ASNet: Attention-stacked deep network for ECG-based cardiac rhythm classification

Vishnu Venkatesh [a], S. Angalaeswari [b], Aravindkumar Sekar [a] [iD],*

[a] *School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Chennai, Tamil Nadu, India*
[b] *School of Electrical Engineering (SELECT), Vellore Institute of Technology, Chennai, Tamil Nadu, India*

ARTICLE INFO

ABSTRACT

Heart disease remains one of the leading causes of worldwide mortality, and early detection is crucial for effective treatment. Automated diagnostic systems can significantly assist healthcare professionals by providing rapid, reliable, and scalable evaluations of patient data. In this proposal, a novel Attention-based Stacked Network model (ASNet) is introduced for cardiac rhythm classification. ASNet consists of two sub-networks: Feature Transformer and an Attentive Transformer. The Feature Transformer helps to extract the complex feature dependencies using multi-head attention mechanisms and feed-forward encoding layers. Attentive Transformer is associated with the ASNet model to enhance interpretability and performance by applying positional encoding, masked self-attention, and a cross-attention mechanism for refined contextual understanding. The performance of our proposed model is evaluated using various metrics, and it achieves an overall accuracy of 91.31%. Furthermore, the proposed model is analyzed with other benchmark datasets, and the obtained results show that the ASNet model outperforms when compared with other state-of-the-art models.

1. Introduction

In the early days, diagnosing cardiovascular-related conditions mainly relied on the manual interpretation of electrocardiogram (ECG) signals by physicians. Although ECGs are critical diagnostic tools, they rely heavily on human interpretation. This dependence often causes variability in accuracy. For example, physician accuracy ranged from a median of 54% before training to 67% after training. Even experienced cardiologists made occasional errors [1]. Observer agreement was only 'substantial' ($\kappa \approx 0.70$), and less experienced clinicians were influenced by computerized interpretations when deciding on thrombolysis eligibility [2]. Misdiagnosis of ST-Elevation Myocardial Infarction (STEMI) is more frequent among less experienced staff, whereas cardiologists showed higher accuracy, especially when interpreting many ECGs weekly [3]. At the same time, early computerized rhythm interpretation had only $approx$88% accuracy, missing about 12% of cases [4]. The lack of automated tools and error-prone analysis methods contributed heavily to elevated mortality rates, where timely intervention could have prevented fatal outcomes.

As shown in Fig. 1, the major components of the ECG signal are the P-wave, QRS complex, and T-wave. The P-wave represents atrial depolarization, which initiates atrial contraction. Following the P-wave, the PR segment corresponds to the electrical conduction from the atria to the ventricles. The PR interval, which includes both the P-wave and the PR segment, measures the total time required for the electrical impulse to travel from the atria to the ventricles.

The QRS complex in an ECG represents the rapid depolarization and contraction of the ventricles, consisting of the Q wave (small negative deflection indicating septal depolarization), R wave (largest positive deflection representing the depolarization of the ventricular myocardium, especially the left ventricle), and S wave (negative deflection following the R wave, marking the final phase of depolarization). The ST segment follows and reflects the period between ventricular depolarization and repolarization, with ST elevation often indicating STEMI or pericarditis, and ST depression suggesting ischemia or digoxin toxicity. The T wave represents ventricular repolarization, typically positive but can be inverted in conditions like ischemia or hyperkalemia. The QT interval, spanning from the start of the Q wave to the end of the T wave, represents the total time for depolarization and repolarization, and is corrected using the QTc formula. Prolonged QT intervals are associated with arrhythmias, including Long QT Syndrome (LQTS), which can lead to torsades de pointes or sudden cardiac death. Monitoring these components is essential for diagnosing cardiac conditions, from ischemic heart disease to arrhythmias, and guiding clinical interventions [5,6].

---

* Corresponding author.
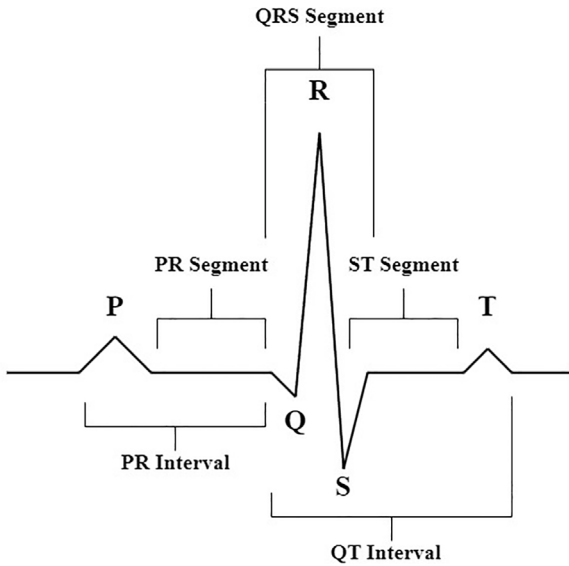*E-mail address:* aravindkumar.s@vit.ac.in (A. Sekar).

Fig. 1. Illustrates the ECG signal with labeled components: P-wave representing atrial contraction, QRS complex indicating ventricular contraction, ST segment illustrating the early phase of ventricular contraction, and T-wave denoting ventricular recovery following contraction.

In recent years, with the emergence of Artificial Intelligence (AI) computing technologies, researchers began exploring traditional machine learning (ML) models for ECG data analysis. Techniques such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Trees, and Random Forests demonstrated promising results in classifying cardiac rhythm conditions from numerical or waveform-derived ECG features. However, these models relied on hand-crafted feature extraction and statistical preprocessing methods, which limited their scalability and adaptability across diverse patient populations. Recently, deep learning techniques have been increasingly applied to ECG signal analysis, with Deep Neural Networks (DNNs), particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), providing a powerful alternative by learning hierarchical and temporal features directly from raw or minimally processed data. These models achieve remarkable performance in detecting subtle abnormalities and classifying various heart conditions, surpassing traditional methods in capturing complex patterns that are difficult to analyze manually. However, deep learning approaches are computationally intensive, often requiring high-performance GPUs, extensive training time, and large annotated datasets to avoid overfitting and ensure generalization. Despite these challenges, the benefits typically outweigh the costs, as deep learning consistently outperforms classical machine learning techniques in ECG analysis.

A novel deep learning architecture is implemented to address some of these limitations by converting ECG images into numerical signals by utilizing a TabNet-based framework as a base model for effective classification. Our model aims to improve the accuracy in four cardiac conditions: Acute Myocardial Infarction (Myocardial Infarction), Historical Infarction Events (History of Myocardial Infarctions), Cardiac Arrhythmia (Abnormal Heartbeats), and Normocardiac (Normal). It also reduces the computational overhead and improves interpretability. The approach bridges the gap between clinical needs and technical advances, offering a practical solution to modern ECG-based cardiac rhythm classification.

The rest of the article is organized as follows. Section 2 highlights the survey of various deep learning models and optimization techniques for the detection and classification of cardiac rhythm through the ECG signal. Section 3 explains the proposed Attention-based Stacked Network model for accurate cardiac rhythm classification. Section 4

presents the experimental setup, and Section 5 highlights the analysis of results obtained for the proposed model. Conclusion of the proposed work is given in Section 6.

2. Related works

2.1. Ensemble learning based approaches for heart disease prediction

The Hybrid Random Forest with Linear Model (HRFLM) approach in [7] combines Random Forest (RF) with Backpropagation Neural Networks (BPNN) and SVM for heart disease prediction. Using feature selection and classification techniques, it achieves 89.01% accuracy on the Cleveland dataset (303 records, 14 attributes), demonstrating the effectiveness of ensemble methods in diagnosis. In [8], Passive Aggressive (PaRSEL), a stacking-based model, integrates Passive Aggressive Classifier (PAC), Ridge Classifier (RC), Stochastic Gradient Descent Classifier (SGDC), and eXtreme Gradient Boosting (XGBoost) as base classifiers, with LogitBoost as a meta-classifier. Employing Recursive Feature Elimination (RFE), Linear Discriminant Analysis (LDA), and Factor Analysis (FA) for feature selection, alongside eight class-balancing techniques such as Proximity Weighted Random Affine Shadow sampling (ProWRAS) and Synthetic Minority Oversampling Technique (SMOTE), PaRSEL achieves 97% accuracy, 80% F1 score, over 90% precision, and 67% recall, highlighting the effectiveness of stacking models in diagnosis.

Deep learning ensemble models, Ensemble-Based Cardiovascular Disease Detection (EnsCVDD)-Net and Blending-Based Cardiovascular Disease Detection (BICVDD)-Net, are explored in [9] for heart disease classification. Using Adaptive Synthetic (ADASYN) sampling for class balancing and Point Biserial Correlation Coefficient (PBCC) for feature selection, BICVDD-Net achieves 91% accuracy, 96% precision, and a 91% F1 score, while EnsCVDD-Net achieves 88% accuracy, demonstrating the value of hybrid deep learning in diagnosis. In [10], five classifiers are evaluated using tenfold cross-validation with hyperparameter tuning via Randomized Search CV and Grid Search CV, achieving 96% accuracy, 0.98 recall, and 0.96 ROC-AUC score, maintaining 96.88% accuracy on an independent set, underscoring the efficacy of ensemble learning with validation. A hybrid approach combining feature engineering and ensemble learning in [11] integrates RFE with SVM for feature selection and uses XGBoost with Bayesian optimization, showing high accuracy and precision across datasets ranging from 333 to 1,025 samples with 9–32 features. Similarly, [12] combines Principal Component Analysis (PCA) and LDA for feature selection, training models such as Random Forest, KNN, Naïve Bayes, and SVM with Stacking, Bagging, Voting, and Boosting techniques; the Bagging Ensemble Machine Learning Algorithm (BEMLA) achieves 92% accuracy on Cleveland and 97% on Framingham and Indicators datasets, demonstrating the strength of ensemble models in diagnosis.

Advanced pre-processing and feature extraction further enhance prediction performance. In [13], min–max normalization, Higher Order Statistics (HOS), entropy-based features, and Mutual Information (MI) are combined with a hybrid CNN-DeepMaxout model with score-level fusion, achieving 97.21% accuracy on the Cleveland dataset with 90% training data and a Matthews Correlation Coefficient (MCC) of 0.9326 at 70% training, indicating strong predictive performance with low false positive and false negative rates. Using a mean-based splitting technique, [14] partitions data into subsets modeled with Classification and Regression Trees (CART), and a homogeneous ensemble with an accuracy-weighted aging classifier achieves 93% and 91% accuracy on the Cleveland and Framingham datasets, highlighting the value of structured ensemble learning. Addressing imbalanced medical data, [15] employs SMOTE for synthetic Coronary Heart Disease (CHD) cases and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for unsupervised segmentation before training a Random Forest classifier. Using 33,831 records from six Chinese counties (2006–2008) with only 78 CHD cases, the model achieves 99% accuracy, 0.8 sensitivity, 0.998 specificity, and 0.91 AUC, underscoring the power of clustering and ensemble learning for high-precision heart disease prediction.

2.2. Deep neural network-based heart disease prediction and classification

Leveraging CNNs for feature extraction and LSTMs for temporal dependencies, [16] proposes a Hybrid Deep Neural Network (HDNN) achieving 98.86% accuracy on the Cleveland HD and larger combined datasets. For early detection, [17] presents the Early Detection (ED) CNN model with multilayer perceptrons and regularization, achieving 99.1% precision on the UCI dataset. Graph-based learning and temporal modeling in [18] with the Optimal Scrutiny Boosted Graph Convolutional (O-SBGC) LSTM model attain 94.8% accuracy, 94.9% sensitivity, and 91.9% specificity. Combining autoencoders with DenseNet, [19] achieves 99.67% mean and 99.99% test accuracy on UCI Cleveland using data augmentation. Integrating XGBoost and Bi-LSTM, [20] achieves 99.4% accuracy on Framingham, Cleveland, and MIMIC-III datasets. Attention-based BiLSTM models with XGBoost and grid search optimization in [21] reach 99.99% accuracy on Amazon SageMaker. Optimization-driven hybrid models like DPA-RNN and LSTM in [22] using GA, PSO, ABO, and GSA achieve 99.21% and 93.56% accuracy on the Cleveland and Clinical Heart Failure datasets. Finally, [23] combines HKMFCM segmentation, ABC-based feature selection, and an Attention-infused BiRecurrenTwin Network with BiGRU and BiLSTM, attaining 96.09% accuracy, 97% ROC, and a 7.75% improvement over traditional methods on UCI datasets, highlighting the effectiveness of hybrid and attention-driven deep learning for accurate heart disease prediction.

