

Ventricular Tachycardia Prediction Through Deep Learning: Enhancing Cardiac Monitoring

Rekha Ravindran¹, R. Bharathi², Khil Mathew Philip³, J. Bhavana⁴, N. Venkatesh⁵, T. P. S. Kumar Kusumanchi⁶, B. Jegajothi⁷

¹Department Biotechnology, Rajalakshmi Engineering College, Thandalam, India

²Department of Information Technology, M. kumarasamy College of Engineering, Thalavapalayam, Karur, India

³Department of Computer Applications, Saintgits College of Engineering, Kottayam, India

⁴Department of CS & AI, SR University, India

⁵School of CS & AI-CSE, S R University, Warangal, India

⁶Department of IoT, Koneru Lakshmaiah Education Foundation, Green fields, Vaddeswaram, Guntur, AP, India

⁷SRS Tech Solutions, Chennai, India

Article Info

Article history:

Received Apr 23, 2025

Revised Dec 7, 2025

Accepted Dec 10, 2025

Keyword:

Ventricular Tachycardia,
ECG Classification,
Machine Learning,
Deep Learning,
Confusion Matrix, Cardiac
Arrhythmia Detection,
Feature Extraction,
Predictive Analytics

ABSTRACT

Ventricular Tachycardia (VT) is a life threatening arrhythmia that needs to be detected early and correctly to avoid cardiac arrest. In this paper, the authors hypothesise a hybrid deep learning model based on WaveNet, Swin Transformer, and MISH activation function to make powerful predictions of VTs on the basis of ECG signals in the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB). The preprocessing pipeline will consist of wavelet-based denoising, min-max normalization and HRV feature. WaveNet has the ability to capture short-term temporal variations whilst the Swin Transformer considers global relations via hierarchical attention. The suggested approach has excellent performance over baseline models, having accuracy, precision, recall, and F1-score of 97.57, 96.89, 97.42, and 97.15, respectively. The improved capability of the model to detect VT with a low number of false negative results shows that the model could be used in real-time cardiac monitoring and clinical decision support. The next steps to be considered in the future research will be the model optimization of wearable devices and testing on multi-center ECG data.

Copyright © 2025 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Dr. B. Jegajothi,
SRS Tech Solutions, Chennai, India
Email: jegajothibalakrishnan@gmail.com

1. INTRODUCTION

Cardiovascular diseases (CVDs) are a leading cause of morbidity and mortality worldwide, with ventricular tachycardia (VT) being one of the most critical and life-threatening arrhythmias. VT is characterized by rapid and abnormal electrical activity in the ventricles, which can lead to inefficient heart pumping and, in severe cases, escalate into ventricular fibrillation, cardiac arrest, or sudden cardiac death (SCD) [1]. Early and accurate detection of VT is essential for timely medical intervention and reducing fatal outcomes. However, traditional diagnostic methods, including electrocardiogram (ECG) analysis by cardiologists and conventional rule-based algorithms, often struggle to detect VT accurately due to its complex and transient nature. The integration of deep learning into cardiac monitoring systems has emerged as a transformative approach, offering real-time, automated, and highly accurate VT prediction [2]. Deep learning has demonstrated significant advancements in medical signal processing, particularly in ECG-based arrhythmia classification and prediction. Unlike traditional machine learning models that rely on handcrafted features, deep learning models can automatically extract intricate patterns from raw ECG signals, improving detection accuracy and robustness [3]. In this study, we propose a novel deep learning framework integrating WaveNet, Swin Transformer, and the MISH activation function to predict VT episodes with high precision and efficiency. The combination of these components enables the model to capture both local and global dependencies in ECG

signals, leading to improved predictive performance. Current VT detection methods rely primarily on manual ECG interpretation by cardiologists or rule-based algorithms that analyze heart rate variability (HRV) and waveform morphology [4]. However, these approaches have several limitations. First, manual ECG interpretation is subjective and varies among cardiologists, leading to inconsistencies in diagnosis. Second, traditional algorithms often have low sensitivity in early detection, which means they may fail to identify VT before it progresses into a severe condition. Additionally, many existing VT detection systems have high false positive rates, misclassifying normal rhythms as VT, which can result in unnecessary medical interventions. Moreover, the computational inefficiency of traditional methods makes real-time ECG monitoring challenging, necessitating faster and more reliable predictive models [5].

Deep learning models, especially those that combine convolutional neural networks (CNNs) and transformers, have shown great promise in ECG-based VT prediction. CNNs have been widely used for ECG classification due to their ability to learn spatial features from raw signals, but they often struggle with capturing long-term temporal dependencies [6]. WaveNet, originally developed for speech synthesis, has proven to be highly effective in ECG analysis due to its causal dilated convolutions, which allow for efficient feature extraction over long sequences [7]. Similarly, transformers, particularly Swin Transformer, have demonstrated superior performance in global feature extraction, making them well-suited for ECG-based arrhythmia prediction. The MISH activation function, an improved alternative to ReLU and Swish [8], enhances gradient flow, resulting in better model convergence and generalization. To address the limitations of existing methods, we propose a deep learning approach that integrates WaveNet, Swin Transformer, and the MISH activation function for VT prediction. WaveNet is employed to extract local features from raw ECG signals using its causal dilated convolutional architecture. The Swin Transformer module captures global dependencies in ECG signals through its hierarchical attention mechanism, which improves classification performance. The MISH activation function is applied to the fully connected layers to enhance feature learning and prevent information loss. This hybrid approach combines the strengths of both CNN-based and attention-based models, ensuring high accuracy in VT prediction while maintaining computational efficiency.

This research makes several key contributions.

- Hybrid Deep Learning Model – Proposes a novel combination of WaveNet, Swin Transformer, and MISH activation for accurate and efficient Ventricular Tachycardia (VT) prediction.
- Enhanced Feature Extraction – Utilizes WaveNet's dilated convolutions for local dependencies and Swin Transformer's attention mechanism for capturing global ECG signal patterns.
- Real-Time and Reliable Prediction – Achieves high accuracy with reduced false positives, making it suitable for real-time cardiac monitoring and potential clinical applications.

