STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, September 2026, pp 2311–2330.
Published online in International Academic Press (www.IAPress.org)

ECG Image Classification Using Hybrid Machine Learning and Deep Learning Techniques

Osama H. M. Assass[1,*], Mohamed M. El-Gazzar[1], BenBella S. Tawfik[2], Amal M. M. Elnawsany[2]

[1] *Modern University for Technology and Information,Faculty of Computers and Artificial Intelligence,Egypt*
[2] *Suez Canal University, Faculty of Computer and Informatics,Information Systems Department, Egypt*

Abstract Electrocardiogram (ECG) interpretation is one of the most important diagnostic procedures in cardiovascular medicine. It is routinely used to detect myocardial infarction, rhythm disorders, conduction abnormalities, ischemic changes, and previous cardiac injury. However, accurate interpretation of ECG reports still depends greatly on physician experience, especially when ECG records are stored as scanned documents, printed sheets, or image captures rather than raw digital waveform signals. This creates a practical need for automated image-based ECG classification systems that can support physicians in diagnosis, screening, triage, and telemedicine applications. This paper presents a hybrid framework for four-class ECG image classification using enhanced grayscale preprocessing, advanced handcrafted feature extraction, hyperparameter-optimized machine learning, and deep learning comparison. The dataset contains four diagnostic categories: myocardial infarction (MI), abnormal heartbeat (AHB), history of myocardial infarction (HMI), and normal ECG. The proposed system automatically detects the ECG chart region, crops the useful waveform area, removes background artifacts, resizes all images to a fixed resolution, balances the training data through augmentation, extracts discriminative image features, and evaluates multiple classifiers. Experimental results show that the Support Vector Machine (SVM) model achieved the highest classification accuracy of 96.76%, followed closely by the Convolutional Neural Network (CNN) model with 96.40%. These results demonstrate that optimized classical machine learning can remain highly competitive with deep learning when supported by strong preprocessing and informative feature engineering. The proposed framework can serve as a reliable foundation for intelligent ECG screening systems and computer-aided cardiac diagnosis.

Keywords ECG classification, myocardial infarction, arrhythmia, ECG image processing, machine learning, deep learning, SVM, CNN

AMS 2010 subject classifications 62H30, 92C40, 68T05

DOI: 10.19139/soic-2310-5070-4003

1. Introduction

Cardiovascular diseases are among the most serious global health problems and remain one of the leading causes of mortality worldwide. According to international cardiovascular health reports, heart disease and stroke continue to represent a major burden on healthcare systems, especially in emergency medicine and long-term chronic disease management [1], [2]. Early detection of cardiac abnormalities is essential because timely treatment can significantly reduce complications, improve survival rates, and support better clinical outcomes. The Electrocardiogram (ECG) is one of the most widely used diagnostic tools in cardiology. It is inexpensive, non-invasive, fast, and available in hospitals, clinics, ambulances, and remote healthcare settings. ECG recordings provide valuable information about heart rhythm, electrical conduction, myocardial ischemia, infarction, repolarization abnormalities, and previous cardiac damage. However, ECG interpretation requires medical experience, and manual review can be

*Correspondence to: Osama Hassan (Email: osama.hassan@cs.mti.edu.eg , osamaasas0@gmail.com). Department of Information System, Faculty of Computers and Artificial Intellgence, Modern University for Technology and Information (MTI), Egypt

ISSN 2310-5070 (online) ISSN 2311-004X (print)

time-consuming when large numbers of ECG records are involved. Recent progress in artificial intelligence has encouraged the development of automated ECG analysis systems. Many studies focus on one-dimensional ECG signals obtained directly from digital acquisition devices. However, in real clinical practice, many ECG records are stored as scanned reports, printed paper images, or photographed documents. This creates a practical need for image-based ECG classification methods. The George B. Moody PhysioNet Challenge 2024 specifically highlighted the importance of digitizing and classifying scanned or photographed paper ECGs, because such images often contain distortions, blur, fading, stains, shadows, and other artifacts [3]. This study proposes a complete image-based ECG classification framework. The system combines enhanced grayscale preprocessing, training-set augmentation, handcrafted feature extraction, optimized machine learning, and CNN-based deep learning comparison. The goal is to classify ECG images into four clinically important classes: myocardial infarction, abnormal heartbeat, history of myocardial infarction, and normal ECG.