2.3. Optimization technique based feature selection for heart disease classification

Integrating LASSO-based feature selection with Gradient Boosting, [24] introduces the Hyperparameter Gradient Boosting (HypGB) system, achieving 97.32% accuracy on the Cleveland dataset through hyperparameter tuning. Using LSTM networks, [25] analyzes 2,621 medical records from UAE hospitals, with PCA for feature selection, achieving 71.5% accuracy for CHD, 84.4% for Dyspnea, and 90.6% for combined symptoms. RSA-RF in [26] employs a Random Search Algorithm for feature selection on the Cleveland heart failure dataset, achieving 93.33% accuracy, outperforming conventional Random Forest by 3.3%. In [27], six classifiers, including DT, RF, and KNN, are optimized via GridSearchCV, with dimensionality reduction using PCA, TSNE, and LDA, achieving 99.8% accuracy for DT and RF, and 99.02% for KNN on Cleveland data. Studies [28,29] implement feature selection methods such as RFE, PCA, and univariate selection, applying models like DT, RF, SVM, and AdaBoost, achieving up to 99.12% accuracy with cross-validation, ensuring robustness. Using Boruta for feature selection, [30] identifies key clinical factors from the Cleveland Clinic dataset, with logistic regression achieving 88.52% accuracy and improved F1, precision, and recall, demonstrating the effectiveness of feature-driven machine learning approaches.

2.4. Privacy-preserving and IoT-based approaches in heart disease prediction

A heart disease prediction model (HDPM) [31] uses DBSCAN for outlier detection and a hybrid SMOTE-ENN to handle class imbalance. XGBoost achieves 95.90% and 98.40% accuracy on Statlog and Cleveland datasets, with high precision, recall, and F1 scores. The Recursion Enhanced Random Forest with an Improved Linear Model (RERF-ILM) [32] integrates SVM, KNN, and RF, optimized with PSO and ACO, achieving 96.6% accuracy, 96.8% stability, and 96.7% F-measure on UCI data. A Two-Layer Voting (TLV) framework [33] uses hard and soft voting with feature selection (ANOVA F-test, Chi-squared, Mutual Information) and GridSearchCV-tuned MLP, DT, SVM, and RF, achieving 99.3% and 88.9% accuracy on Kaggle and UCI datasets. DeFedHDP [34] is a decentralized online aggregation method, enabling hospitals to share model parameters in a peer-to-peer setup while maintaining privacy. Decentralized Federated Heart Disease Prediction (DeFedHDP) [35] addresses fragmented data across institutions. LSTM models capture temporal dependencies for cardiology, predicting blood pressure and hospital stay using 12,036 surgical cases, outperforming KNN and SVM, but limited public datasets pose evaluation challenges. A Hierarchical Bayesian Model [36] improves personalized survival prediction by sampling location and shape of distributions, evaluated on UNOS and public datasets using C-index and Brier Score. An Intelligent Decision Support System (IDSS) [37] with Fused Layered Bidirectional LSTM analyzes HRV and IoT data in real-time, achieving 94.17% accuracy, 83.18% sensitivity, and 96% specificity. A study [38] comparing FRS and PAR in rural Ningxia (2,209 individuals) shows limited discrimination (C-index 0.65–0.67) and poor calibration, underestimating CHD by 22%–46%. An IoT-based system [39] collects real-time patient data, preprocessing with Median Studentized Residual and selecting features via Harris Hawk Optimization (HHO).

A Modified Deep LSTM aggregates an Improved Spotted Hyena Optimization (ISHO) for classifying data, achieving 98.01% accuracy on the Hungarian dataset. Additional metrics include 98.33% sensitivity, 97.33% specificity, a 98.1 F-score, and an execution time of 11.2 s, demonstrating strong diagnostic performance. A study on heart disease diagnosis evaluated Naïve Bayes, SVM, and XGBoost using a private dataset [40] (200 samples from Egyptian hospitals) and a public dataset. XGBoost achieved the highest accuracy of 97.57%, demonstrating superior predictive performance. Model evaluation included sensitivity, specificity, precision, and F1-score, validating its effectiveness in heart disease prediction. An IoT-enabled heart disease prediction framework [41] utilizes a Finch Hunt Optimization (FHO)-modified BiLSTM classifier for real-time patient monitoring and diagnosis. Key features like total harmonic distortion, heart rate, and entropy enhance predictive accuracy, while FHO optimizes BiLSTM hyperparameters. Evaluated on Coronary Artery Disease Biomarker (CADB) dataset (1,602 samples), it achieved 97.24% accuracy, 95.17% F1-score, 96.52% precision, and 93.86% recall, with similar results in 8-fold cross-validation.

Based on the above analysis, deep learning models outperform traditional methods. The proposed work introduces an attention-based mechanism to extract relevant features from ECG signals for classifying cardiac rhythms.

3. Proposed Attention-based Stacked Network (ASNet) model

The proposed Attention-based Stacked Network identifies different cardiac rhythms from ECG signals of heart-disease patients. Initially, raw ECG images are preprocessed, and heart rate variability (HRV) features are extracted. These features are then passed to ASNet, which combines Feature Transformers and an Attentive Transformer. The Feature Transformer uses an encoder–decoder design. The encoder first normalizes and embeds the ECG features, then uses multi-head self-attention and feed-forward layers to capture both local and global patterns. The decoder then combines and aligns these features using cross-attention and masked multi-head attention to highlight important signal parts and refine the features for classification. The Attentive Transformer gives different levels of importance to the features using feature attention, cross-feature attention, and masked multi-head self-attention to improve understanding and classification accuracy. Finally, the refined features pass through fully connected layers to accurately classify multiple types of ECG rhythms. Fig. 2 represents the overall architecture of the proposed ASNet model.

3.1. Pre-processing

Pre-processing helps to convert the raw ECG images into a suitable format for feature extraction and classification. The preprocessing stage includes several steps such as grayscale conversion, image resizing, contrast enhancement, and denoising. Algorithm 1 highlights
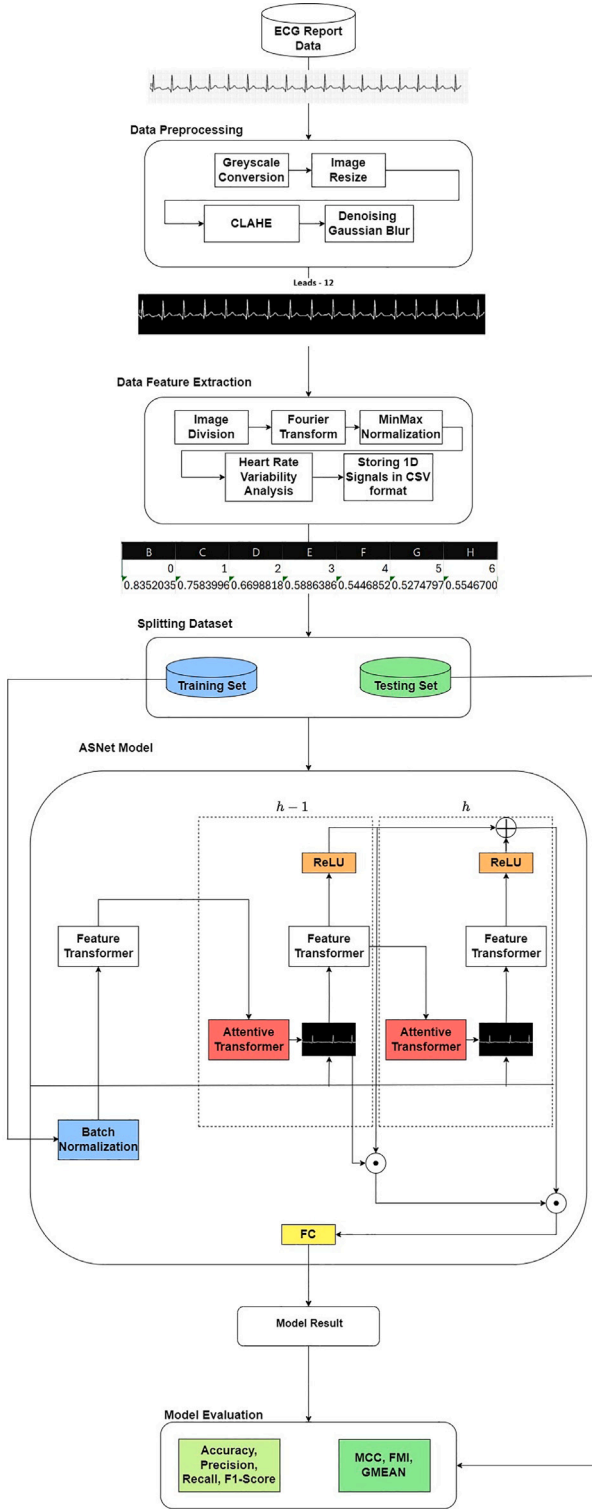
Fig. 2. Illustrates the overall architecture of the proposed ASNet model for cardiac rhythm classification.

Fig. 3. Step-wise Stacked Feature Extraction: Sequentially extracting time-domain, frequency-domain, statistical, morphological, and HRV features from normalized ECG segments, followed by concatenation into a comprehensive feature vector.

the training set or the testing set. Based on this partitioning strategy, no overlap between the training and testing sets is allowed. This approach prevents random image-level splitting, which could place different leads of the same patient into both sets. Such overlap would artificially inflate performance due to leakage of patient-specific waveform morphology or noise patterns. With patient-level separation, the reported metrics reflect the model's ability to generalize across unseen patients rather than memorizing intra-patient correlations. After applying this partitioning, the dataset consisted of 8,918 training images and 2,230 testing images. These patient-level splits were consistently maintained during preprocessing, feature extraction, and model training. This ensured reproducibility and fairness of evaluation. The raw image with a pixel size of 1162×827 was used as the input image (I). It was then converted into grayscale (I') using Eq. (1).

$$I' = G(I_i) \tag{1}$$

where i indicates the i_{th} input of the raw image. The grayscale converted image I' is then resized I'' to 256×256 using Eq. (2).

$$I'' = R_s(I', 256, 256) \tag{2}$$

where R_s indicates the resize function, then the contrast of the image is adjusted using Contrast Limited Adaptive Histogram Equalization (CLAHE). It also helps to improve the visibility of the ECG waveforms (I_c). It is computed using Eq. (3).

$$I_c = CLAHE(I'') \tag{3}$$

In order to reduce the noise (n_r), a Gaussian blur (G_b) is used. The final noise removal image is computed using Eq. (4).

$$I_{nr} = G_b(I_c) \tag{4}$$

3.2. Stacked transformation for feature extraction

The pre-processed 1D normalized sequences are then passed to stacked feature extraction. Stacked feature extraction for each ECG Segment has the following process. The ECG image is segmented into individual leads then it is passed to the Fourier transformation to extract frequency-domain features. Time-domain features (amplitudes and intervals) are extracted, followed by the calculation of statistical features (mean and variance). Morphological features (wave slopes and durations) and Heart Rate Variability (HRV) features (from RR intervals) are then derived. All extracted features are concatenated into a single vector, normalized, and stored as a 1D sequence in CSV format. This ensures efficient extraction of meaningful features from each ECG segment, supporting the performance of our proposed ASNet

the procedure carried out for pre-processing in the ASNet model. The dataset used in this study is structured as $N_{patients} \times 12$ leads. Each patient contributes twelve-lead ECG images in the standard diagnostic configuration. To ensure robust evaluation and eliminate the possibility of data leakage, the dataset is split into training (80%) and testing (20%) subsets at the patient level. All twelve-lead ECG images from a single patient were grouped together and assigned entirely to either
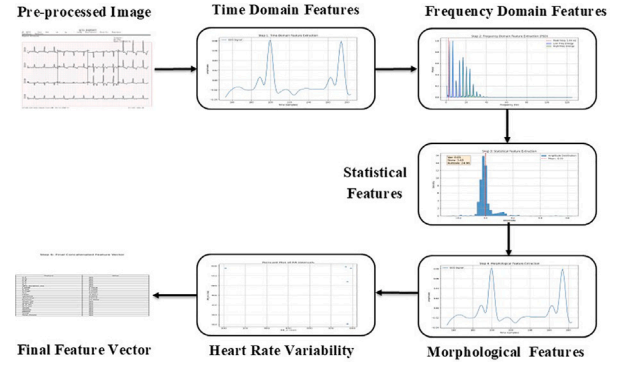

Algorithm 1 ECG Image Preprocessing

Require: Raw ECG dataset D with input images I_i of size 1162×827 pixels

Step 1: Image Preprocessing

for each image $I_i \in D$ **do**

    Convert to grayscale using equation (1)

    Resize to 256×256 using equation (2)

    Enhance contrast with CLAHE using equation(3)

    Apply Gaussian blur for denoising using equation (4)

end for

Step 2: Lead Extraction

for each I_{nr} **do**

    Divide into 12 bounding regions: $B = \{B_1, B_2, \ldots, B_{12}\}$

    Extract lead segments: $S_l = I_{nr}(x_l : x_l + w, y_l : y_l + h)$ for $l = 1, \ldots, 12$

end for

Step 3: Conversion to 1D Signals

for each lead segment S_l **do**

    Apply Discrete Fourier Transform (DFT) using equation (5)

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1 \quad (5)$$

    Normalize signals using equation (6) and store as 1D CSV files

$$F'(k) = \frac{F(k) - \min(F)}{\max(F) - \min(F)} \quad (6)$$

end for

Step 4: Sampling Rate Mapping

Temporal resolution:

$$T_s = \frac{25 \text{ mm/s}}{P_x} \quad (7)$$

Amplitude scaling:

$$A_s = \frac{10 \text{ mm/mV}}{P_y} \quad (8)$$

Step 5: Bounding and Axis Detection

Detect waveform region using equation (9)

$$B(x, y) = \begin{cases} 1, & \text{if } (x, y) \in \Omega_{\text{waveform}} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Remove grid lines, text, and annotations for clean waveform extraction

Ensure: Cleaned, preprocessed ECG signals stored as 1D normalized sequences

---

Table 1

Notation and description for Algorithm 2.