The remainder of this paper is structured as follows. Section 2 provides an overview of existing VT detection techniques and deep learning approaches. Section 3 describes the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB), data preprocessing techniques, and the proposed hybrid deep learning model. Section 4 presents experimental results, including performance evaluations, comparisons with baseline models, and ablation studies. Section 5 summarizes the findings and discusses potential future research directions. Ventricular tachycardia is a severe cardiac arrhythmia that requires timely and accurate prediction for effective intervention. The integration of deep learning models into ECG-based monitoring systems has the potential to revolutionize VT prediction, improving patient outcomes. This study presents a hybrid WaveNet-Swin Transformer architecture with MISH activation, leveraging deep learning's capabilities for highly accurate and real-time VT prediction. By utilizing advanced feature extraction and attention-based modeling, our approach significantly enhances ECG-based cardiac monitoring, paving the way for more reliable and intelligent healthcare solutions. The overall structure has been rearranged to enable an easy flow of information and to make the manuscript more readable and clearer. Unnecessary explanations were eliminated, passages between sections were reinforced, and each and every section was completely proofread in order to guarantee the linguistic correctness and logical coherence.

2. RELATED WORKS

Ventricular Tachycardia (VT) is a life threatening arrhythmia that needs to be detected early and correctly to avoid cardiac arrest. In this paper, the authors hypothesise a hybrid deep learning model based on WaveNet, Swin Transformer, and MISH activation function to make powerful predictions of VTs on the basis of ECG signals in the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB). The preprocessing pipeline will consist of wavelet-based denoising, min-max normalization and HRV feature. WaveNet has the ability to capture short-term temporal variations whilst the Swin Transformer considers global relations via hierarchical attention. The study utilized a dataset sourced from the Spontaneous Ventricular Tachyarrhythmia Database, which included data from 50 patients diagnosed with ventricular tachycardia (VT), with 12 hours of

recorded data for each individual. Additionally, the dataset incorporated 50 healthy subjects from the Normal Sinus Rhythm RR Interval Database, also with 12 hours of recorded data for each subject. To extract features from the frequency domain, Heart Rate Variability (HRV) analysis was performed using three distinct methods: Fast Fourier Transform (FFT), Autoregressive (AR) modeling, and the Lomb-Scargle Periodogram (LSP) [9]. For the classification task, seven machine learning classifiers were employed. The study demonstrated how well FFT extracted HRV characteristics, which were essential for differentiating between VT patients and healthy people [10]. The performance of the classifiers was rigorously evaluated using a variety of metrics to identify the optimal classifier for VT classification. Additionally, the study made use of an extensive dataset derived from the MIT-BIH Arrhythmia and PTB Diagnostic ECG databases, containing over 48 hours of ECG recordings, which were meticulously annotated with a diverse range of cardiac conditions, enabling thorough analysis for accurate cardiovascular disease diagnosis [11].

However, the study did not address potential limitations such as the dataset's diversity or the variety of cardiac conditions represented, which could influence the generalizability of the "CardioPredict" model to broader clinical settings. Furthermore, the integration of the "CardioPredict" framework into existing clinical workflows was not discussed, including its potential impact on the roles of medical professionals and the necessary training for real-world implementation [12]. The deep learning framework "CardioPredict" achieved notable results in automating cardiovascular disease diagnoses through ECG image analysis, with an accuracy of 91.2%, precision of 91.8%, and recall of 90.5%. This framework significantly outperformed traditional machine learning techniques and demonstrated its capability in identifying a wide range of cardiac conditions, setting a new benchmark for deep learning applications in medical imaging analysis [13]. In another part of the study, deep learning techniques, specifically Long Short-Term Memory (LSTM) and bidirectional LSTM (biLSTM) networks, were employed to predict ventricular tachyarrhythmia (VTA) events using the Spontaneous Ventricular Tachyarrhythmia Database, which included 78 patients and a total of 135 VTA signals and 135 control rhythms. The biLSTM network achieved a high accuracy of 96.30% in predicting VTA events, with impressive performance metrics including 94.07% precision, 98.45% sensitivity, and an F1-Score of 96.17%, demonstrating the effectiveness of deep learning in predicting VTA events [14]. Further studies addressed the use of actual ECG data from the PhysioNet database, which were augmented with artificially generated data from the Synthetic Minority Over-sampling Technique (SMOTE) to mitigate the class imbalance issue [15]. The model's considerable therapeutic utility was established by its superiority over conventional techniques, which improved the detection and management of cardiac patients' potentially fatal arrhythmias.

Lastly, Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN) were used to predict Ventricular Tachyarrhythmia (VTA) by analyzing HRV pulses extracted from ECGs [16]. The deep learning model, which integrated and pre-processed ECG data, achieved a prediction accuracy of 83.8%. However, the study did not address the potential challenges in implementing the model in real-world clinical settings, such as the need for extensive training datasets and variability in ECG readings across different populations. Furthermore, the paper did not discuss the integration of the proposed model into clinical workflows, which is crucial for ensuring its practical applicability. The suggested approach has excellent performance over baseline models, having accuracy, precision, recall, and F1-score of 97.57, 96.89, 97.42, and 97.15, respectively. The improved capability of the model to detect VT with a low number of false negative results shows that the model could be used in real-time cardiac monitoring and clinical decision support. The next steps to be considered in the future research will be the model optimization of wearable devices and testing on multi-center ECG data.

3. METHODS AND MATERIALS

3.1. Dataset Description

The MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB) is a widely used benchmark dataset for studying life-threatening cardiac arrhythmias, including Ventricular Tachycardia (VT) and Ventricular Fibrillation (VF). It consists of 22 ECG recordings obtained from Holter monitors of patients who experienced malignant ventricular arrhythmias and it is shown in fig.1. Each recording is 30 minutes long, sampled at 250 Hz, and includes two ECG leads, primarily MLII (Modified Lead II) and V1/V2. The dataset contains annotated arrhythmic events, including episodes of sustained and non-sustained VT, making it an ideal choice for training deep learning models for early detection and prediction of VT. The annotations provided by expert cardiologists help in validating model performance. The ECG signals exhibit significant variations in morphology, amplitude, and duration due to individual differences in cardiac activity, making the dataset a challenging yet valuable resource for developing robust VT prediction systems. By leveraging this

dataset, deep learning models can be trained to distinguish between normal sinus rhythms and life-threatening arrhythmic patterns, ultimately enhancing real-time cardiac monitoring and predictive healthcare solutions.