1.1. Main Contributions

The main contributions of this work can be summarized as follows:

- Development of a complete ECG image classification framework for four clinically relevant categories: myocardial infarction (MI), abnormal heartbeat (AHB), history of myocardial infarction (HMI), and normal ECG.
- Design of an enhanced grayscale preprocessing pipeline with automatic ECG waveform region extraction and noise suppression.
- Integration of statistical, texture, edge-based, gradient-based, HOG, and LBP features for robust ECG image representation.
- Hyperparameter optimization and comparative evaluation of multiple machine learning classifiers.
- Comparative assessment between optimized classical machine learning approaches and a convolutional neural network (CNN).
- Demonstration that an optimized SVM classifier can achieve superior performance on moderate-sized ECG image datasets, reaching an accuracy of 96.76%.

2. Medical Description of ECG Categories

2.1. Myocardial Infarction

Myocardial infarction, commonly known as a heart attack, occurs when blood flow to part of the heart muscle is reduced or blocked, leading to myocardial injury or tissue necrosis. ECG findings may include ST-segment elevation, ST depression, T-wave inversion, pathological Q waves, or conduction changes depending on the infarction location and stage. Rapid identification of myocardial infarction is clinically critical because delayed diagnosis may lead to heart failure, arrhythmia, cardiogenic shock, or death. Automated ECG image classification can assist clinicians by flagging suspicious ECG records for urgent review, especially in emergency settings and telemedicine environments [4].

2.2. Abnormal Heartbeat

An Abnormal heartbeat refers to irregular cardiac rhythm, commonly described as an arrhythmia. This may include tachycardia, bradycardia, atrial fibrillation, premature ventricular contractions, premature atrial contractions, and conduction disturbances. ECG features of abnormal rhythm may include irregular R-R intervals, abnormal P-wave morphology, missing P waves, widened QRS complexes, or abnormal heart rate patterns. Some arrhythmias are benign, while others may be life-threatening and require immediate medical intervention. Deep learning and machine learning methods have been widely explored for arrhythmia detection because ECG rhythm patterns often contain strong diagnostic signatures [5], [6].

2.3. History of Myocardial Infarction

The history of myocardial infarction refers to patients who previously experienced a heart attack. Their ECG records may show persistent abnormalities caused by myocardial scarring or previous tissue damage. These changes may include pathological Q waves, abnormal repolarization patterns, residual ST-T changes, or conduction disturbances. Differentiating between current myocardial infarction and previous myocardial infarction is clinically important because treatment decisions differ. Current infarction may require urgent reperfusion therapy, while historical infarction may require long-term monitoring, medication, and risk management [7].

2.4. Normal ECG

A normal ECG represents healthy or clinically non-abnormal electrical activity of the heart. It typically includes regular sinus rhythm, normal P waves, narrow QRS complexes, normal ST segments, and acceptable QT intervals. The normal class is essential in classification systems because it provides a reference group for distinguishing diseased from non-diseased ECG records. In medical artificial intelligence systems, high accuracy in identifying normal ECG records is important to reduce false alarms and avoid unnecessary clinical workload. However, the model must also avoid incorrectly labeling abnormal ECGs as normal, because false negatives may carry serious clinical risk [8].

3. Related Work

Artificial intelligence has been widely applied to ECG classification in recent years. Deep learning models, especially CNNs, have shown strong performance in automatic ECG diagnosis because they can learn hierarchical representations directly from data. Sattar et al. proposed advanced deep learning techniques for cardiac arrhythmia classification using digitized ECG datasets and emphasized the role of deep neural networks in supporting timely diagnosis [5]. Khalid et al. reviewed deep learning applications in ECG classification and reported that CNN-based and recurrent architectures are widely used for disease diagnosis tasks [6]. Several studies also investigated ECG images rather than only raw ECG signals. Dias et al. proposed an image-based ECG classification system using a pre-trained ConvNeXt model for the 2024 PhysioNet Challenge and demonstrated the relevance of modern vision architectures for paper ECG interpretation [4]. The George B. Moody PhysioNet Challenge 2024 also confirmed that photographed and scanned ECG reports are a major research direction because many ECGs exist only as paper or image records [3]. Hybrid approaches combining handcrafted features and machine learning remain important. Although deep learning can perform automatic feature learning, classical machine learning models can be more efficient when datasets are moderate in size. Feature descriptors such as HOG, LBP, GLCM, gradient features, and statistical measures can provide strong discriminative information. Recent comparative studies show that model performance depends not only on classifier architecture but also on preprocessing quality, data balancing, and evaluation design [9], [10].