Notation	Description
$s_i(t)$	Normalized ECG segment for the ith heartbeat (1D signal)
$\mathbf{f}_i$	Final feature vector for the ith heartbeat
$\mathbf{f}_i^{(1)}$	Time-domain feature subvector for the ith heartbeat
$\mathbf{f}_i^{(2)}$	Frequency-domain feature subvector for the ith heartbeat
$\mathbf{f}_i^{(3)}$	Statistical feature subvector for the ith heartbeat
$\mathbf{f}_i^{(4)}$	Morphological feature subvector for the ith heartbeat
$\mathbf{f}_i^{(5)}$	Heart Rate Variability (HRV) feature subvector for the ith heartbeat
A_P, A_Q, A_R, A_S, A_T	Amplitudes of the P, Q, R, S, and T waves
PR_i	PR interval duration
QRS_i	QRS complex duration
QT_i	QT interval duration
RR_i	RR interval between consecutive R-peaks
$\mathcal{F}\{s_i(t)\}$	Fourier Transform of ECG segment
$\mathcal{W}\{s_i(t)\}$	Wavelet Transform of ECG segment
f_{peak}	Dominant frequency component
$E_{\text{low}}, E_{\text{high}}$	Energy in low and high frequency bands
H_s	Spectral entropy
μ_i	Mean value of the ECG segment
σ_i^2	Variance of the ECG segment
skew_i	Skewness of the ECG segment
kurt_i	Kurtosis of the ECG segment
slope_{QR}	Slope between Q and R peaks
slope_{RS}	Slope between R and S peaks
A_{QRS}	Area under the QRS complex
D_P, D_T	Duration of P wave and T wave
SDNN	Standard deviation of NN intervals (HRV feature)
RMSSD	Root mean square of successive RR differences (HRV feature)
pNN50	Percentage of RR intervals differing by > 50 ms (HRV feature)
LF/HF	Ratio of low- to high-frequency HRV power
Total Power	Total spectral power of HRV (overall variability)
d	Total number of extracted features

model. Algorithm 2 outlines the step-by-step process for extracting stacked transformation features, while the corresponding notations are provided in Table 1. Fig. 3 indicates the step-wise stacked feature extraction.

3.3. Feature transformer architecture

The Feature Transformer Architecture in the ASNet model employs an Encoder-Decoder setup for ECG signal processing. The encoder normalizes and embeds ECG features, applying multi-head self-attention to capture both local and global dependencies, followed by feed-forward networks for further refinement. The decoder aggregates and aligns features using cross-attention, followed by masked multi-head attention to focus on key signal components, and final feed-forward networks to refine the features for classification. This structure enables effective extraction, alignment, and classification of heart rhythm from ECG signals.

3.3.1. Batch normalization

The ECG image I is first batch-normalized with a fixed batch size $B = 64$ to stabilize training by normalizing the input distribution. The batch-normalized image is denoted as X_{bn}.

3.3.2. Feature transformer blocks for representation learning

The normalized input X_{bn} is then processed through a sequence of Feature Transformer Blocks (T_f), each following the encoder–decoder design described in Algorithm 3. Within each block, the **encoder** outputs Z_e by applying multi-head self-attention and feed-forward layers with residual connections, while the **decoder** refines these representations to produce Z_d using cross-feature attention, masked multi-head attention, and additional feed-forward layers. These successive transformations enable the extraction of high-level, discriminative ECG features. Fig. 4 depicts the Feature Transformer architecture for heart rhythm classification, and Table 2 summarizes the detailed network configuration.

Algorithm 2 Stacked Feature Extraction for Each ECG Segment

Require: Normalized ECG segment $s_i(t)$

 1: **Step 1: Extract Time-domain Features**
 2: $\mathbf{f}_i^{(1)} \leftarrow [A_P, A_Q, A_R, A_S, A_T, \mathrm{PR}_i, \mathrm{QRS}_i, \mathrm{QT}_i, \mathrm{RR}_i]$
 3: **Step 2: Extract Frequency-domain Features**
 4: Apply the Fourier Transform to the signal:
 5: $S_i(f) = \mathcal{F}\{s_i(t)\}$
 6: Extract the following frequency-domain features:
 7: $\mathbf{f}_i^{(2)} \leftarrow [f_{\mathrm{peak}}, E_{\mathrm{low}}, E_{\mathrm{high}}, H_s]$
 8: **Step 3: Extract Statistical Features**
 9: $\mathbf{f}_i^{(3)} \leftarrow [\mu_i, \sigma_i^2, \mathrm{skew}_i, \mathrm{kurt}_i]$
10: **Step 4: Extract Morphological Features**
11: $\mathbf{f}_i^{(4)} \leftarrow [\mathrm{slope}_{QR}, \mathrm{slope}_{RS}, A_{QRS}, D_P, D_T]$
12: **Step 5: Extract HRV Features**
13: $\mathbf{f}_i^{(5)} \leftarrow [\mathrm{SDNN}, \mathrm{RMSSD}, \mathrm{pNN50}, LF/HF, \mathrm{Total\ Power}]$
14: **Step 6: Concatenate All Features**
15: $\mathbf{f}_i \leftarrow [\mathbf{f}_i^{(1)}, \mathbf{f}_i^{(2)}, \mathbf{f}_i^{(3)}, \mathbf{f}_i^{(4)}, \mathbf{f}_i^{(5)}]$

Ensure: Feature vector $\mathbf{f}_i$ for the ECG segment $s_i(t)$ after stacked transformations

Table 2
Configuration of Feature Transformer.

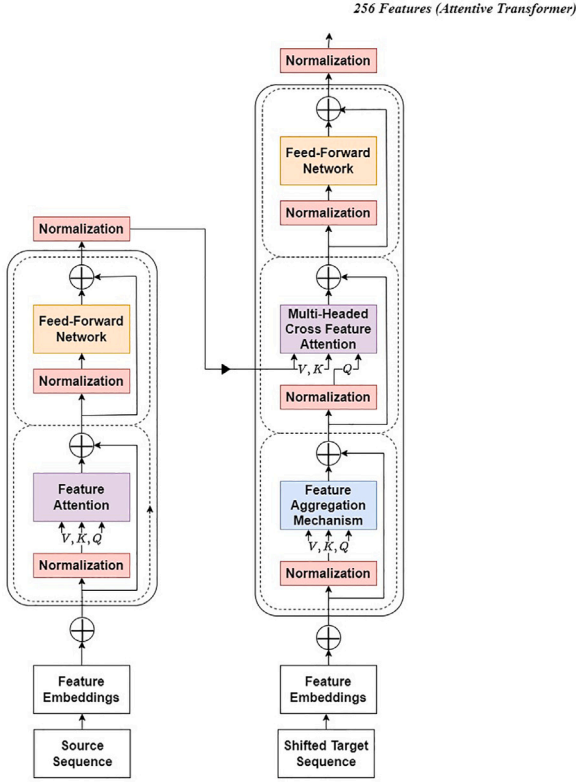| Feature Transformer			
Layer name	Operation	No. of filters	Parameters
Input Layer	Dense Layer	256	128 features
Feature Attention Layer	Multi-Head Attention	512	H=8, d_k=32
Feed-Forward Encoder Layer 1	Dense Layer + ReLU	1024	–
Feed-Forward Encoder Layer 2	Dense Layer + ReLU	1024	–
Normalization Layer	Layer Normalization	256	–
Feature Aggregation Layer	Multi-Head Attention	256	H=8, d_k=32
Cross Feature Attention Layer	Cross-Attention	256	H=8, d_k=32
Feature Transformer Output Layer	Dense Layer	256	–

Fig. 4. Architecture of the Feature Transformer, incorporating feature attention for the source sequence, a feature aggregation mechanism for the target sequence, and multi-headed cross-feature attention to enhance feature interactions. Each layer is followed by normalization and a feed-forward network to ensure robust feature representation.

The algorithm extracts comprehensive features from each normalized ECG heartbeat segment by applying stacked transformations. It first derives time-domain features such as peak amplitudes and intervals, then captures frequency-domain characteristics using Fourier or Wavelet transforms. Next, it computes statistical descriptors including mean and variance, and finally extracts morphological features such as slopes and wave durations. All these features are concatenated into a single feature vector for each heartbeat, enabling effective ECG signal analysis and classification.

Algorithm 3 Feature Transformer in the ASNet Model

Require: Preprocessed ECG image
Ensure: Refined feature representations for classification
    Encoder:
 1: **for** each input ECG image **do**
 2:     Apply batch normalization: $X_{bn} \leftarrow \mathrm{BN}(\mathrm{Image})$
 3:     Embed features: $E_f \leftarrow \mathrm{Embed}(X_{bn})$
 4:     Normalize: $N_{e1} \leftarrow \mathrm{Norm}(E_f)$
 5:     Compute feature self-attention: $A_f \leftarrow \mathrm{SelfAttention}(N_{e1})$
 6:     Residual connection: $R_{e1} \leftarrow A_f + N_{e1}$
 7:     Normalize: $N_{e2} \leftarrow \mathrm{Norm}(R_{e1})$
 8:     Feed-forward network: $F_e \leftarrow \mathrm{FFN}(N_{e2})$
 9:     Residual connection: $R_{e2} \leftarrow F_e + R_{e1}$
10:     Final encoder output: $Z_e \leftarrow \mathrm{Norm}(R_{e2})$
11: **end for**
    Decoder:
12: **for** each encoder output Z_e **do**
13:     Apply batch normalization: $Z_{bn} \leftarrow \mathrm{BN}(Z_e)$
14:     Embed features: $E_d \leftarrow \mathrm{Embed}(Z_{bn})$
15:     Normalize: $N_{d1} \leftarrow \mathrm{Norm}(E_d)$
16:     Aggregate features: $A_d \leftarrow \mathrm{FeatureAggregation}(N_{d1})$
17:     Residual connection: $R_{d1} \leftarrow A_d + N_{d1}$
18:     Normalize: $N_{d2} \leftarrow \mathrm{Norm}(R_{d1})$
19:     Cross-feature attention with encoder output:
        $C_a \leftarrow \mathrm{CrossFeatureAttention}(N_{d2}, Z_e)$
20:     Residual connection: $R_{d2} \leftarrow C_a + N_{d2}$
21:     Normalize: $N_{d3} \leftarrow \mathrm{Norm}(R_{d2})$
22:     Feed-forward network: $F_d \leftarrow \mathrm{FFN}(N_{d3})$
23:     Residual connection: $R_{d3} \leftarrow F_d + R_{d2}$
24:     Final decoder output: $Z_d \leftarrow \mathrm{Norm}(R_{d3})$
25: **end for**
    Output Layer:
26: Apply ReLU activation: $O \leftarrow \mathrm{ReLU}(Z_d)$
27: Compute classification loss: $L_{\mathrm{ASNet}} \leftarrow -\sum y \log \hat{y}$
28: Update ASNet parameters with Adam optimizer:

$$\theta_{\mathrm{ASNet}} \leftarrow \theta_{\mathrm{ASNet}} - \eta \nabla L_{\mathrm{ASNet}}$$

3.3.3. Feature attention mechanism — Encoder

In the encoder, the batch-normalized input X_{bn} is embedded to obtain dense feature representations E_f. These embeddings form the basis for learning patterns in the ECG waveforms. The embedded features are normalized to produce N_{e1}, which is then processed by

the Feature Attention Mechanism to compute multi-head self-attention scores, using query (Q), key (K), and value (V) matrices. The resulting attention output is denoted as A_f. A residual connection is applied to combine the attention output with the normalized input: $R_{e1} = A_f + N_{e1}$. This is followed by a second normalization step to produce N_{e2}, which is then passed to the feed-forward network as described in Algorithm 3.

3.3.4. Feed-forward networks with skip connections

After the attention mechanism, a Feed-Forward Network (FFN) further refines the features by capturing nonlinear patterns in the ECG data. The FFN takes as input the normalized encoder output N_{e2} and produces F_e. A residual (skip) connection is then applied to preserve the original feature information, and the result is finally normalized to obtain the encoded source features, as expressed in Eq. (10):

$$Z_e = \text{Norm}(F_e + R_{e1}). \tag{10}$$

This formulation helps to maintain feature integrity while enhancing the learned representation.