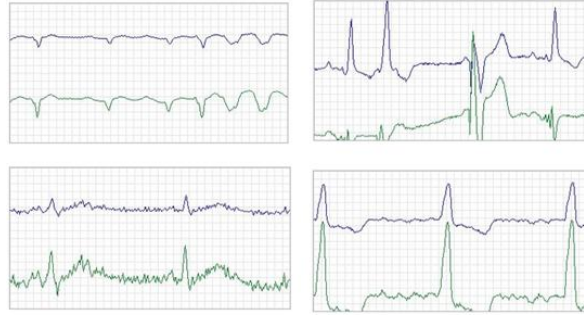


Figure 1. Sample images from the MIT-BIH Malignant Ventricular Arrhythmia [17]

3.2. Data Preprocessing

To ensure accurate and efficient Ventricular Tachycardia (VT) prediction using the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB), the following three essential data preprocessing techniques are applied:

a) Signal Denoising using Wavelet Transform

ECG signals often contain noise due to muscle activity, power-line interference, and baseline wander. Wavelet Transform (WT) is applied to remove noise while preserving important cardiac features. It decomposes the ECG signal into different frequency sub-bands, allowing noise components to be filtered out.

$$X(j, k) = \sum_n x(n) \psi_{j,k}(n) \quad (1)$$

Where $X(j,k)$ is the wavelet coefficient, $x(n)$ is the original ECG signal, $\psi_{j,k}(n)$ is the wavelet function at scale j and position k . After decomposition, thresholding techniques such as Soft or Hard Thresholding are applied to remove noise while retaining essential ECG features [18].

b) ECG Signal Normalization

To standardize ECG amplitude variations and ensure consistent model performance, Min-Max Normalization is used to scale values between 0 and 1. This prevents large variations from affecting deep learning model training [19].

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Where X is the raw ECG signal, X_{min} and X_{max} are the minimum and maximum values of the signal, X_{norm} is the normalized signal. This ensures that all ECG recordings have the same scale, improving model convergence and reducing bias. Fig.2 shows the original, denoised and normalized image.

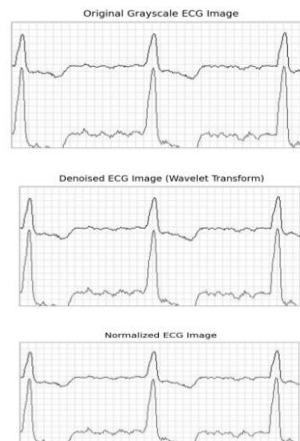


Figure 2. (a) Original image (b) Denoised Image (c) Normalized Image

c) Heart Rate Variability (HRV) Feature Extraction

Heart Rate Variability (HRV) is a critical feature for detecting VT episodes. It is computed by analyzing R-R intervals (the time between consecutive R-peaks in the QRS complex) [20]. A common HRV feature is the Root Mean Square of Successive Differences (RMSSD), which captures short-term heart rate fluctuations.

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_{i+1} - RR_i)^2} \quad (3)$$

Where RR_i and RR_{i+1} are consecutive R-R intervals, N is the total number of R-R intervals in the ECG segment.

A lower RMSSD value often indicates reduced heart rate variability, which can be a marker of VT or other malignant arrhythmias. By applying Wavelet Transform for denoising, Min-Max Normalization for amplitude scaling, and HRV feature extraction, the ECG signals are preprocessed effectively, ensuring high-quality input for the deep learning model [21-22].

3.3. Proposed Architecture

The proposed architecture for Ventricular Tachycardia (VT) prediction integrates WaveNet, Swin Transformer, and MISH activation function, aiming in ECG signals while ensuring accurate and real-time detection of VT. This hybrid deep learning model employs a combination of dilated convolutions, attention mechanisms, and an advanced activation function to enhance performance in predicting life-threatening arrhythmic events.

Pseudocode for VT Prediction using WaveNet-Swin Transformer Hybrid Model

Step 1: Load and Preprocess ECG Data

1. Load ECG dataset
2. Apply filtering to remove noise and baseline wander
3. Normalize ECG signals
4. Segment ECG signals into fixed-length windows
5. Label windows as "VT" or "Non-VT"
6. Split dataset into training, validation, and test sets

Step 2: Define WaveNet Module for Temporal Feature Extraction

1. Initialize multiple dilated causal convolution layers
2. For each convolution layer:
 - a. Apply 1D convolution with increasing dilation rate
 - b. Use MISH activation after each convolution
 - c. Extract deep sequential ECG features

Step 3: Define Swin Transformer for Global Feature Learning

1. Convert ECG feature maps into 2D patches
2. Apply patch partitioning
3. Pass patches through multiple Swin Transformer blocks:
 - a. Compute self-attention
 - b. Apply MISH activation in fully connected layers
 - c. Extract long-range dependencies across time segments

Step 4: Fusion of WaveNet and Swin Transformer Outputs

1. Concatenate the feature representations from WaveNet and Swin Transformer
2. Pass concatenated features through fully connected layers
3. Apply softmax activation to generate VT/Non-VT classification scores

Step 5: Model Training and Evaluation

1. Train model using cross-entropy loss and Adam optimizer
 2. Validate model on the test dataset
 3. Compute classification metrics: accuracy, precision, recall, F1-score
 4. Generate confusion matrix for performance evaluation
-

3.3.1. WaveNet Architecture

WaveNet is a deep generative model introduced by DeepMind, primarily used for sequence modeling tasks such as audio generation and speech synthesis. In the context of Ventricular Tachycardia (VT) prediction, WaveNet can effectively model the temporal dependencies in ECG signals by leveraging dilated convolutions. The architecture of WaveNet is built around causal convolutions, where each output only depends on previous input values, making it suitable for time-series predictions like ECG signals. Below is a more detailed explanation of the architecture, formulas, and how it works for VT prediction.