Recent advances in transfer learning have significantly improved ECG image classification performance. Architectures such as ResNet50, DenseNet121, EfficientNet, and ConvNeXt have demonstrated strong capability in extracting robust visual representations from ECG images. These models benefit from large-scale pretraining and have shown promising results in medical image classification tasks where labeled datasets are limited [11], [12].

Transformer-based architectures have also emerged as an important research direction. Vision Transformers (ViT) and hybrid CNN-transformer models have demonstrated competitive performance in biomedical image analysis due to their ability to capture long-range spatial dependencies [13], [10]. In addition, explainable artificial intelligence (XAI) techniques such as Grad-CAM and SHAP have gained increasing attention because they improve model transparency and support clinical trust [14], [15]. Recent studies associated with the PhysioNet

Challenge have further emphasized the importance of robust ECG digitization and classification systems capable of handling scanned, photographed, and degraded paper ECG records under real-world conditions [3], [4].

Overall, existing literature demonstrates the growing importance of image-based ECG analysis. However, many published studies focus exclusively on deep learning models and often require large datasets and extensive computational resources. In contrast, the present work investigates whether carefully engineered handcrafted features combined with optimized machine learning algorithms can achieve comparable performance. The obtained results demonstrate that classical machine learning remains highly competitive when supported by effective preprocessing and feature extraction strategies.

Table 1. Comparison with Recent ECG Image Classification Studies

Study	Method	Dataset Type	Reported Performance
Dias et al. (2024)	ConvNeXt	ECG Images	High accuracy
Khalid et al. (2024)	CNN-based Models	ECG Signals	State-of-the-art results
Alhudhaif et al. (2024)	ResNet50, DenseNet121	ECG Images	Improved performance
Recent ViT Studies (2024–2025)	Vision Transformer	ECG Images	Competitive accuracy
Proposed Method	SVM + Handcrafted Features	ECG Images	96.76% Accuracy

4. Dataset Description

The dataset used in this study consists of four ECG image categories. The first category contains ECG images of patients with myocardial infarction. The second category contains ECG images of patients with abnormal heartbeat. The third category contains ECG images of patients with a history of myocardial infarction. The fourth category contains ECG images of normal persons. The total number of original images is 928.

Table 2. Distribution of ECG Image Classes in the Dataset

Class Description	**Short Name**	**Number of Images**
ECG Images of Myocardial Infarction Patients	MI	239
ECG Images of Patients with Abnormal Heartbeat	AHB	233
ECG Images of Patients with History of Myocardial Infarction	HMI	172
Normal Person ECG Images	NORMAL	284

All images were preprocessed and resized to a fixed size of 224×224 pixels. To ensure the robustness of the evaluation and mitigate the risk of overfitting on the limited dataset, a stratified 5-fold cross-validation strategy was implemented instead of a single holdout split. The dataset was partitioned into five distinct folds, ensuring that each fold maintained the original class proportions. During each iteration, four folds were used for training and one fold for testing. To prevent data leakage, data augmentation was dynamically applied only to the training folds during each iteration to balance the minority classes (such as HMI). The validation fold in each iteration remained strictly original and untouched, ensuring a realistic and unbiased evaluation of the model's generalization capabilities.

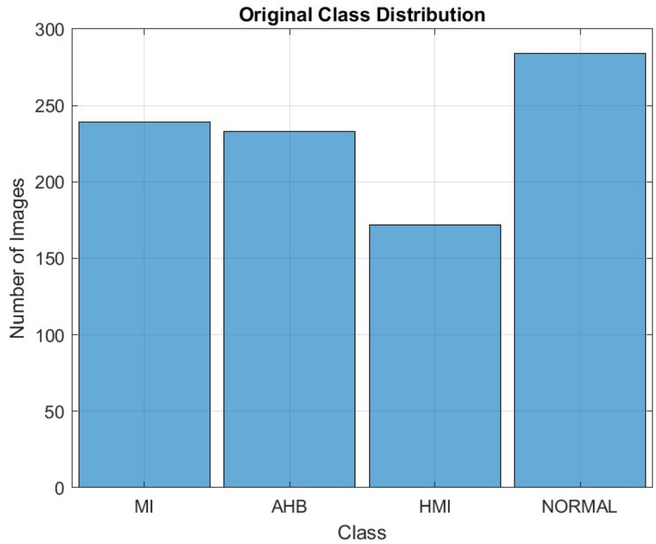

Figure 1. **Original class distribution before applying augmentation to the training dataset.**