3.3.5. Feature aggregation with decoder

On the decoder side, the architecture refines the encoder's output representations for accurate classification. Initially, the encoder output Z_e is batch-normalized to obtain Z_{bn}, then embedded to produce E_d, which is subsequently normalized to N_{d1}. The Feature Aggregation Mechanism computes aggregated features and applies multi-head cross-feature attention between the decoder's normalized features and the encoder output, as denoted in Eq. (11).

$$C_a = \text{CrossFeatureAttention}(\text{Norm}(A_d + N_{d1}), Z_e) \tag{11}$$

where A_d is computed using Eq. (12).

$$A_d = \text{FeatureAggregation}(N_{d1}). \tag{12}$$

Eq. (11) therefore represents the key decoder step that integrates the encoder output into the aggregated decoder features.

The decoder output is computed through a series of residual and normalization steps. First, a residual connection combines the cross-feature attention output C_a with the normalized decoder representation N_{d2}:

$$R_{d2} = C_a + N_{d2}. \tag{13}$$

The result is normalized to obtain

$$N_{d3} = \text{Norm}(R_{d2}), \tag{14}$$

and then passed through the Feed-Forward Network (FFN) to produce

$$F_d = \text{FFN}(N_{d3}). \tag{15}$$

A second residual connection gives the final decoder output:

$$Z_d = \text{Norm}(F_d + R_{d2}). \tag{16}$$

The multi-class classification loss of the proposed ASNet model is defined by the cross-entropy function (Eq. (17)):

$$L_{\text{ASNet}} = -\sum y \log(\hat{y}). \tag{17}$$

Finally, the model parameters are updated using the Adam optimizer as expressed in Eq. (18):

$$\theta_{\text{ASNet}} \leftarrow \theta_{\text{ASNet}} - \eta \cdot \nabla L_{\text{ASNet}}. \tag{18}$$

Eqs. (13)–(16) describe the decoder's residual and normalization operations, while Eqs. (17) and (18) define the training objective and the parameter update step, respectively.

3.3.6. Variance scaling for numerical stability

To preserve the integrity of the smoothed features, the outputs of each transformation stage are concatenated. Let X_1 and X_2 denote the feature representations obtained after the sequence of operations FC → BN → GLU in two consecutive transformations. The variance of each transformed feature set is then computed as

$$\text{Var}(X_1) = \frac{1}{n} \sum_{i=1}^{n} (X_{1,i} - \mu_{X_1})^2, \tag{19}$$

$$\text{Var}(X_2) = \frac{1}{n} \sum_{i=1}^{n} (X_{2,i} - \mu_{X_2})^2, \tag{20}$$

where μ_{X_1} and μ_{X_2} are the mean values of X_1 and X_2, respectively, and n is the number of feature elements.

Eqs. (19) and (20) define the variance scaling step, which ensures numerical stability and prevents feature dominance during subsequent encoder–decoder processing.

3.4. Attentive transformer architecture

The Attentive Transformer in ASNet consists of three main components: the Feature Attention Mechanism on the encoder side, which uses multi-head self-attention to learn intra-sequence dependencies through query, key, and value matrices; the Cross-Feature Attention on the decoder side, which models inter-sequence dependencies between the encoder output and decoder input using multi-head attention; and the Masked Multi-Head Self-Attention on the decoder side, which prevents the model from accessing future time steps during training by applying a mask. Algorithm 4 details the stepwise procedure of the Attentive Transformer.

3.4.1. Encoder for sequential dependency modeling

The encoder processes the source sequence of ECG features, while the decoder generates refined features contributing to accurate cardiac rhythm classification. Fig. 5 illustrates the architecture of the Attentive Transformer, and Table 3 shows the network configuration, where the embedding dimension $d_{\text{model}} = 256$, the number of attention heads $H = 8$, and the key/value dimensions per head are d_k/d_v. Initially, positional encoding P_e is applied to the input sequence, which is then normalized to N_{e1} and passed through the standard multi-head self-attention module A_e. A residual connection forms $R_{e1} = A_e + N_{e1}$, followed by normalization $N_{e2} = \text{Norm}(R_{e1})$. The result is fed to the Feed-Forward Network F_e, and a residual connection produces the final encoder output $Z_e = \text{Norm}(F_e + R_{e1})$. This formulation allows the model to capture both local and global temporal dependencies in the ECG signal, generating encoded features Z_e that are subsequently used in the decoder for cross-feature attention and masked self-attention operations.

3.4.2. Decoder with masked multi-head attention

The decoder of the Attentive Transformer refines the encoder features for accurate classification. Eight multi-head attention mechanisms are introduced in the decoder to extract features required for classification. All batches undergo positional encoding P_d when passed to the input sequence. The encoded sequence is normalized to N_{d1} and passed through a series of residual blocks to produce R_{d1}. This normalized value N_{d2} is then input to the multi-head cross-feature attention mechanism to generate C_a. The extracted features are passed through the final feed-forward network F_c and concatenated with R_{d2} to produce Z_d. Finally, the refined features Z_d are passed through a fully connected (FC) layer with ReLU activation to produce the output O_h for cardiac rhythm classification.

4. Experiments

This section explains the dataset used for both training and validation. Also, indicates the various evaluation metrics used to validate the proposed ASNet model along with its implementation details.
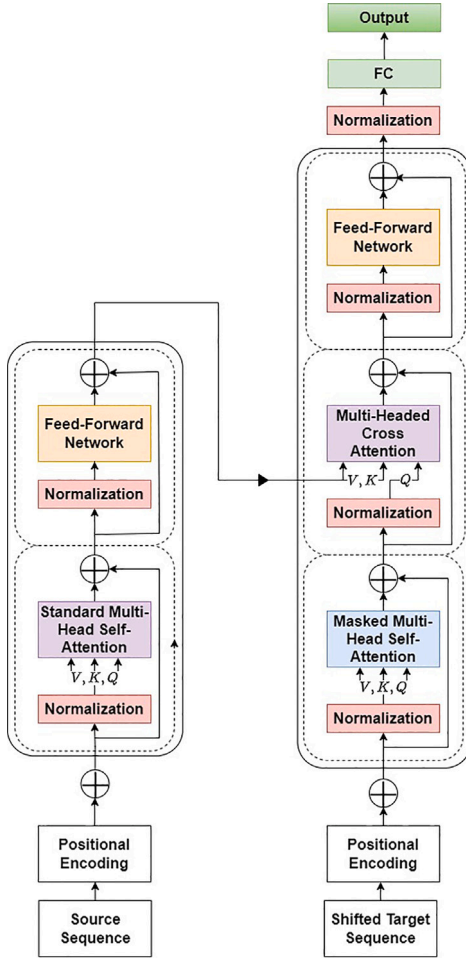
Fig. 5. Architecture of the Attentive Transformer, comprising standard multi-head self-attention for the source sequence, masked multi-head self-attention for the target sequence, and multi-headed cross-attention to align features. Each layer is followed by normalization and feed-forward networks, ensuring effective sequence representation for ECG classification.

Table 3
Configuration of Attentive transformer.

Feature Transformer			
Layer name	Operation	No. of filters	Parameters
Input Layer	Positional Encoding	256	256 features from FT
Self-Attention Layer	Multi-Head Attention	512	H=8, d_k=32
Feed-Forward Encoder Layer 1	Dense Layer + ReLU	1024	–
Feed-Forward Encoder Layer 2	Dense Layer + ReLU	1024	–
Masked Self-Attention Layer	Multi-Head Attention	512	H=8, d_k=32
Cross Feature Attention Layer	Cross-Attention	256	H=8, d_k=32
Output Classification Layer	Dense Layer + Softmax	4	–

4.1. ECG images dataset of cardiac patients

Our proposed ASNet model uses the ECG Image dataset of Cardiac Patients, which is publicly available in [42]. The data set consists

---

Algorithm 4 Attentive Transformer in ASNet Model

Require: Batch of preprocessed ECG images B
Ensure: Refined attentive features for classification

  Encoder:
1: **for** each input batch B **do**
2:   Apply positional encoding:
      $P_e \leftarrow \text{PositionalEncoding}(B)$
3:   Normalize encoded input: $N_{e1} \leftarrow \text{Norm}(P_e)$
4:   Multi-head self-attention:
      $A_e \leftarrow \text{MultiHeadSelfAttention}(N_{e1})$
5:   Residual connection:
      $R_{e1} \leftarrow A_e + N_{e1}$
6:   Normalize:
      $N_{e2} \leftarrow \text{Norm}(R_{e1})$
7:   Feed-Forward Network (FFN):
      $F_e \leftarrow \text{FFN}(N_{e2})$
8:   Residual connection and final encoder output:
      $Z_e \leftarrow \text{Norm}(F_e + R_{e1})$
9: **end for**
  Decoder:
10: **for** each input batch B **do**
11:   Apply positional encoding:
       $P_d \leftarrow \text{PositionalEncoding}(B)$
12:   Normalize: $N_{d1} \leftarrow \text{Norm}(P_d)$
13:   Masked multi-head self-attention:
       $M_d \leftarrow \text{MaskedMultiHeadSelfAttention}(N_{d1})$
14:   Residual connection: $R_{d1} \leftarrow M_d + N_{d1}$
15:   Normalize: $N_{d2} \leftarrow \text{Norm}(R_{d1})$
16:   Multi-head cross-feature attention with encoder output:
       $C_a \leftarrow \text{CrossFeatureAttention}(N_{d2}, Z_e)$
17:   Residual connection: $R_{d2} \leftarrow C_a + N_{d2}$
18:   Normalize: $N_{d3} \leftarrow \text{Norm}(R_{d2})$
19:   Feed-Forward Network (FFN): $F_c \leftarrow \text{FFN}(N_{d3})$
20:   Residual connection and final decoder output:
       $Z_d \leftarrow \text{Norm}(F_c + R_{d2})$
21: **end for**
  Output Layer:
22: Apply ReLU activation: $O_h \leftarrow \text{ReLU}(Z_d)$
23: Compute ASNet loss: $L_{\text{ASNet}} \leftarrow -\sum y \log(\hat{y})$
24: Update parameters with Adam optimizer:

   $\theta_{\text{ASNet}} \leftarrow \theta_{\text{ASNet}} - \eta \nabla L_{\text{ASNet}}$

---

of four different categories of ECG images. 1. Patients with myocardial infarction ($240 \times 12 = 2880$ images), 2. Abnormal Heartbeat ($233 \ times 12 = 2796$ images), 3. history of Myocardial Infarction (MI) ($172 \times$The $12 = 2064$ images), and 4. Normal person ($284 \times 12 = 3408$ images). The proposed model is trained and validated with the split ratio of 80% (8918 images) and 20% (2230 images), respectively.

4.2. Other benchmark datasets

The ASNet model is also validated using two different benchmark datasets, as mentioned in the following.

- **Cornell University ECG Image Database:** ECG Image Kit is introduced in [43], it consists of 35,000 ECG images derived from the PTB-XL and Emory Healthcare databases with artifacts and distortions. They also provided a ground-truth time series of ECG images.
- **PTB Diagnostic ECG Database:** PhysioBank is introduced in [44], namely PTB Diagnostic ECG Database. It consists of 549 high-resolution images of ECG recordings. These records were taken

from 294 individual patients with nine different classes. In order to validate the proposed ASNet model, we used only four classes (Myocardial infarction, Abnormal Heartbeat, history of Myocardial Infarction, and Normal) out of nine.

4.3. Evaluation metrics

To comprehensively assess the performance of our ASNet-based ECG classification framework, we employ multiple evaluation metrics to ensure an accurate and balanced evaluation of the effectiveness of the model. Since medical diagnosis is highly sensitive, it is crucial not only to measure overall classification performance but also to focus on detecting abnormalities with high precision while minimizing false positives and false negatives.

- TP_{AS} **(True Positive)**: It correctly identifies a normal heartbeat as normal.
- TN_{AS} **(True Negative)**: It correctly identifies a abnormal heartbeat as abnormal.
- FP_{AS} **(False Positive)**: It incorrectly identifies a normal heartbeat as abnormal.
- FN_{AS} **(False Negative)**: It incorrectly identifies a abnormal heartbeat as normal.