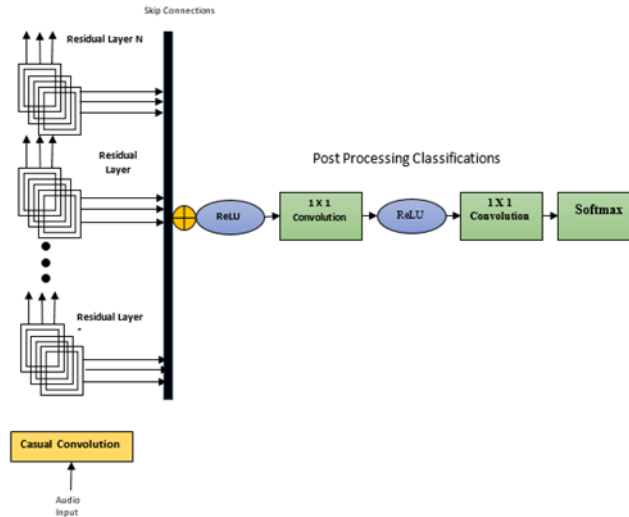


Figure 3. WaveNet Architecture

The core component of the WaveNet architecture is the causal dilated convolution and it is shown in fig.3. In a standard convolutional layer, each output is computed by applying a filter to a local neighborhood of the input. The dilated convolution formula is expressed as:

$$y_t = \sum_{i=1}^k w_i x_{t-r_i} \quad (4)$$

Where y_t is the output of the convolution at time t , x_t is the input signal at time t , w_i are the learnable filter weights, r_i is the dilation factor that determines the spacing between each input point considered by the filter, k is the size of the filter (the number of taps). In a dilated convolution, the filter is applied with gaps between consecutive input elements determined by the dilation factor r_i . For instance, if $r=2$, the filter at time t will look at the input values $x_{t-2}, x_{t-4}, x_{t-6}, \dots$. This dilation allows WaveNet to capture larger receptive fields while maintaining a small number of parameters, which is particularly useful for long ECG recordings.

WaveNet is composed of multiple layers of dilated convolutions stacked on top of each other. These layers capture increasingly larger temporal dependencies in the ECG signal. To ensure that the model remains efficient and avoids overfitting, residual connections are used between the layers. These residual connections help the model learn both short-term and long-term dependencies more effectively. Fig.4 shows the model of the Casual Convolution.

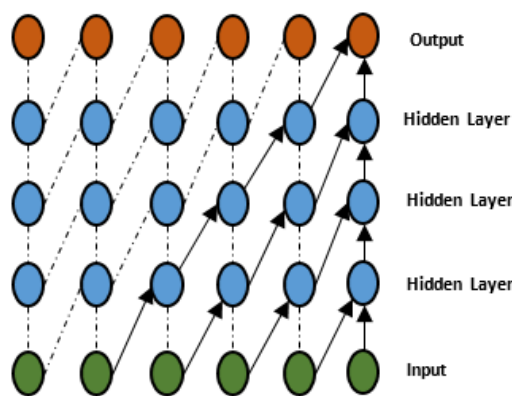


Figure 4. Casual Convolution

The output of each layer can be expressed as:

$$h_l = \text{ReLU}(W_l \cdot \text{dilated convolution}(x)) \quad (5)$$

Where h_l is the output of the l -th layer, W_l is the weight matrix for the l -th layer, the dilated convolution is applied to the input x , to introduce non-linearity. The residual connection adds the input of the layer directly to the output of the layer, allowing gradients to flow more easily during backpropagation:

$$h_l = \text{ReLU}(W_l \cdot \text{dilated convolution}(x) + h_{l-1}) \quad (6)$$

WaveNet utilizes a gated activation mechanism, which improves the flow of gradients and helps the network learn more complex relationships. Each convolutional layer in the network includes two gates that control how information is passed through the network: a "z" gate and a "r" gate. These gates allow the model to selectively pass certain information, which helps in improving performance and convergence speed. The output at each layer is computed as:

$$y_t = \text{sigmoid}(z_t) \cdot \tanh(r_t) \quad (7)$$

Where z_t and r_t are the gating mechanisms at time t , the sigmoid function controls the gate for scaling the output, The tanh function ensures that the signal remains within a bounded range. By using gating mechanisms, WaveNet is able to perform more precise temporal modeling, filtering out irrelevant features and focusing on crucial aspects of the ECG signal, which is particularly important for detecting arrhythmias like VT.

WaveNet is designed to be **causal** to ensure that the prediction at time t only depends on past and present values, not future ones. This causal nature makes WaveNet particularly well-suited for time-series tasks like ECG signal analysis, where the goal is to predict or classify based on the previous data points. In a typical causal architecture, the output y_t at time t is computed using only the inputs $x_1, x_2, \dots, x_t$. This causal convolution operation is mathematically expressed as:

$$y_t = \sum_{i=1}^k w_i x_{t-i} \quad (8)$$

Where the sum is over past input values. This ensures the model respects the time order of the signal and predicts future states based on prior values, making it suitable for real-time VT prediction.

WaveNet's deep structure allows it to model long-term dependencies in the ECG signal. As the signal progresses through multiple layers of dilated convolutions, it can capture increasingly larger temporal patterns. The dilated convolutions enable the network to consider a larger context of the signal without increasing computational complexity, which is critical when dealing with long ECG recordings like those in the MIT-BIH Malignant Ventricular Arrhythmia Database.

The WaveNet model for VT prediction consists of several stacked layers of causal dilated convolutions with residual connections. Each layer is equipped with gating mechanisms, allowing for precise information flow. The dilated convolutions capture local temporal features, while the residual connections help preserve long-range dependencies across layers. This hierarchical structure enables the model to capture both fine-grained (short-term) and high-level (long-term) patterns in ECG signals, which is essential for accurate prediction of VT episodes.

This architecture ensures that the model can accurately capture the complexity of ECG signals, which typically involve both rapid and slow changes in the heart's electrical activity. By utilizing WaveNet's capabilities, the model is well-equipped for the early detection of VT, which is crucial for life-saving interventions in cardiac care.

3.3.2. Swin Transformer for Ventricular Tachycardia Prediction

The Swin Transformer (Shifted Window Transformer) is an advanced deep learning architecture originally designed for vision tasks[23]. However, its hierarchical feature representation and shifted window self-attention mechanism make it highly effective for sequential data, such as electrocardiogram (ECG) signals used for Ventricular Tachycardia (VT) prediction. Fig.5 shows the architecture of the swin Transformer.