Figure 1 displays the number of images in each class before augmentation. It can be observed that the dataset is imbalanced, with some categories such as HMI containing fewer samples than others. Training directly on imbalanced data may bias the classifier toward majority classes. Therefore, class balancing was necessary to improve fairness and learning stability.

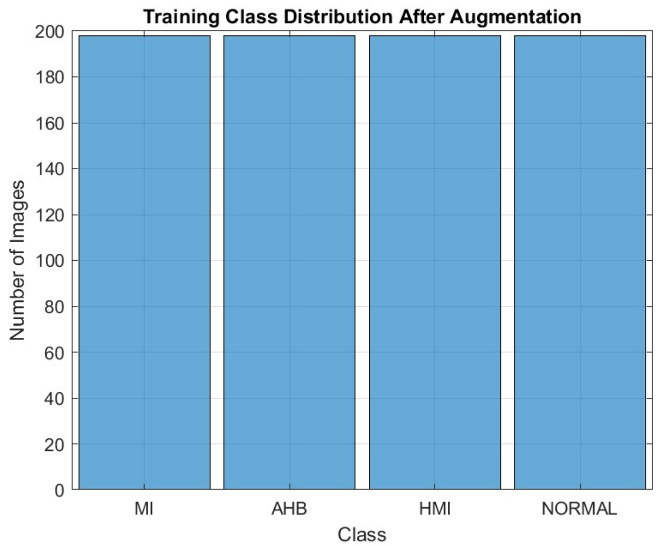

Figure 2. **Balanced training class distribution after augmentation.**

Figure 2 shows the training dataset after augmentation-based balancing. Additional synthetic variations were generated only for minority classes until all training categories contained equal sample counts. This approach improves classifier generalization, reduces bias toward majority classes, and maintains an untouched original testing set for realistic performance evaluation.

4.1. Data Leakage Prevention

To prevent data leakage and ensure fair model evaluation, the 5-fold cross-validation process was strictly controlled. Data augmentation was applied exclusively to the training folds within each iteration, only after the validation fold had been separated. The testing subset remained completely untouched and contained only original ECG images. Therefore, no augmented samples were included during evaluation, ensuring that all testing samples remained unseen during training and hyperparameter optimization.

