal Processing and Control, 71, 103228. doi:10.1016/j.bspc.2021.103228.

[26] Wang, B., Chen, G., Rong, L., Liu, Y., Yu, A., He, X., Wen, T., Zhang, Y., & Hu, B. (2023). Arrhythmia Disease Diagnosis Based on ECG Time-Frequency Domain Fusion and Convolutional Neural Network. IEEE Journal of Translational Engineering in Health and Medicine, 11, 116–125. doi:10.1109/JTEHM.2022.3232791.

[27] Dong, X., & Si, W. (2023). Heartbeat Dynamics: A Novel Efficient Interpretable Feature for Arrhythmias Classification. IEEE Access, 11, 3305473. doi:10.1109/ACCESS.2023.3305473.

[28] Kaniraja, C. P., M, V. D., & Mishra, D. (2024). A deep learning framework for electrocardiogram (ECG) super resolution and arrhythmia classification. Research on Biomedical Engineering, 40(1), 199-211. doi:10.1007/s42600-024-00343-w.

[29] Chen, M. C., & Chen, C. I. (2024). Atrial Fibrillation Detection in Single-Lead ECG Signals: A Comparative Study of CNN with LSTM and Residual Neural Network Models. ACM International Conference Proceeding Series, 138–142. doi:10.1145/3673971.3673980.

[30] Kachuee, M., Fazeli, S., & Sarrafzadeh, M. (2018). ECG heartbeat classification: A deep transferable representation. In Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, 443–444. doi:10.1109/ICHI.2018.00092.

[31] Xu, X., Jeong, S., & Li, J. (2020). Interpretation of Electrocardiogram (ECG) Rhythm by Combined CNN and BiLSTM. IEEE Access, 8, 125380–125388. doi:10.1109/ACCESS.2020.3006707.

[32] Peimankar, A., Ebrahimi, A., & Wiil, U. K. (2024). xECG-Beats: an explainable deep transfer learning approach for ECG-based heartbeat classification. Network Modeling Analysis in Health Informatics and Bioinformatics, 13(1), 1–13. doi:10.1007/s13721-024-00481-2.

[33] Guo, L., Sim, G., & Matuszewski, B. (2019). Inter-patient ECG classification with convolutional and recurrent neural networks. Biocybernetics and Biomedical Engineering, 39(3), 868–879. doi:10.1016/j.bbe.2019.06.001.

[34] Hao, S., Xia, Y., & Ye, Y. (2023). Generative Adversarial Network with Transformer for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters, 20, 104276. doi:10.1109/LGRS.2023.3322139.

[35] Serhani, M. A., Ismail, H., El-Kassabi, H. T., & Breiki, H. Al. (2025). Enhancing arrhythmia prediction through an adaptive deep reinforcement learning framework for ECG signal analysis. Biomedical Signal Processing and Control, 101, 107155. doi:10.1016/j.bspc.2024.107155.

[36] Karthikeyani, S., Sasipriya, S., & Ramkumar, M. (2025). An Evaluation of Dimensionality Reduction and Classification Techniques for Cardiac Disease Diagnosis from ECG Signals with Various Deep Learning Classifiers. Circuits, Systems, and Signal Processing, 44(1), 416–446. doi:10.1007/s00034-024-02845-5.

[37] Zhao, Y., Kang, J., Zhang, T., Han, P., & Chen, T. (2025). ECG-Chat: A Large ECG-Language Model for Cardiac Disease Diagnosis. Proceedings - IEEE International Conference on Multimedia and Expo, 1–6. doi:10.1109/ICME59968.2025.11209476.

[38] Abdullayev, I. N., Jiyanbayev, O. E., & Nasimov, R. K. (2026). Early Detection and Prognostic Assessment of Ischemic Heart Diseases Based on Multimodal Artificial Intelligence Algorithms. Technical Science Integrated Research, 2(1), 33-42.

[39] Meeran, S. B. U., & Munaf, N. A. (2026). Comparative analysis of unidirectional and bidirectional RNNs for ECG arrhythmia detection using augmented MIT-BIH data. Bulletin of Electrical Engineering and Informatics, 15(1), 712–722. doi:10.11591/eei.v15i1.10893.

[40] Akbar, A., & Utami, E. (2026). Systematic Review of the Use of the MIT-BIH Polysomnography Database for the Detection and Classification of Sleep Disorders. Sistemasi: Jurnal Sistem Informasi, 15(1), 308-320.

[41] Wagner, P., Strodthoff, N., Bousseljot, R. D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. Scientific Data, 7(1), 154. doi:10.1038/s41597-020-0495-6.

[42] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 770–778. doi:10.1109/CVPR.2016.90.

[43] Fawzy, A. M., Rivera-Caravaca, J. M., Underhill, P., Fauchier, L., & Lip, G. Y. H. (2023). Incident heart failure, arrhythmias and cardiovascular outcomes with sodium-glucose cotransporter 2 (SGLT2) inhibitor use in patients with diabetes: Insights from a global federated electronic medical record database. Diabetes, Obesity and Metabolism, 25(2), 602–610. doi:10.1111/dom.14854.

[44] Xu, J., Sun, X., Zhang, Z., Zhao, G., & Lin, J. (2019). Understanding and improving layer normalization. Advances in Neural Information Processing Systems, 32.

[45] Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine, 20(3), 45–50. doi:10.1109/51.932724.

[46] Bousseljot, R., Kreiseler, D., & Schnabel, A. (1995). Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, 40(S1), 317–318. doi:10.1515/bmte.1995.40.s1.317.