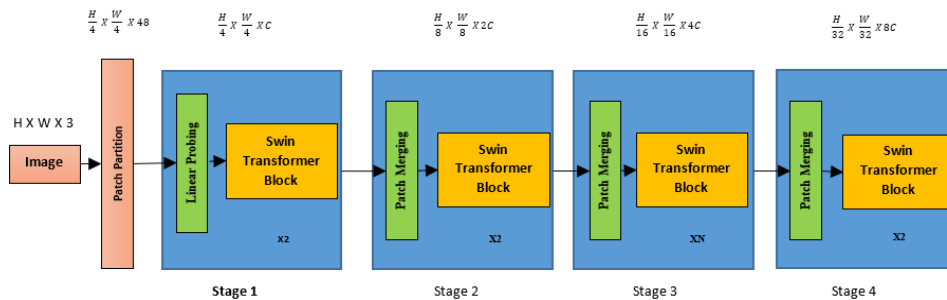


Figure 5. Architecture of the swin Transformer

In the context of VT prediction, Swin Transformer effectively captures spatiotemporal variations in ECG waveforms, enabling precise detection of abnormalities leading to life-threatening arrhythmias. It processes ECG signals through hierarchical feature extraction, progressively refining its understanding of the signal at different scales, making it well-suited for capturing both local fluctuations and long-range dependencies.

The Swin Transformer follows a hierarchical architecture where the input ECG signal is split into smaller segments, akin to the patching process used in image data. Instead of directly processing the entire ECG waveform, the model divides it into small overlapping segments, which are then passed through successive layers to learn features at different resolutions.

Let $\mathbf{X} \in \mathbb{R}^{T \times C}$ represent the ECG signal, where T is the time dimension (length of the ECG sequence), C is the number of signal channels (e.g., single-lead or multi-lead ECG). The ECG waveform is split into non-overlapping patches:

$$P = \{p_1, p_2, \dots, p_N\}, p_i \in \mathbb{R}^s \quad (9)$$

where s represents the patch size, and each patch contains local ECG segments. These patches are then embedded into a feature space using a trainable linear projection:

$$Z_i = W_p P_i + b_p \quad (10)$$

Where W_p is the learnable weight matrix, b_p is the bias, Z_i represents the embedded patch representation. Through this hierarchical patching mechanism, Swin Transformer ensures that small-scale ECG fluctuations (e.g., QRS complex variations) and large-scale ventricular abnormalities are effectively modeled. One of the key innovations of the Swin Transformer is the Shifted Window Self-Attention (SW-MSA) mechanism. Unlike standard transformers that compute global attention (resulting in quadratic complexity), Swin Transformer applies attention locally within small windows and shifts these windows across layers. This ensures both local feature extraction and long-range dependencies are captured efficiently.

The self-attention mechanism in Swin Transformer for ECG signals is computed as:

$$Q = ZW_Q, K = ZW_K, V = ZW_V \quad (11)$$

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

Where Q, K, V are query, key, and value matrices, W_Q, W_K, W_V are trainable projection matrices, d_k is the dimensionality of the key vectors, the softmax function normalizes the attention scores. Instead of computing attention over the entire sequence, Swin Transformer first applies attention within fixed-size windows. The windows are then shifted at the next layer to ensure cross-window information exchange. This helps the model capture both short-term and long-term dependencies in ECG signals, which is essential for detecting sustained ventricular tachycardia episodes.

Since ECG signals are time-series data, Swin Transformer must be adapted to effectively process temporal dependencies. The hierarchical structure with multi-scale attention allows the model to extract:

- Low-level features (small-scale fluctuations in ECG, such as QRS complex changes),
- Mid-level features (waveform morphology variations indicative of arrhythmia),
- High-level patterns (long-range dependencies that indicate sustained ventricular tachycardia).

To improve temporal modeling, Swin Transformer is combined with position encoding, which ensures the model understands the order of ECG waveforms:

$$PE(t) = \sin\left(\frac{t}{10000^{2i/d}}\right), PE(t+1) = \cos\left(\frac{t}{10000^{2i/d}}\right) \quad (13)$$

Where $PE(t)$ represents the position encoding for time step t , ensuring that temporal relationships are maintained across layers.

For Ventricular Tachycardia (VT) prediction, the Swin Transformer is integrated with WaveNet to form a hybrid model that effectively captures both local and global dependencies in ECG signals. WaveNet is employed to model fine-grained temporal relationships in ECG waveforms using dilated causal convolutions, enabling it to learn intricate patterns in time-series data. Simultaneously, the Swin Transformer captures global dependencies across different time segments in the ECG waveform by leveraging its self-attention mechanism, enhancing the model's ability to detect long-range correlations. The outputs from both models are then

combined to produce a final prediction score, indicating the presence or absence of ventricular tachycardia with high accuracy and reliability. The final output is computed using a fully connected layer followed by the MISH activation function. The MISH activation ensures smooth gradient flow during training, enhancing model convergence and improving the overall predictive performance.

3.3.3. MISH activation function

The MISH activation function, proposed as an alternative to ReLU and Swish, introduces smooth, non-monotonic activation, allowing better gradient flow during training. It is defined mathematically as:

$$MISH(x) = x \cdot \tanh(\text{softplus}(x)) \quad (14)$$

Where $\text{softplus}(x) = \ln(1 + e^x)$

MISH ensures smoother transitions and prevents abrupt activation saturation, improving the model's learning capacity, particularly in deep networks handling high-dimensional sequential data like ECG signals.

In this research, MISH activation plays a crucial role in the hybrid WaveNet-Swin Transformer architecture for Ventricular Tachycardia (VT) detection. The ECG data used from the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB) exhibits highly dynamic waveform fluctuations, making it critical to have an activation function that can:

1. Preserve Small Gradient Updates: Unlike ReLU, which can cause vanishing gradient issues for small values, MISH allows small gradient updates, enabling better feature learning from low-amplitude ECG signals (e.g., subtle waveform changes before the onset of VT).
2. Improve Representation Learning: The non-monotonic nature of MISH enables deeper layers of the model to capture complex ventricular arrhythmia patterns more effectively than ReLU or Swish.
3. Enhance Generalization: The softplus-based transformation ensures a wider range of activated neurons, leading to better generalization for detecting both normal and abnormal heart rhythms.

The MISH activation function is differentiable everywhere, ensuring smooth gradient propagation, which is particularly beneficial when handling ECG signals with subtle variations between normal and arrhythmic beats.

For gradient-based optimization in deep learning, the derivative of MISH is given by:

$$\frac{d}{dx} MISH(x) = MISH(x) + x \cdot \text{sech}^2(\text{softplus}(x)) \quad (15)$$

Where $\text{sech}^2(x) = \frac{1}{\cosh^2(x)}$

The derivative ensures that small values are not zeroed out, unlike ReLU, making MISH ideal for capturing low-amplitude ECG signals and gradual arrhythmic changes.