The most fundamental metric is accuracy, which evaluates the proportion of correctly classified instances relative to the total number of instances. Here we calculate the metric using (21)

$$A_{AS_{net}} = \frac{TP_{AS} + TN_{AS}}{TP_{AS} + TN_{AS} + FP_{AS} + FN_{AS}} \tag{21}$$

We calculate Precision using Eq. (22), also known as Positive Predictive Value (PPV), which measures how many of the predicted positive cases are actually correct.

$$P_{AS_{net}} = \frac{TP_{AS}}{TP_{AS} + FP_{AS}} \tag{22}$$

Precision is particularly important in ECG classification, as high precision ensures that false alarms (false positives) are minimized, reducing the risk of unnecessary medical intervention. However, precision alone is not sufficient, as it does not account for missed positive cases. The recall is computed using Eq. (23). It measures the ability of the model to correctly detect actual positive cases.

$$R_{AS_{net}} = \frac{TP_{AS}}{TP_{AS} + FN_{AS}} \tag{23}$$

A high recall is critical in medical diagnosis since missing an actual case (false negative) could lead to severe health consequences for the patient. Since there is often a trade-off between precision and recall, we use the F1-score. F1-score is computed using (24).

$$F1_{AS_{net}} = 2 \times \frac{P_{AS_{net}} \times R_{AS_{net}}}{P_{AS_{net}} + R_{AS_{net}}} \tag{24}$$

A high F1-score indicates a good balance between precision and recall, ensuring that the model neither misses too many actual cases nor generates excessive false alarms.

$$
\begin{aligned}
P_1 &= TP_{AS} + FP_{AS}, \quad P_2 = TP_{AS} + FN_{AS}, \\
P_3 &= TN_{AS} + FP_{AS}, \quad P_4 = TN_{AS} + FN_{AS}, \\
MCC_{AS_{net}} &= \frac{(TP_{AS} \times TN_{AS}) - (FP_{AS} \times FN_{AS})}{\sqrt{P_1 \times P_2 \times P_3 \times P_4}}
\end{aligned} \tag{25}
$$

To further enhance the reliability of evaluation, we incorporate the Matthews Correlation Coefficient ($MCC_{AS_{net}}$) using Eq. (25), which provides a more balanced performance measure especially for imbalanced datasets.

$MCC_{AS_{net}}$ is highly valuable in medical classification tasks as it takes into account all four values of the confusion matrix and is particularly effective in scenarios where class distributions are skewed.

Table 4
Computational complexity.

Component	Specification/Value
GPU	NVIDIA RTX 3090
CPU	Intel Core i9-12900K
RAM	32 GB DDR5
Training Time	Approximately 4 h (50 epochs, batch size = 64)
Testing Time	8 min (on 20% test split)
Time Taken for Single Image Inference	0.219 s
Class Prediction Time	0.144 s
Tools/Software Used	TensorFlow & Keras

In addition, the Fowlkes-Mallows Index (FMI) is used to evaluate the geometric mean of precision and recall. Using the Eqs. (22) and (23), we calculate its geometric mean using (26).

$$FMI_{AS_{net}} = \sqrt{P_{AS_{net}} \times R_{AS_{net}}} \tag{26}$$

A high FMI value indicates that the model maintains a good balance between positive predictive power and recall, making it a reliable metric for classification performance evaluation.

Finally, G-Mean (Geometric Mean) ensures that the model performs well on both classes. It is computed using Eq. (27).

$$G_{AS_{net}} = \sqrt{Sensitivity \times Specificity} \tag{27}$$

G-Mean is particularly useful in handling class imbalances, as it ensures that both positive and negative class predictions are considered equally in the evaluation. By incorporating these diverse evaluation metrics, we provide a comprehensive assessment of our ASNet-based ECG classification framework, ensuring that the model effectively detects cardiac abnormalities while minimizing the risks of misclassifications. These metrics collectively enhance the reliability and interpretability of our classification results, making them suitable for real-world medical applications.

4.4. Implementation details

Proposed AS_{net} is implemented using a high-performance computing system equipped with an NVIDIA GPU (RTX 3090) to accelerate deep learning model training through CUDA-optimized TensorFlow. The system has a specification of Intel Core i9 with 32 GB RAM. The ASNet model is trained using the Adam optimizer with a learning rate of 0.001. Training is performed over 50 epochs with a batch size of 64. To prevent overfitting and improve generalization, a dropout rate of 0.5 is applied to the fully connected layers. These hyperparameters were carefully selected to ensure stable convergence and optimal performance of the network. The complete training process took approximately 4 h. Testing was done in 8 min using a 20% split from the ECG Images dataset of the Cardiac Patients. The computational time complexity of the proposed ASNet model is indicated in Table 4. ASNet takes 0.219 s to infer a single ECG report, while the class prediction time takes 0.144 s. The open source TensorFlow and Keras frameworks were utilized in building the proposed model.

Table 5 represents the comparative analysis of computational complexity with various existing models. The proposed ASNet model demonstrates favorable computational efficiency with reliable classification, achieving 91.31% accuracy and an average inference latency of 0.219 s per signal. This is significantly faster than FFT+CNN-LSTM's 213 s and AF Detection's 2.052 s, showing a 9× to 1000× improvement. LSTM-based models have high time complexity, approximately $\mathcal{O}(n \times m \times h)$, where n is the input length, m is the feature dimension, and h is hidden units, making them less efficient for long sequences. ASNet

Table 5

Comparative analysis of computational complexity and performance metrics of ECG-based cardiac rhythm classification models.

Model	Accuracy	Platform	Inference time & Tools used	Key insight
AF Detection Model [45]	84.5%	CPU	CPU: 2.052 s per signal	Highlights inefficiency without GPU acceleration
ECG-TCN [46]	94.2%	STM32L475/GAP8 MCU	0.10 mJ; 23×more efficient than ARM M4F	Energy-optimized for embedded systems
Binarized CNN [47]	95.67%	Not specified	12.65×speedup, 24.8×compression, 25% memory savings	Extreme speed and efficiency gains
Tiny Transformer [48]	98.97%	GAP9 MCU	4.28 ms, 0.09 mJ	Ultra-efficient inference for wearables
Quantized CNN [49]	99.6%	ARM Cortex A55	7.65 ms, 5.85 mJ, 93 kB memory	Low power, compact, high-performing model
FFT+CNN-LSTM [50]	97.4%	NVIDIA RTX 2060, Intel Core i7-10750H, 16 GB DDR4 RAM	Testing time: 213s; TensorFlow & Keras	Integrates temporal-frequency learning for preventing overfitting
ASNet (proposed)	**91.31%**	NVIDIA RTX 3090, Intel Core i9-12900K, 32 GB DDR5	Testing: 8 mins (20% test data); Single inference: 0.219s; TensorFlow & Keras	Balanced deep learning model with good accuracy and speed

Table 6

Quantitative analysis of ASNet using ECG Images dataset of Cardiac Patients ($A_{AS_{net}}$, $P_{AS_{net}}$, $R_{AS_{net}}$, $F1_{AS_{net}}$, $MCC_{AS_{net}}$, $FMI_{AS_{net}}$ and $G_{AS_{net}}$.

Model	$A_{AS_{net}}$	$P_{AS_{net}}$	$R_{AS_{net}}$	$F1_{AS_{net}}$	$MCC_{AS_{net}}$	$FMI_{AS_{net}}$	$G_{AS_{net}}$
Base Model -TabNet (BM)	89.53	87.31	90.16	88.73	0.7208	0.9132	0.8847
BM+ FT	90.13	88.04	90.81	89.33	0.7432	0.9224	0.8848
BM + AT	90.64	88.52	91.33	89.86	0.7524	0.9248	0.8866
BM+FT+AT	90.93	89.12	91.81	90.43	0.7611	0.9351	0.8881
ASNet (proposed)	91.31	89.72	92.54	91.14	0.7702	0.9419	0.8902

Table 7

Quantitative analysis of ASNet using Cornell University ECG Image Database ($A_{AS_{net}}$, $P_{AS_{net}}$, $R_{AS_{net}}$, $F1_{AS_{net}}$, $MCC_{AS_{net}}$, $FMI_{AS_{net}}$ and $G_{AS_{net}}$.

Model	$A_{AS_{net}}$	$P_{AS_{net}}$	$R_{AS_{net}}$	$F1_{AS_{net}}$	$MCC_{AS_{net}}$	$FMI_{AS_{net}}$	$G_{AS_{net}}$
Base Model -TabNet (BM)	88.21	86.53	87.93	87.23	0.7014	0.9030	0.8714
BM+ FT	88.91	87.33	88.45	87.80	0.7287	0.9102	0.8802
BM + AT	89.632	87.910	89.103	88.530	0.7359	0.9147	0.8849
BM+FT+AT	90.301	88.720	89.803	89.210	0.7532	0.9211	0.8966
ASNet (proposed)	92.71	91.28	92.10	91.60	0.7871	0.9325	0.8713

avoids recurrence and uses parallel convolutional operations with linear complexity $\mathcal{O}(k \times n)$, where k is the kernel size, improving speed and scalability. Its stacked convolutional layers extract localized features with reduced temporal dependencies, lowering memory bandwidth during training and inference. Running on NVIDIA RTX 3090 and Intel Core i9-12900K, it benefits from GPU acceleration and high memory throughput. Compared to MCU-based models like Tiny Transformer and ECG-TCN, optimized for ultra-low latency and energy, ASNet balances scalability and deeper classification capability. This makes it well-suited for clinical decision support systems requiring real-time performance and moderate computational resources.

5. Results

In this section, an extensive analysis of results is discussed for the AS_{net} model for classifying heart disease. The proposed AS_{net} model is evaluated using various metrics for measuring the accuracy of classified ECG signals. Performance of the AS_{net} model is compared with Base Model (BM) + Feature Transformer (FT) + Attentive Transformer (AT), BM+AT, and BM+FT models were evaluated for various benchmark datasets.

5.1. Extensive analysis of results obtained for ASNet model using ECG images dataset of cardiac patients
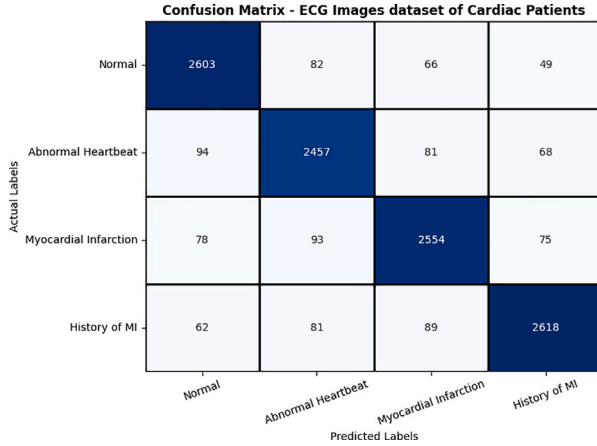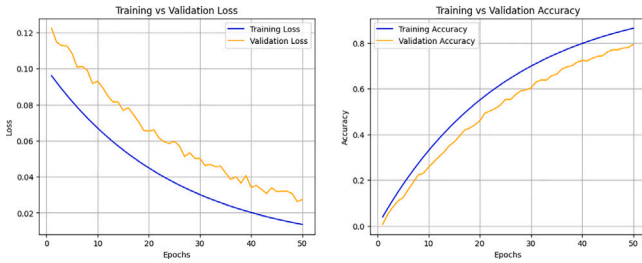
Table 6 highlights the results obtained for the AS_{net} model using various evaluation metrics for accurate heart disease classification through ECG signals. It also highlights the accurate results obtained for classification.

Based on Table 6 following observations are made. Initially, we designed a base model (BM) using a tabular network (TabNet) for classifying heart disease. It achieves value of $A_{AS_{net}}$ 89.53% and $F1_{AS_{net}}$ 88.73%. Then, we integrate the Feature Transformer (FT) network with the base model in order to achieve a better score. By incorporating the Feature Transformer (FT) network with the base model, it slightly increases the $A_{AS_{net}}$, which proves that the FT network helps to extract only the required feature for accurate classification of the heart disease. In order to achieve a higher accuracy score, we concatenate the features extracted from the FT with the Attentive Transformer (AT) network. The results obtained for BM+FT+AT achieve a better accuracy score $F1_{AS_{net}}$ of 90.43% when compared with BM+FT and BT+AT. Then we implemented stacked BM+FT+AT, namely AS_{net}, to obtain a higher score.

The results obtained for ECG Images dataset of Cardiac Patient shows that the AS_{net} model (BM+FT+ AT+Stacked) performs well when compared with BM+ FT+AT, BM+AT and BM+FT models in terms of $A_{AS_{net}}$ and $F1_{AS_{net}}$ scores, it achieves the value of 91.31% and 91.14% respectively. Fig. 6 indicates the confusion matrix obtained for the ECG Images of Cardiac Patients dataset. It was clearly observed that the true positive values are high when compared with other values. Fig. 7 shows the training and validation curves for both accuracy and loss. Fig. 8 showcases the few sample output images obtained for the ECG Images dataset of the Cardiac Patients dataset, along with the classified labels. The proposed ASNet model demonstrates strong alignment with expert annotations, achieving an IoU score of 0.8790 on the ECG Images dataset of cardiac patients. This high overlap metric indicates that the model effectively identifies clinically relevant regions, closely matching

Table 8

Quantitative analysis of ASNet using PTB diagnostic ECG database ($A_{AS_{net}}$, $P_{AS_{net}}$, $R_{AS_{net}}$, $F1_{AS_{net}}$, $MCC_{AS_{net}}$, $FMI_{AS_{net}}$ and $G_{AS_{net}}$.