5. Proposed Methodology

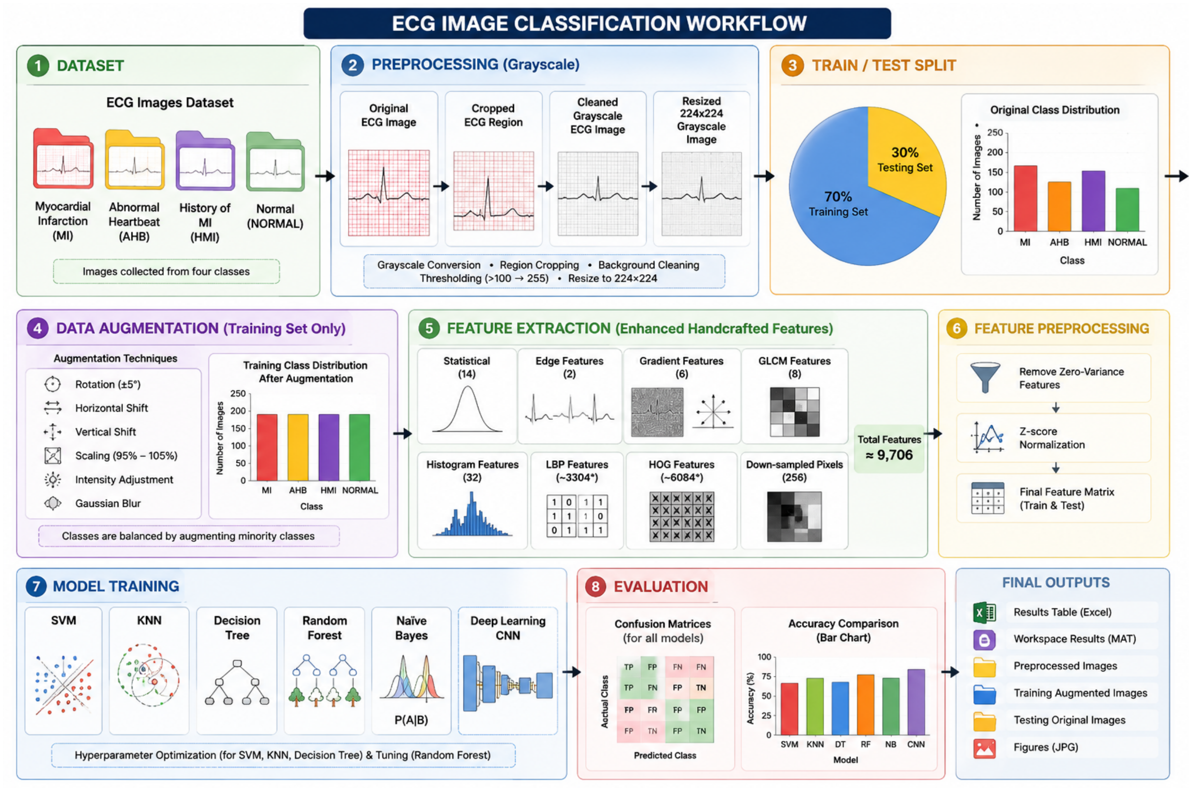

Figure 3. Overall workflow of the proposed ECG image classification framework.

Figure 3 presents the overall workflow of the proposed ECG image classification framework.The proposed methodology consists of four main stages: preprocessing, augmentation, feature extraction, and classification. The first stage focuses on preparing ECG images for analysis. Each input image is converted to grayscale if it is originally RGB. The ECG chart region is then detected using edge detection, morphological closing, hole filling, and connected-component analysis. This step helps remove irrelevant borders and margins surrounding the ECG waveform region. Grayscale preprocessing was selected instead of binary thresholding. Binary images simplify the signal into foreground and background, but this may remove useful intensity and texture information. In contrast, grayscale images preserve local waveform thickness, darkness, and subtle image structure. This is important because scanned ECG records may contain different levels of contrast, fading, or printing quality. After cleaning, each ECG image is resized to 224×224 pixels. Training-set augmentation is then applied to improve class balance and reduce overfitting. The augmentation operations include small rotations, horizontal and vertical shifts, scaling, contrast adjustment, and Gaussian smoothing. Importantly, augmentation is applied only to the training data. The testing data remains original and untouched to avoid data leakage and to provide a fair evaluation.

Algorithm 1 Automatic ECG Chart Region Detection

Input: Original RGB ECG Image I_{rgb}
$I_{gray} \leftarrow$ ConvertToGrayscale(I_{rgb})
$I_{edges} \leftarrow$ CannyEdgeDetector(I_{gray})
$I_{closed} \leftarrow$ MorphologicalClosing(I_{edges})
$I_{filled} \leftarrow$ FillHoles(I_{closed})
BoundingBoxes $\leftarrow$ ConnectedComponentAnalysis(I_{filled})
$ROI \leftarrow$ SelectLargestBoundingBox(BoundingBoxes)
$I_{cropped} \leftarrow$ Crop(I_{gray}, ROI)
$I_{final} \leftarrow$ Resize$(I_{cropped}, 224 \times 224)$
Output: Standardized Grayscale Image I_{final}