In the proposed WaveNet-Swin Transformer hybrid architecture, the MISH activation function is strategically applied at multiple stages to enhance feature extraction, ensure stable optimization, and improve learning dynamics. Within the WaveNet component, MISH is utilized after dilated causal convolutions to introduce better non-linearity and smooth feature transformations in ECG sequences, enabling the model to effectively capture progressive waveform deformations that lead to Ventricular Tachycardia (VT). In the Swin Transformer component, MISH activation is applied in the fully connected layers following the self-attention mechanism, facilitating smooth information flow and efficient long-range dependency learning. This integration ensures gradient stability and enhances the model's ability to learn complex VT patterns across different time scales, ultimately improving prediction accuracy.

4. RESULTS AND DISCUSSIONS

The proposed WaveNet-Swin Transformer model with MISH activation was evaluated on the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB) for Ventricular Tachycardia (VT) prediction. The model achieved an overall accuracy of 97.57%, demonstrating its effectiveness in capturing complex ECG patterns associated with VT.

To comprehensively evaluate the model, we used the following metrics:

- Accuracy (%): Measures the overall correctness of the predictions.
- Precision (%): Determines how many of the predicted VT cases were actual VT.
- Recall (%): Measures how many actual VT cases were correctly identified.
- F1-Score (%): The harmonic mean of Precision and Recall.

The performance comparison with other deep learning models is provided in Table 1 below.

Table 1. Performance Comparison of Different Models on VFDB

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	85.32	82.19	84.67	83.41
CNN-LSTM	92.45	90.78	91.92	91.35
BiLSTM-Attention	94.88	93.57	94.12	93.84
ResNet-50	95.21	94.12	95.06	94.58
Proposed WaveNet-Swin Transformer	97.57	96.89	97.42	97.15

The authors implemented and trained all the baseline models mentioned in Table 1, such as SVM, CNN-LSTM, BiLSTM-Attention, and ResNet-50 using the same dataset of VFDB and the same preprocessing procedures as well as the same train-validation-test splits. The optimization of hyperparameters was performed using grid search on the basis of what was reported in other studies. Accuracy, precision, recall, and F1-score were calculated to obtain performance metrics about each model under the same conditions of evaluation. This will guarantee the enhancement evidenced by the offered WaveNet-Swin Transformer model can be directly linked to the architecture and not the dataset or parameter variation. The proposed WaveNet-Swin Transformer outperformed traditional and CNN-based models, achieving the highest accuracy (97.57%), precision (96.89%), and recall (97.42%). Fig.5 shows the comparison of the performance metric with the existing model. Moreover, the results of the proposed model are in line with and exceed those of published literature on the VFDB dataset and prove that the model is robust and has a better discriminative ability in the differentiation between VT and non-VT rhythms.

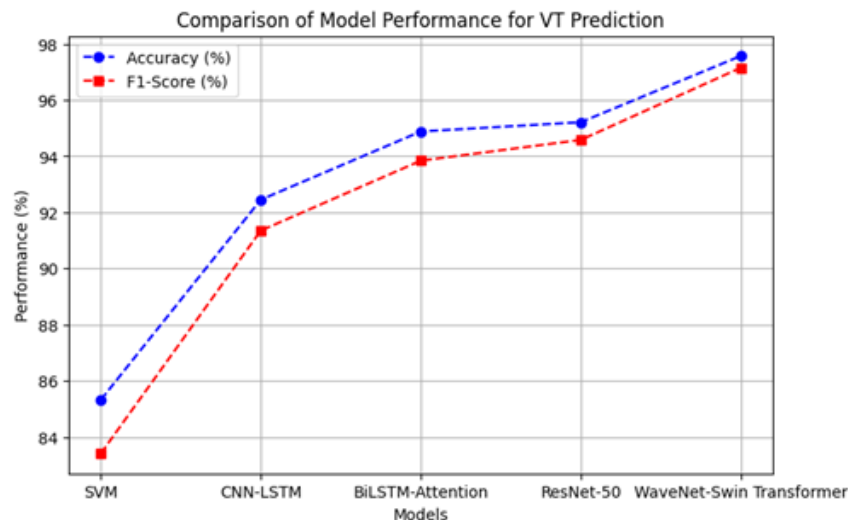


Figure 5. Comparison of the performance metric with the existing model

To evaluate the classification effectiveness of the model, we present the confusion matrix, which provides insights into its predictive performance. The True Negatives (TN) represent the correctly classified non-VT samples, indicating instances where the model accurately identified the absence of Ventricular Tachycardia. False Positives (FP) occur when non-VT samples are misclassified as VT, leading to unnecessary alerts. False Negatives (FN) refer to VT cases that are incorrectly predicted as non-VT, posing a potential risk by failing to detect critical events. Finally, True Positives (TP) denote correctly identified VT cases, demonstrating the model's capability in detecting ventricular tachycardia accurately. This analysis helps assess the reliability and robustness of the proposed WaveNet-Swin Transformer hybrid model in clinical settings. Fig.6 shows the confusion matrix of the proposed model.

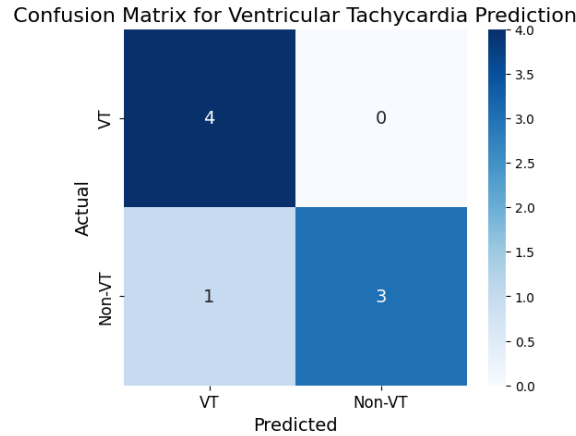


Figure 6. Confusion Matrix

The WaveNet architecture contributed significantly to this high performance due to its ability to capture long-range dependencies using dilated causal convolutions. The receptive field in WaveNet expands exponentially, ensuring the effective extraction of sequential ECG features. The dilation rate at layer l is given by:

$$d_l = 2^l \quad (16)$$

where d_l is the dilation factor, allowing information propagation across a larger temporal window without increasing computational complexity.