Model	$A_{AS_{net}}$	$P_{AS_{net}}$	$R_{AS_{net}}$	$F1_{AS_{net}}$	$MCC_{AS_{net}}$	$FMI_{AS_{net}}$	$G_{AS_{net}}$
Base Model -TabNet (BM)	75.00	75.50	74.53	75.00	0.7027	0.7211	0.7426
BM+ FT	76.213	76.801	75.517	76.112	0.7209	0.7427	0.7533
BM + AT	78.011	78.559	77.213	77.801	0.7433	0.7614	0.7707
BM+FT+AT	79.533	79.801	78.567	79.112	0.7626	0.7884	0.7902
ASNet (proposed)	83.00	83.52	82.51	83.01	0.8038	0.8231	0.8280

Fig. 6. Confusion Matrix - ECG Images dataset of Cardiac Patients.

Fig. 7. Training and Validation - Accuracy and Loss. Demonstrates the significant reduction in loss while an increase in accuracy throughout the training and validation process of 50 epochs.

expert-defined areas and validating its reliability for cardiac feature localization.

5.2. Comprehensive analysis of heart disease classification using Cornell university ECG image database

The results in Table 7 demonstrate that the Cornell University dataset leads to a significant 1.4% increase in accuracy for the proposed AS_{net} model compared to its original training dataset. The proposed ASNet significantly outperforms the Base Model (TabNet) across all evaluated metrics, including accuracy (92.71% vs. 88.21%), precision (91.28% vs. 86.53%), recall (92.10% vs. 87.93%), F1-score (91.60% vs. 87.23%), MCC (0.7871 vs. 0.7014), FMI (0.9325 vs. 0.9030), and G-mean (0.8713 vs. 0.8714). Although adding the Feature Transformer (BM + FT) improves performance over the base model, it still underperforms compared to the complete ASNet, which achieves superior results across all metrics. The Attentive Transformer alone (BM + AT) achieves better results than BM + FT but remains inferior to ASNet, highlighting the benefit of integrating both components. Combining both Feature and Attentive Transformers (BM + FT + AT) further improves performance, approaching that of ASNet, though the proposed ASNet with stackable mechanisms still achieves the highest precision, recall, and F1-score.

The F1-score of the proposed AS_{net} shows a slight improvement, reaching 91.60%, compared to 91.14% from the original dataset. While there is a slight dip in performance for BM, BM + FT, BM + AT, and BM + FT + AT models on the Cornell dataset, the final AS_{net} demonstrates consistent improvement in accuracy, F1-score, and MCC. Fig. 9 presents the confusion matrix for the Cornell University ECG Image Database, indicating enhanced identification of ECG records related to previously diagnosed myocardial infarction (History of MI) compared to the training dataset. Fig. 10 showcases sample outputs from the Cornell University ECG Image Database, along with their corresponding class labels. The proposed ASNet model achieved an IoU score of 0.8242 on the Cornell University ECG Image Database, demonstrating strong agreement with expert annotations and confirming its effectiveness in accurately localizing diagnostically relevant ECG regions.

5.3. Comprehensive analysis of heart disease classification using PTB diagnostic ECG database

Table 8 presents the performance analysis of the proposed ASNet model using the PTB diagnostic ECG database. The following observations are made from Table 8. The proposed ASNet significantly outperforms the Base Model (TabNet) across all metrics, with an increase of 8.00% in accuracy (83.00% vs. 75.00%), 8.02% in precision (83.52% vs. 75.50%), 7.98% in recall (82.51% vs. 74.53%), 8.01% in F1-score (83.01% vs. 75.00%), an MCC increase of 0.1011 (0.8038 vs. 0.7027), 10.2% in FMI (0.8231 vs. 0.7211), and 8.54% in G-mean (0.8280 vs. 0.7426). While the addition of the Feature Transformer (BM + FT) yields improvements over the base model with +1.21% accuracy (76.21% vs. 75.00%), +1.3% precision, +0.98% recall, and +1.11% F1-score, the proposed ASNet surpasses BM + FT by 6.79%, 6.72%, 6.99%, and 6.9%, respectively, in these metrics, confirming the benefit of the full architecture. The BM + Attentive Transformer (BM + AT) further improves metrics compared to BM + FT, achieving 78.01% accuracy (+1.8%), 78.56% precision (+1.76%), 77.21% recall (+1.69%), and 77.80% F1-score (+1.69%). However, ASNet still improves these by 4.99%, 4.96%, 5.3%, and 5.21%, respectively, highlighting the advantage of its combined mechanisms.

The combination of both transformers (BM + FT + AT) achieves 79.53% accuracy (+1.52% over BM + AT), 79.8% precision (+1.24%), 78.57% recall (+1.36%), and 79.11% F1-score (+1.31%). Despite these gains, ASNet still outperforms BM + FT + AT by 3.47%, 3.72%, 3.53%, and 3.9%, respectively, showcasing the effectiveness of the stackable mechanisms in the ASNet model. The confusion matrix provided in Fig. 11 shows that the true positive and true negative classification capabilities are on par with the AS_{net} models trained on different datasets as displayed in Figs. 6 and 9. Fig. 13 highlights a few sample images of the PTB Diagnostic ECG dataset. Accuracy drop of approximately 8%: ASNet achieves around 91% accuracy on the ECG Images dataset of Cardiac Patients, but only about 83% on the PTB dataset. Domain shift: Variations in data sources, patient demographics, and ECG image characteristics cause reduced generalization of the model to the PTB dataset. Potential overfitting: The model may have learned dataset-specific features from the ECG Images dataset of Cardiac Patients, limiting its transferability to PTB. Class imbalance and subset selection: The PTB dataset uses only 4 out of 9 classes, potentially leading to imbalance and lower representativeness, which impacts performance.
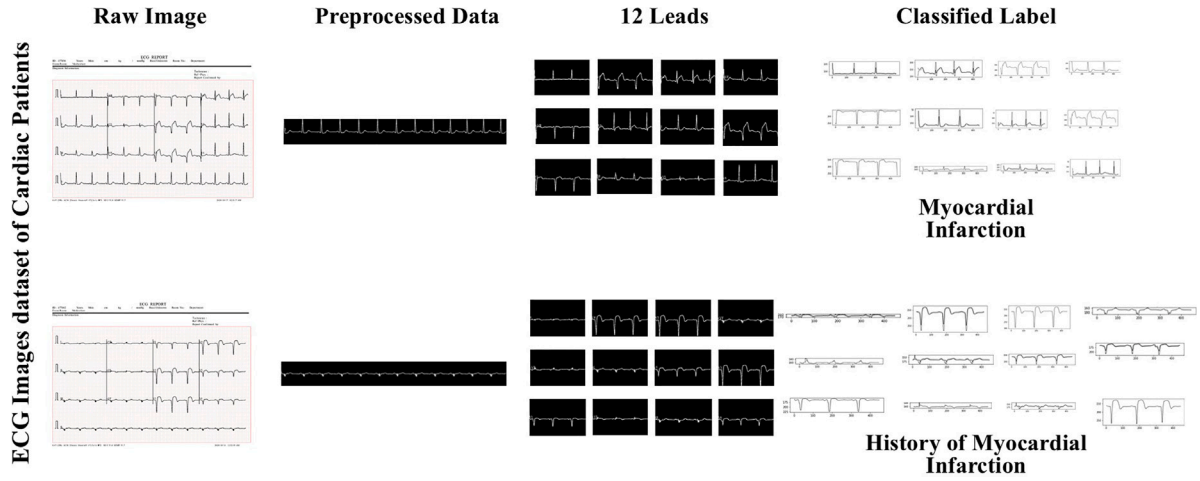
Fig. 8. Sample output images obtained for the ECG Images dataset of the Cardiac Patients dataset using the ASNet proposed model.

Table 9
Performance comparison of the proposed ASNet model with existing state-of-the-art approaches on ECG image datasets of cardiac patients.

Approach	$A_{AS_{net}}$	$P_{AS_{net}}$	$R_{AS_{net}}$	$F1_{AS_{net}}$	$MCC_{AS_{net}}$	$FMI_{AS_{net}}$	$G_{AS_{net}}$
DNN + Embedded Feature Selection [51]	88.12	89.07	91.37	90.11	0.5311	–	–
CardiAI [52]	90.71	86.80	95.51	90.11	–	–	–
Metaheuristic Algorithms [53]	89.11	90.24	91.29	90.77	–	0.877	–
ANN [54]	87.19	89.18	83.19	86.22	0.5219	–	–
ANN + SVM [55]	86.32	77.40	91.35	87.47	0.6443	–	–
ASNet (proposed)	91.31	89.72	92.54	91.14	0.7702	0.9419	0.8902

Dataset size and diversity: PTB contains 549 images compared to thousands in the ECG Images dataset of Cardiac Patients, reducing training diversity and model robustness. Lack of domain adaptation: Absence of domain adaptation or transfer learning techniques contributes to the observed performance decline across datasets.

To address these challenges more comprehensively in future work, mitigation strategies can be expanded to include domain adaptation methods such as adversarial alignment to reduce discrepancies between datasets; transfer learning to leverage large ECG datasets for enhanced robustness; and advanced data augmentation techniques specifically tailored for ECG images to increase dataset diversity and address class imbalance. Additionally, hybrid approaches that combine these techniques are proposed to further strengthen model resilience to domain shifts and improve ASNet's generalization across diverse ECG datasets. The ASNet model achieves an IoU score of 0.7685 on the PTB Diagnostic ECG Database. This shows moderate agreement between model highlights and expert annotations. It confirms the model's effectiveness in detecting key ECG features.

5.4. Comparative analysis of results obtained for heart disease classification

Table 9 compares the performance of various existing ECG classification models from the literature against the proposed ASNet model, evaluated on the ECG images dataset of cardiac patients.

DNN + Embedded Feature Selection [51]: ASNet outperforms this model in all major accuracy-related metrics ($A_{AS_{net}}$: 91.31% vs. 88.12%, $P_{AS_{net}}$: 89.72% vs. 89.07%, $R_{AS_{net}}$: 92.54% vs. 91.37%, $F1_{AS_{net}}$: 91.14% vs. 90.11%), and also achieves a significantly higher MCC (0.7702 vs. 0.5311), suggesting improved overall predictive quality. ASNet also reports additional robustness metrics like FMI and G-mean, which are not provided in the original baseline.

CardiAI [52]: Although CardiAI achieves a slightly higher recall (95.51% vs. 92.54%), ASNet demonstrates superior performance in accuracy (91.31% vs. 90.71%), precision (89.72% vs. 86.80%), and

Fig. 9. Confusion Matrix - Cornell University ECG Image Dataset.

F1-score (91.14% vs. 90.11%). Notably, CardiAI does not report other critical metrics like MCC, FMI, or G-mean, making a complete comparison difficult. ASNet's inclusion of these robustness metrics highlights its balanced and comprehensive performance.

Metaheuristic Algorithms [53]: ASNet achieves a higher accuracy (91.31% vs. 89.11%) and improved F1-score and recall (91.14% and 92.54% vs. 90.77% and 91.29%, respectively). Although the precision is comparable (89.72% vs. 90.24%), ASNet significantly outperforms in FMI (0.9419 vs. 0.877), indicating stronger performance in imbalanced settings.

ANN [54]: The ASNet model shows a notable improvement over this baseline, with around a 4% gain in accuracy and a significant
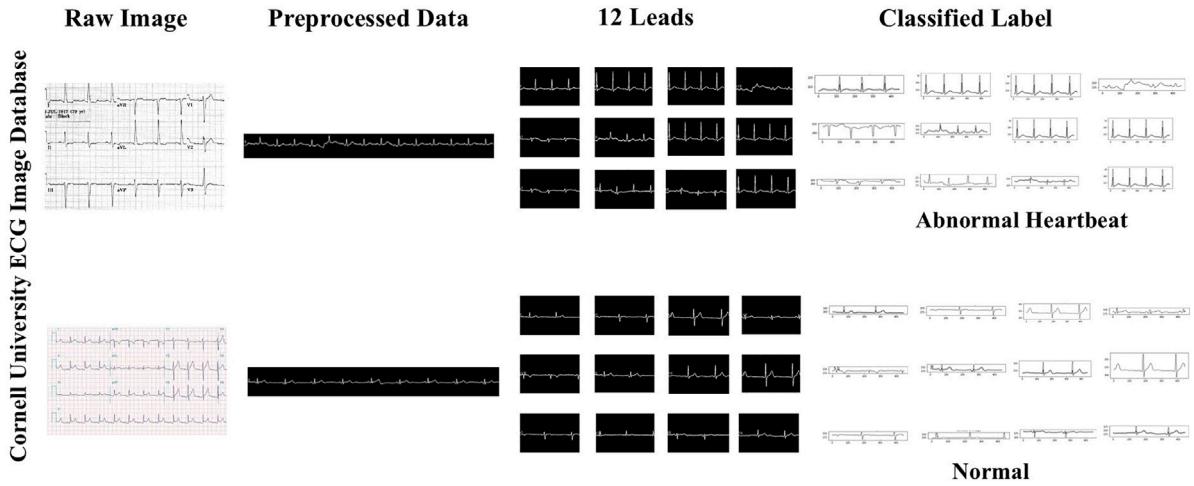
Fig. 10. Sample output images obtained for Cornell University dataset.

Fig. 11. Confusion Matrix - PTB Diagnostic ECG Database.

increase in recall (92.54% vs. 83.19%). The MCC also increases substantially (0.7702 vs. 0.5219), emphasizing ASNet's enhanced classification reliability.

ANN + SVM [55]: ASNet surpasses this hybrid model across all evaluated metrics, particularly in precision (89.72% vs. 77.40%) and F1-score (91.14% vs. 87.47%), while also showing a stronger MCC (0.7702 vs. 0.6443). This highlights the effectiveness of ASNet's integrated architecture in both feature extraction and classification.

To further assess the individual contributions of the Feature Transformer and Attentive Transformer modules within ASNet, an AUC-ROC benchmark was conducted as shown in Fig. 12. The base TabNet model achieves an AUC of 0.7552. Adding the Feature Transformer increases this to 0.7731, while the Attentive Transformer improves it to 0.7874. When both are combined (BM+FT+AT), the AUC reaches 0.8001. The full ASNet model, incorporating stackable mechanisms, achieves the highest AUC of 0.8153, demonstrating cumulative gains from the architectural enhancements.

5.5. Statistical analysis of the proposed ASNet model

To strengthen the evaluation protocol for the proposed ASNet model, we conducted repeated random train-test splits ($n = 10$) and computed mean performance with standard deviations, confidence intervals, and formal hypothesis tests. Table 10 represents the statistical

performance comparison of the proposed ASNet model. On the ECG Images dataset of Cardiac Patients, ASNet achieved an average accuracy of $91.40 \pm 0.24\%$ (95% CI [91.23, 91.58]), F1 of $91.17 \pm 0.36\%$, and MCC of 0.772 ± 0.003. On the Cornell ECG dataset, ASNet reached $92.79 \pm 0.47\%$ accuracy (95% CI [92.46, 93.13]), F1 of $91.58 \pm 0.26\%$, and MCC of 0.788 ± 0.004. On the PTB dataset, ASNet obtained $82.97 \pm 0.40\%$ accuracy (95% CI [82.68, 83.26]), F1 of $83.06 \pm 0.34\%$, and MCC of 0.805 ± 0.003. Across all datasets, χ^2 tests of independence on the confusion matrices (Fig. 5, 8, 10) were overwhelmingly significant (all $p < 0.0001$), confirming that predicted labels strongly depend on actual classes.

Friedman tests across five methods (BM, BM+FT, BM+AT, BM+FT +AT, and ASNet) consistently rejected the null hypothesis of no difference: for Cardiac Patients, $\chi^2 = 34.24$ ($p = 6.65 \times 10^{-7}$) on accuracy; for Cornell, $\chi^2 = 37.04$ ($p = 1.77 \times 10^{-7}$); and for PTB, $\chi^2 = 40.00$ ($p = 4.33 \times 10^{-8}$). Similar results held for F1 and MCC (all $p < 10^{-6}$). The combined reporting of mean ± standard deviation, confidence intervals, χ^2 tests, and Friedman tests enhances the rigor and reproducibility of our evaluation. Post-hoc Wilcoxon signed-rank tests confirmed that the ASNet model significantly outperformed all baselines across datasets. For the Cardiac dataset, differences remained significant even against the strongest baseline BM+FT+AT ($p = 0.027$). On Cornell and PTB datasets, ASNet achieved highly significant improvements over all baselines (all $p \leq 0.002$). These findings demonstrate that ASNet's improvements are not only numerically higher but also statistically significant, with stable variability across repeated runs.

6. Conclusion

In this study, the proposed ASNet model is introduced for effective heart disease classification, accurately categorizing ECG signals into four distinct classes: Normal, Myocardial Infarction (MI), History of MI, and Abnormal Heartbeat. This model is lightweight and easily applicable to real-time applications such as labs, clinics, or a hospital setting. ASNet delivers strong generalization and interpretability across multiple benchmark datasets. Extensive evaluation in three standard ECG datasets demonstrates the model's effectiveness. On the ECG Images dataset of Cardiac Patients, ASNet achieves an accuracy of 91.31%, F1-score of 91.14%, MCC of 0.7702, and FMI of 0.9419. When tested on the Cornell University ECG Image Database, the model outperforms its baseline with the highest accuracy of 92.71%, F1-score of 91.60%, MCC of 0.7871, and FMI of 0.9325. Lastly, evaluation on the PTB Diagnostic ECG Database yields a consistent performance with an accuracy of 83.00%, F1-score of 83.01%, MCC of 0.8038, and FMI of 0.8231.

Table 10

Statistical Performance comparison of different methods across datasets.

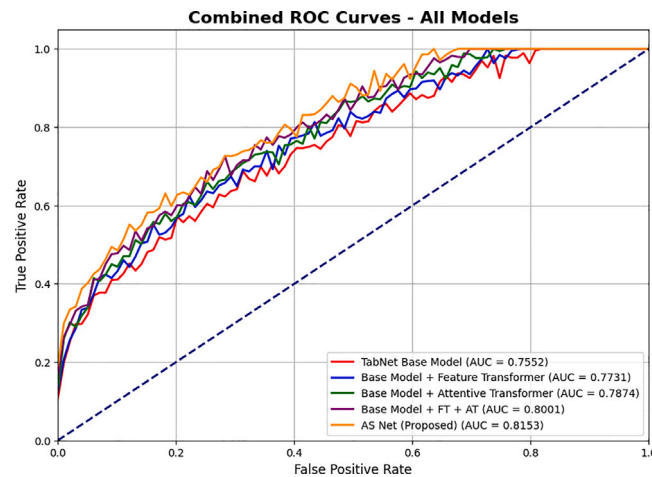| Dataset	Model	Accuracy (mean ± SD)	95% CI (Accuracy)	F1 (mean ± SD)	MCC (mean ± SD)
Cardiac Patients	BM	89.44 ± 0.46	[89.11, 89.76]	88.90 ± 0.40	0.720 ± 0.004
	BM+FT	90.05 ± 0.35	[89.82, 90.28]	89.27 ± 0.34	0.743 ± 0.004
	BM+AT	90.64 ± 0.31	[90.42, 90.87]	89.84 ± 0.34	0.753 ± 0.005
	BM+FT+AT	90.91 ± 0.35	[90.66, 91.16]	90.40 ± 0.49	0.761 ± 0.005
	ASNet	**91.40 ± 0.24**	**[91.23, 91.58]**	**91.17 ± 0.36**	**0.772 ± 0.003**
Cornell ECG	BM	88.33 ± 0.43	[88.02, 88.64]	87.40 ± 0.40	0.702 ± 0.004
	BM+FT	89.02 ± 0.35	[88.77, 89.28]	87.84 ± 0.34	0.729 ± 0.004
	BM+AT	89.78 ± 0.36	[89.52, 90.04]	88.56 ± 0.46	0.736 ± 0.003
	BM+FT+AT	90.23 ± 0.39	[89.95, 90.50]	89.08 ± 0.32	0.754 ± 0.003
	ASNet	**92.79 ± 0.47**	**[92.46, 93.13]**	**91.58 ± 0.26**	**0.788 ± 0.004**
PTB ECG	BM	74.90 ± 0.20	[74.76, 75.04]	75.07 ± 0.33	0.701 ± 0.003
	BM+FT	76.29 ± 0.35	[76.04, 76.54]	75.98 ± 0.20	0.722 ± 0.004
	BM+AT	77.90 ± 0.36	[77.64, 78.16]	77.96 ± 0.46	0.744 ± 0.006
	BM+FT+AT	79.37 ± 0.40	[79.08, 79.66]	78.96 ± 0.25	0.762 ± 0.004
	ASNet	**82.97 ± 0.40**	**[82.68, 83.26]**	**83.06 ± 0.34**	**0.805 ± 0.003**

Fig. 12. Illustrates the comparison of the AUC-ROC curve of various models with the proposed AS_{net} model.
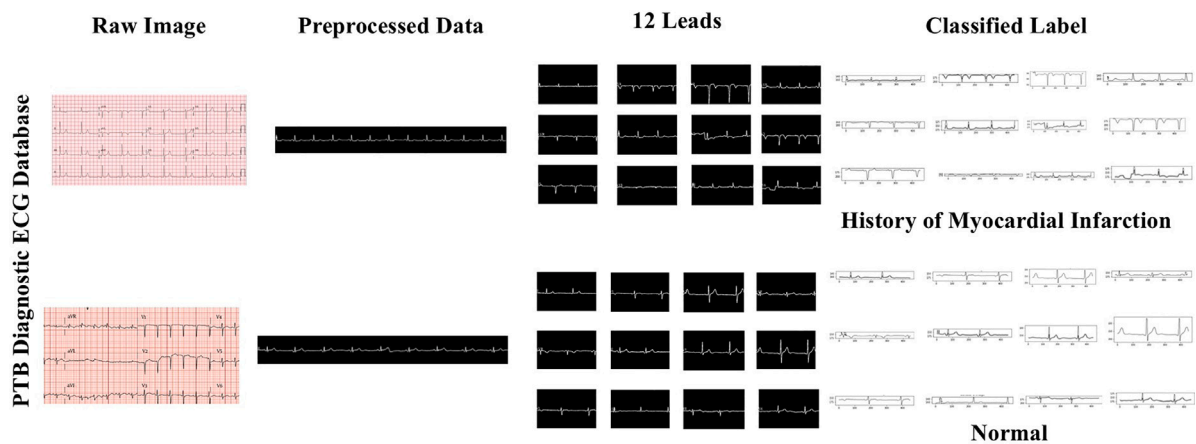

Fig. 13. Sample output images obtained for PTB Diagnostic dataset.

CRediT authorship contribution statement

Vishnu Venkatesh: Software, Methodology, Investigation. **S. Angalaeswari:** Supervision, Investigation. **Aravindkumar Sekar:** Writing – review & editing, Writing – original draft, Supervision, Methodology.

Declaration of competing interest

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript.

Data availability

Data will be made available on request.

References

[1] D.A. Cook, S.Y. Oh, M.V. Pusic, Accuracy of physicians' electrocardiogram interpretations: A systematic review and meta-analysis, JAMA Intern. Med. 180 (11) (2020) 1461–1471, http://dx.doi.org/10.1001/jamainternmed.2020.3989.

[2] D. Massel, Observer variability in ECG interpretation for thrombolysis eligibility: experience and context matter, J. Thromb. Thrombolysis 15 (3) (2003) 131–140, http://dx.doi.org/10.1023/B:THRO.0000011368.55165.97.

[3] A.A. Huitema, T. Zhu, M. Alemayehu, S. Lavi, Diagnostic accuracy of ST-segment elevation myocardial infarction by various healthcare providers, Int. J. Cardiol. 177 (3) (2014) 825–829, http://dx.doi.org/10.1016/j.ijcard.2014.11.032.

[4] A.P. Shah, S.A. Rubin, Errors in the computerized electrocardiogram interpretation of cardiac rhythm, J. Electrocardiol. 40 (5) (2007) 385–390, http://dx.doi.org/10.1016/j.jelectrocard.2007.03.008.

[5] M.N. Attar, K. Wong, D.G. Groves, N. Newall, D.R. Ramsdale, R.K. Moore, Clinical implications of QRS duration and QT peak prolongation in patients with suspected coronary disease referred for elective cardiac catheterization, Ann. Noninvasive Electrocardiol. 13 (2) (2008) 106–112, http://dx.doi.org/10.1111/j.1542-474X.2008.00209.x.

[6] S.J.G.C. Welten, et al., Prolongation of the heart rate-corrected QT interval is associated with cardiovascular diseases: Systematic review and meta-analysis, Arch. Cardiovasc. Dis. 116 (2) (2023) 69–78, http://dx.doi.org/10.1016/j.acvd.2022.11.007.

[7] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7 (2019) 81542–81554, http://dx.doi.org/10.1109/ACCESS.2019.2923707.

[8] A. Noor, N. Javaid, N. Alrajeh, B. Mansoor, A. Khaqan, S.H. Bouk, Heart disease prediction using stacking model with balancing techniques and dimensionality reduction, IEEE Access 11 (2023) 116026–116045, http://dx.doi.org/10.1109/ACCESS.2023.3325681.

[9] H. Khan, N. Javaid, T. Bashir, M. Akbar, N. Alrajeh, S. Aslam, Heart disease prediction using novel ensemble and blending based cardiovascular disease detection networks: EnsCVDD-Net and BlCVDD-Net, IEEE Access 12 (2024) 109230–109254, http://dx.doi.org/10.1109/ACCESS.2024.3421241.

[10] S. Mondal, R. Maity, Y. Omo, S. Ghosh, A. Nag, An efficient computational risk prediction model of heart diseases based on dual-stage stacked machine learning approaches, IEEE Access 12 (2024) 7255–7270, http://dx.doi.org/10.1109/ACCESS.2024.3350996.