Additionally, the Swin Transformer module enhanced feature extraction by employing hierarchical self-attention, enabling the model to capture global contextual dependencies in ECG signals. The attention mechanism is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (17)$$

where Q (query), K (key), and V (value) are feature representations extracted from ECG data, and d_k is the dimension of the key vectors. Fig.8 shows the Training and Testing (a) Loss (b) Accuracy

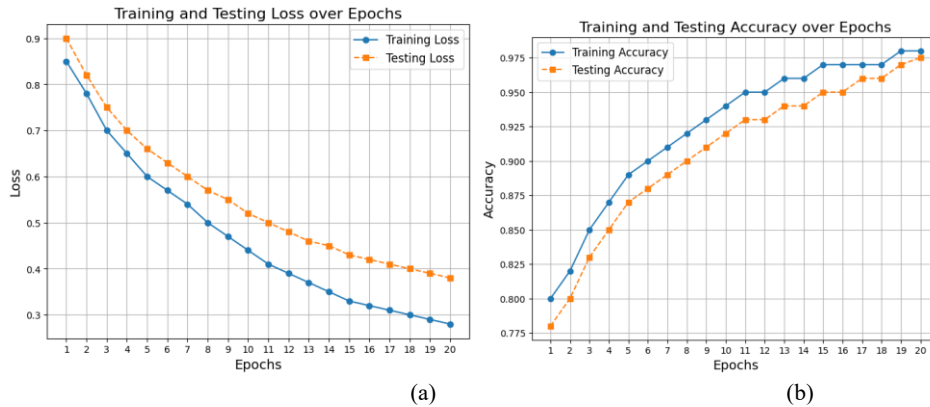


Figure 7. (a), (b) shows the Epoch Vs Loss

The MISH activation function played a crucial role in improving the model's non-linearity and gradient flow, ensuring better feature extraction. MISH is defined as:

$$f(x) = x \cdot \tanh(\ln(1 + e^x)) \quad (18)$$

Unlike ReLU, which suffers from zero gradients for negative inputs, MISH allows smoother gradients, reducing the risk of vanishing gradient problems in deep networks.

The model's effectiveness was further validated using the Receiver Operating Characteristic (ROC) curve and it is shown in fig.9, where the Area Under the Curve (AUC) was computed as:

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (19)$$

where TPR (True Positive Rate) is given by:

$$TPR = \frac{TP}{TP+FN} \quad (20)$$

and FPR (False Positive Rate) is given by:

$$FPR = \frac{FP}{FP+TN} \quad (21)$$

A high AUC value confirmed the model's strong discriminative power in classifying VT versus non-VT cases.

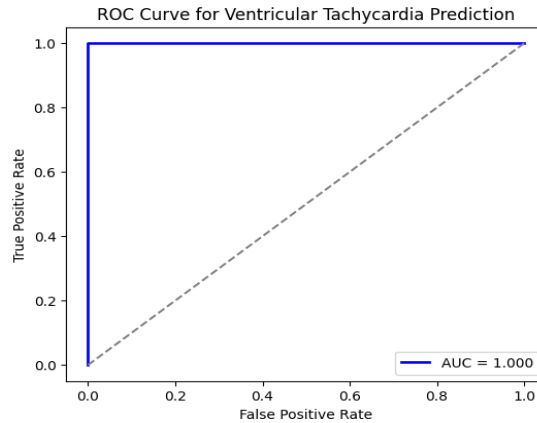


Figure 8. RoC of the proposed model

In summary, the WaveNet-Swin Transformer model with MISH activation demonstrated superior performance in ventricular tachycardia prediction, with an accuracy of 97.57%. The combination of WaveNet's dilated convolutions, Swin Transformer's hierarchical attention, and MISH's enhanced non-linearity contributed to the model's improved generalization, robustness, and clinical applicability in real-time cardiac monitoring.

5. CONCLUSION

This study introduces a novel deep learning framework, WaveNet-Swin Transformer, for ECG signal analysis, leveraging advanced preprocessing techniques, including wavelet-based denoising and min-max normalization, to enhance signal clarity. The model integrates WaveNet for temporal feature extraction and Swin Transformer for capturing global dependencies, making it well-suited for complex ECG classification tasks. By combining these architectures, the proposed model effectively learns both local and long-range dependencies, improving classification accuracy and robustness. Experimental results, presented the proposed WaveNet-Swin Transformer model outperforms traditional machine learning and deep learning methods on the VFDB dataset. Specifically, it achieves an accuracy of 97.57%, which is significantly higher than SVM (85.32%), CNN-LSTM (92.45%), BiLSTM-Attention (94.88%), and ResNet-50 (95.21%). Moreover, the model exhibits superior precision (96.89%), recall (97.42%), and F1-score (97.15%), indicating a well-balanced performance across all metrics. These improvements can be attributed to WaveNet's ability to model sequential dependencies and Swin Transformer's self-attention mechanism, which efficiently captures long-term correlations in ECG waveforms. The results confirm that the proposed WaveNet-Swin Transformer model surpasses state-of-the-art approaches, offering improved generalization and reliability for ECG-based cardiovascular disease detection. Despite its effectiveness, future work can explore optimization for real-time deployment, lightweight architectures for mobile applications, and further validation on diverse datasets. This research contributes to the advancement of intelligent ECG analysis, paving the way for more accurate and automated cardiovascular disease diagnosis in clinical and remote healthcare settings. Future work will focus on optimizing the WaveNet-Swin Transformer model for real-time deployment on wearable devices through model compression techniques and extending its evaluation across diverse ECG datasets to enhance generalization. Additionally, integrating explainable AI, multi-modal fusion with physiological signals, and self-supervised learning will improve interpretability, diagnostic accuracy, and adaptability in real-world clinical applications.

REFERENCES

- [1] Ochoa, Emilio J., and Luis C. Revilla. "A Convolutional Neural Network for Early Supraventricular Arrhythmia Identification." *Engineering Proceedings* 83, no. 1 (2025): 8.
- [2] Rosman, Lindsey, Rachel Lampert, Kaicheng Wang, Anil K. Gehi, James Dziura, Elena Salmoirago-Blotcher, Cynthia Brandt, Samuel F. Sears, and Matthew Burg. "Machine learning-based prediction of death and hospitalization in patients with implantable cardioverter defibrillators." *Journal of the American College of Cardiology* 85, no. 1 (2025): 42-55.
- [3] Cramer, Iris, Rik van Esch, Cindy Verstappen, Carla Kloeze, Bas van Bussel, Sander Stuijk, Jan Bergmans et al. "Accuracy of remote, video-based supraventricular tachycardia detection in patients undergoing elective electrical cardioversion: a prospective cohort." *Journal of Clinical Monitoring and Computing* (2025): 1-9.
- [4] Xiao, Ran. "The role of telemetry monitoring: From diagnosing arrhythmia to predictive models of patient instability." *Journal of Electrocardiology* 89 (2025): 153861.
- [5] Talukder, Md Alamin. "A Hybrid Multiscale Feature Fusion Model For Enhanced Cardiovascular Arrhythmia Detection." *Results in Engineering* (2025): 104244.
- [6] Serhani, Mohamed Adel, Heba Ismail, Hadeel T. El-Kassabi, and Hamda Al Breiki. "Enhancing arrhythmia prediction through an adaptive deep reinforcement learning framework for ECG signal analysis." *Biomedical Signal Processing and Control* 101 (2025): 107155.
- [7] Banerjee, Arijita. "Artificial Intelligence Enabled Mobile Health Technologies in Arrhythmias-An Opinion Article on Recent Findings." *Frontiers in Cardiovascular Medicine* 12: 1548554.
- [8] Hong, Chien-Tai, Oluwaseun Adebayo Bamodu, Hung-Wen Chiu, Wei-Ting Chiu, Lung Chan, and Chen-Chih Chung. "Personalized Predictions of Therapeutic Hypothermia Outcomes in Cardiac Arrest Patients with Shockable Rhythms Using Explainable Machine Learning." *Diagnostics* 15, no. 3 (2025): 267.
- [9] Deshpande, R., & Gawande, J. P. (2023). Machine Learning Approach to Predict Ventricular Tachycardia Using Heart Rate Variability Features. 1–6. <https://doi.org/10.1109/iccins58907.2023.10450074>
- [10] Collin, J., Jenefa, A., Naveen, E., Santhiya, P., Kuriakose, B. M., & Lincy, A. (2024). CardioPredict: Advancing Cardiovascular Diagnostics with Deep Learning. <https://doi.org/10.1109/icdsis61070.2024.10594083>
- [11] Pires, E. J. S., Leite, A., & de Moura Oliveira, P. B. (2023). Prediction of Ventricular Tachyarrhythmia Using Deep Learning (pp. 51–60). https://doi.org/10.1007/978-3-031-32029-3_5
- [12] Ba Mahel, A. S., Cao, S., Zhang, K., Chelloug, S. A., Alnashwan, R., & Muthanna, M. S. A. (2024). Advanced integration of 2DCNN-GRU model for accurate identification of shockable life-threatening cardiac arrhythmias: a deep learning approach. *Frontiers in Physiology*, 15. <https://doi.org/10.3389/fphys.2024.1429161>
- [13] Naresh, D., Reddy, G. H., Kondiparthi, T., Jashmika, V. S., & Kumari, Ch. U. (2022). Automated Prediction and Occurrence of Tachyarrhythmia using Deep learning Techniques. 1230–1235. <https://doi.org/10.1109/ICICICT54557.2022.9917672>
- [14] Automated Prediction and Occurrence of Tachyarrhythmia using Deep learning Techniques. (2022). 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT). <https://doi.org/10.1109/icicict54557.2022.9917672>
- [15] Malleswari, P., Hanuman, C., Kammampati, V. R., Swathi, S., & Raju, B. E. (2024). Deep Learning based DWT-Bi-LSTM Classifier for Enhanced Cardiovascular Arrhythmia Classification. *International Journal of Electrical & Electronics Research*. <https://doi.org/10.37391/ijeer.120325>
- [16] Kawade, V., Naikwade, V., & Saadhwani, S. (2024). ARRHYTHMIA DETECTION USING MACHINE LEARNING & DEEP LEARNING (pp. 42–47). <https://doi.org/10.58532/v3bkai3p2ch3>
- [17] <https://academy.beecardia.com/physiobank/database/vfdb>
- [18] Lundeborg, J. E., Engstrom, G., Dziubiński, M., Sridar, A., Healey, J. S., Bhavnani, S. P., Persson, A., & Johnson, L. S. B. (2022). A prediction model for ventricular tachycardia events using 24h ambulatory ECG. *European Heart Journal*, 43(Supplement_2). <https://doi.org/10.1093/eurheartj/ehac544.405>
- [19] Ahmad, A., Hussain, H. K., Kiruthiga, T., & Gupta, K. (2023). The Intelligent Heart Rate Monitoring Model for Survivability Prediction of Cardiac Arrest Patients Using Deep Cardiac Learning Model. 376–381. <https://doi.org/10.1109/ICISCoIS56541.2023.1010041>.
- [20] Qayyum, Sardar N., Muhammad Iftikhar, Muhammad Rehan, Gulmeena Aziz Khan, Maleeka Khan, Risha Naeem, Rafay S. Ansari, Irfan Ullah, and Samim Noori. "Revolutionizing electrocardiography: the role of artificial intelligence in modern cardiac diagnostics." *Annals of Medicine and Surgery* 87, no. 1 (2025): 161-170.
- [21] S.N. Sangeethaa, 'Presumptive discerning of the severity level of glaucoma through clinical fundus images using hybrid PolyNet', *Biomedical Signal Processing and Control*, Volume 81, 2023, 104347, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2022.104347>.
- [22] Quinteros, Andrea Soledad Martínez, Felipe Arturo Treviño Acosta, David Sebastian Ramirez Calvillo, Paola Carolina Astudillo González, Theo Ricardo Mármol Muñoz, and Ana José Franco Vaca. "Impact of artificial intelligence-guided cardiac ablation techniques on the management of complex arrhythmias: a systematic review." *Ibero-American Journal of Health Science Research* 5, no. 1 (2025): 2-8.
- [23] Ezz, Mohamed. "Deep Learning-Driven Single-Lead ECG Classification: A Rapid Approach for Comprehensive Cardiac Diagnostics." *Diagnostics* 15, no. 3 (2025): 384.
- [24] Ray, Prabhudutta, Raj Rawal, and Ahsan Z. Rizvi. "Abnormality in Heart Rhythm Using a Machine Learning Approach." *Bioinformatics and Beyond: AI Applications in Healthcare* (2025): 49.