[11] H.A. Al-Jamimi, Synergistic feature engineering and ensemble learning for early chronic disease prediction, IEEE Access 12 (2024) 62215–62233, http://dx.doi.org/10.1109/ACCESS.2024.3395512.

[12] M.S.H. Rabbi, M.M. Bari, T. Debnath, A. Rahman, A.K. Das, M.P. Hossain, G. Muhammad, Performance evaluation of optimal ensemble learning approaches with PCA and LDA-based feature extraction for heart disease prediction, Biomed. Signal Process. Control. 101 (2025) 107138, http://dx.doi.org/10.1016/j.bspc.2024.107138.

[13] S.G. Taj, K. Kalaivani, Hybrid prediction model with improved score level fusion for heart disease diagnosis, Comput. Biol. Chem. 113 (2024) 108278, http://dx.doi.org/10.1016/j.compbiolchem.2024.108278.

[14] I.D. Mienye, Y. Sun, Z. Wang, An improved ensemble learning approach for the prediction of heart disease risk, Inform. Med. Unlocked 20 (2020) 100402, http://dx.doi.org/10.1016/j.imu.2020.100402.

[15] I. Kaur, T. Ahmad, A cluster-based ensemble approach for congenital heart disease prediction, Comput. Methods Programs Biomed. 243 (2024) 107922, http://dx.doi.org/10.1016/j.cmpb.2023.107922.

[16] M.S.A. Reshan, S. Amin, M.A. Zeb, A. Sulaiman, H. Alshahrani, A. Shaikh, A robust heart disease prediction system using hybrid deep neural networks, IEEE Access 11 (2023) 121574–121591, http://dx.doi.org/10.1109/ACCESS.2023.3328909.

[17] Y. Pan, M. Fu, B. Cheng, X. Tao, J. Guo, Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform, IEEE Access 8 (2020) 189503–189512, http://dx.doi.org/10.1109/ACCESS.2020.3026214.

[18] B. Ramesh, K. Lakshmanna, A novel early detection and prevention of coronary heart disease framework using hybrid deep learning model and neural fuzzy inference system, IEEE Access 12 (2024) 26683–26695, http://dx.doi.org/10.1109/ACCESS.2024.3366537.

[19] N.S. Alghamdi, M. Zakariah, A. Shankar, W. Viriyasitavat, Heart disease prediction using autoencoder and DenseNet architecture, Egypt. Inform. J. 28 (2024) 100559, http://dx.doi.org/10.1016/j.eij.2024.100559.

[20] S.A. Alzakari, A.A. Menaem, N. Omer, A. Abozeid, L.F. Hussein, I.A.M. Abass, R. Ayadi, A. Elhadad, Enhanced heart disease prediction in remote healthcare monitoring using IoT-enabled cloud-based XGBoost and Bi-LSTM, Alex. Eng. J. 105 (2024) 280–291, http://dx.doi.org/10.1016/j.aej.2024.06.036.

[21] Y.K. Saheed, T.T. Salau-Ibrahim, M. Abdulsalam, I.A. Adeniji, B.F. Balogun, Modified bi-directional long short-term memory and hyperparameter tuning of supervised machine learning models for cardiovascular heart disease prediction in mobile cloud environment, Biomed. Signal Process. Control. 94 (2024) 106319, http://dx.doi.org/10.1016/j.bspc.2024.106319.

[22] V. Baviskar, M. Verma, P. Chatterjee, G. Singal, Efficient heart disease prediction using hybrid deep learning classification models, IRBM 44 (5) (2023) 100786, http://dx.doi.org/10.1016/j.irbm.2023.100786.

[23] R.R. Aswathi, K.P. Kumar, B. Ramakrishnan, An innovative attention infused-BiRecurrenTwin network assisted hybrid segmentation technique for accurate heart disease prediction, Eng. Appl. Artif. Intell. 138 (2024) 109389, http://dx.doi.org/10.1016/j.engappai.2024.109389.

[24] A. Jafar, M. Lee, HypGB: High accuracy GB classifier for predicting heart disease with HyperOpt HPO framework and LASSO FS method, IEEE Access 11 (2023) 138201–138214, http://dx.doi.org/10.1109/ACCESS.2023.3339225.

[25] S. Ghorashi, K. Rehman, A. Riaz, H.K. Alkahtani, A.H. Samak, I. Cherrez-Ojeda, A. Parveen, Leveraging regression analysis to predict overlapping symptoms of cardiovascular diseases, IEEE Access 11 (2023) 60254–60266, http://dx.doi.org/10.1109/ACCESS.2023.3286311.

[26] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, R. Nour, An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection, IEEE Access 7 (2019) 180235–180243, http://dx.doi.org/10.1109/ACCESS.2019.2952107.

[27] A. Javeed, Zhou, An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection, IEEE Access 7 (2019) 180235–180243, http://dx.doi.org/10.1109/ACCESS.2019.2952107.

[28] M.A. Bouqentar, O. Terrada, S. Hamida, S. Saleh, D. Lamrani, B. Cherradi, A. Raihani, Early heart disease prediction using feature engineering and machine learning algorithms, Heliyon 10 (19) (2024) e38731, http://dx.doi.org/10.1016/j.heliyon.2024.e38731.

[29] M.A. Islam, M.Z.H. Majumder, M.S. Miah, S. Jannaty, Precision healthcare: A deep dive into machine learning algorithms and feature selection strategies for accurate heart disease prediction, Comput. Biol. Med. 176 (2024) 108432, http://dx.doi.org/10.1016/j.compbiomed.2024.108432.

[30] G. Manikandan, B. Pragadeesh, V. Manojkumar, A.L. Karthikeyan, R. Manikandan, A.H. Gandomi, Classification models combined with Boruta feature selection for heart disease prediction, Inform. Med. Unlocked 44 (2024) 101442, http://dx.doi.org/10.1016/j.imu.2023.101442.

[31] N.L. Fitriyani, M. Syafrudin, G. Alfian, J. Rhee, HDPM: An effective heart disease prediction model for a clinical decision support system, IEEE Access 8 (2020) 133034–133050, http://dx.doi.org/10.1109/ACCESS.2020.3010511.

[32] C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han, J. Yu, Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the internet of medical things platform, IEEE Access 8 (2020) 59247–59256, http://dx.doi.org/10.1109/ACCESS.2020.2981159.

[33] D.Y. Omkari, K. Shaik, An integrated two-layered voting (TLV) framework for coronary artery disease prediction using machine learning classifiers, IEEE Access 12 (2024) 56275–56290, http://dx.doi.org/10.1109/ACCESS.2024.3389707.

[34] M. Wei, J. Yang, Z. Zhao, X. Zhang, J. Li, Z. Deng, DeFedHDP: Fully decentralized online federated learning for heart disease prediction in computational health systems, IEEE Trans. Comput. Soc. Syst. 11 (5) (2024) 6854–6867, http://dx.doi.org/10.1109/TCSS.2024.3406528.

[35] P. Bizopoulos, D. Koutsouris, Deep learning in cardiology, IEEE Rev. Biomed. Eng. 12 (2019) 168–193, http://dx.doi.org/10.1109/RBME.2018.2885714.

[36] A. Bellot, M. van der Schaar, A hierarchical Bayesian model for personalized survival predictions, IEEE J. Biomed. Health Inf. 23 (1) (2019) 72–80, http://dx.doi.org/10.1109/JBHI.2018.2832599.

[37] V. Chole, M. Thawakar, M. Choudhari, S. Chahande, S. Verma, A. Pimpalkar, Enhancing heart disease risk prediction with GdHO fused layered BiLSTM and HRV features: A dynamic approach, Biomed. Signal Process. Control. 95 (2024) 106470, http://dx.doi.org/10.1016/j.bspc.2024.106470.

[38] J. Qiu, Z. Chang, K. Wang, K. Chen, Q. Wang, J. Zhang, J. Li, C. Yang, Y. Zhao, Y. Zhang, The predictive accuracy of coronary heart disease risk prediction models in rural Northwestern China, Prev. Med. Rep. 36 (2023) 102503, http://dx.doi.org/10.1016/j.pmedr.2023.102503.

[39] G. Rajkumar, T. Gayathri Devi, A. Srinivasan, Heart disease prediction using IoT based framework and improved deep learning approach: Medical application, Med. Eng. Phys. 111 (2023) 103937, http://dx.doi.org/10.1016/j.medengphy.2022.103937.

[40] H.F. El-Sofany, Predicting heart diseases using machine learning and different data classification techniques, IEEE Access 12 (2024) 106146–106160, http://dx.doi.org/10.1109/ACCESS.2024.3437181.

[41] Y.S. Chichani, S.L. Kasar, An efficient IoT enabled heart disease prediction model using Finch hunt optimization modified BiLSTM classifier, Biomed. Signal Process. Control. 100 (2025) 107170, http://dx.doi.org/10.1016/j.bspc.2024.107170.

[42] A.H. Khan, M. Hussain, ECG Images Dataset of Cardiac Patients, Mendeley Data, Version 2, 2021. http://dx.doi.org/10.17632/gwbz3fsgp8.2.

[43] M.A. Reyna, et al., ECG-image-database: A dataset of ECG images with real-world imaging and scanning artifacts; A foundation for computerized ECG image digitization and analysis, 2024, arXiv, [Online]. Available: https://arxiv.org/abs/2409.16612.

[44] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P.C. Ivanov, R. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220, http://dx.doi.org/10.1161/01.CIR.101.23.e215.

[45] L. Khriji, M. Fradi, M. Machhout, A. Hossen, Deep learning-based approach for atrial fibrillation detection, in: M. Jmaiel, M. Mokhtari, B. Abdulrazak, H. Aloulou, S. Kallel (Eds.), The Impact of Digital Technologies on Public Health in Developed and Developing Countries, Springer International Publishing, Cham, 2020, pp. 100–113, http://dx.doi.org/10.1007/978-3-030-51517-1_9.

[46] T.M. Ingolfsson, X. Wang, M. Hersche, A. Burrello, L. Cavigelli, L. Benini, ECG-TCN: Wearable cardiac Arrhythmia detection with a temporal convolutional network, in: 2021 IEEE 3rd Int. Conf. Artificial Intelligence Circuits and Systems, AICAS, 2021, pp. 1–4, http://dx.doi.org/10.1109/AICAS51828.2021.9458520.

[47] A. Wang, W. Xu, H. Sun, N. Pu, Z. Liu, H. Liu, Arrhythmia classifier using binarized convolutional neural network for resource-constrained devices, in: 2022 4th Int. Conf. Communications, Information System and Computer Engineering, CISCE, 2022, pp. 213–220, http://dx.doi.org/10.1109/CISCE55963.2022.9851002.

[48] P. Busia, M.A. Scrugli, V.J.-B. Jung, L. Benini, P. Meloni, A tiny transformer for low-power Arrhythmia classification on microcontrollers, IEEE Trans. Biomed. Circuits Syst. 19 (1) (2025) 142–152, http://dx.doi.org/10.1109/TBCAS.2024.3401858.

[49] G. Silva, P. Silva, G. Moreira, V. Freitas, J. Gertrudes, E. Luz, A systematic review of ECG arrhythmia classification: Adherence to standards, fair evaluation, and embedded feasibility, 2025, arXiv, [Online]. Available: https://arxiv.org/abs/2503.07276.

[50] A. Eleyan, E. Alboghbaish, Electrocardiogram signals classification using deep-learning-based incorporated convolutional neural network and long short-term memory framework, Computers 13 (2) (2024) http://dx.doi.org/10.3390/computers13020055, Art. no. 55.

[51] D. Zhang, Y. Chen, Y. Chen, Heart disease prediction based on the embedded feature selection method and deep neural network, J. Heal. Eng. (2021) http://dx.doi.org/10.1155/2021/6260022.

[52] A.T. Mulaguri, S. Katta, V.K. Karanam, Y. Pachipala, S. Narisety, Enhancing heart disease prediction using CardiAI: With key performance metrics accuracy, precision, recall and F1-score, Int. J. Intell. Syst. Appl. Eng. 12 (18s) (2024) 646–655, https://ijisae.org/index.php/IJISAE/article/view/5013.

[53] H. Zhang, R. Mu, Refining heart disease prediction accuracy using hybrid machine learning techniques with novel metaheuristic algorithms, Int. J. Cardiol. 416 (2024) 132506, http://dx.doi.org/10.1016/j.ijcard.2024.132506.

[54] R.R. Sarra, A.M. Dinar, M.A. Mohammed, Enhanced accuracy for heart disease prediction using artificial neural network, Indones. J. Electr. Eng. Comput. Sci. 29 (2023) 375–383, http://dx.doi.org/10.11591/ijeecs.v29.i1.pp375-383.

[55] R.R. Sarra, A.M. Dinar, M.A. Mohammed, Enhanced accuracy for heart disease prediction using artificial neural network, Indones. J. Electr. Eng. Comput. Sci. (IJEECS) 29 (1) (2023) 375–383, http://dx.doi.org/10.11591/ijeecs.v29.i1.pp375-383.