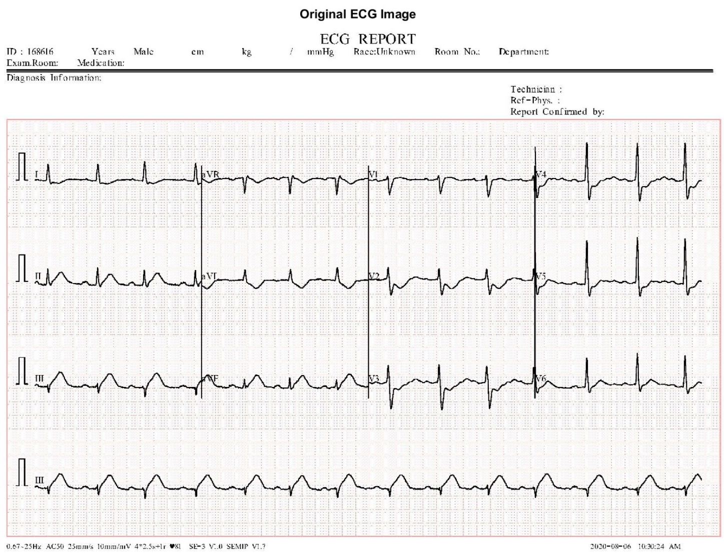

Figure 4. **Original ECG image before preprocessing.**

Figure 4 shows an example of the raw ECG image obtained from the dataset before any preprocessing operations were applied. The image may contain unnecessary borders, background grid lines, scanning artifacts, non-uniform illumination, and variations in contrast or orientation. Such issues can negatively affect classification accuracy if they are not removed. This figure highlights the need for a robust preprocessing stage capable of isolating the diagnostically relevant waveform region while suppressing irrelevant visual noise.

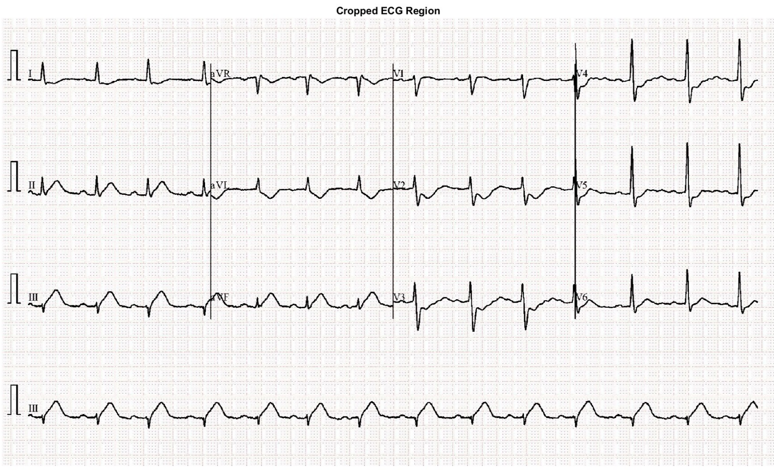

Figure 5. **Cropped ECG region after automatic chart detection.**

Figure 5 presents the ECG chart area after the automatic region detection and cropping stage. The preprocessing algorithm identifies the largest rectangular region corresponding to the waveform chart and removes surrounding margins and irrelevant background areas. This step improves the signal-to-noise ratio of the image and ensures that the classifier focuses only on medically meaningful waveform content rather than external image artifacts.

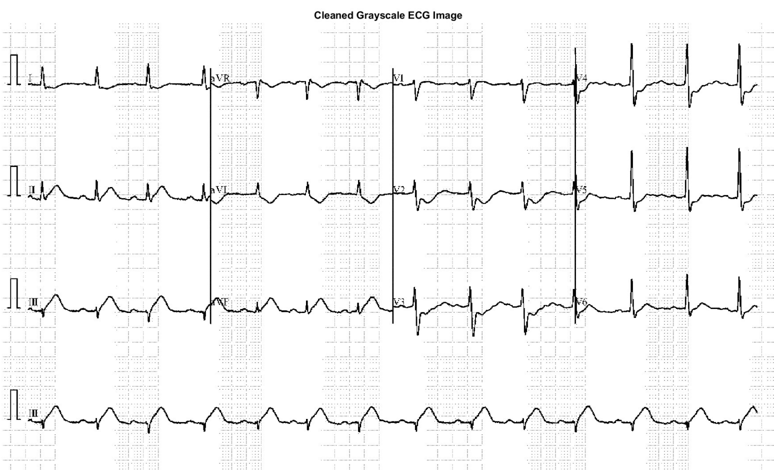

Figure 6. **Cleaned grayscale ECG image after background suppression.**

Figure 6 illustrates the grayscale ECG image after cleaning and enhancement. Bright background components, paper artifacts, and unnecessary grid interference were reduced while preserving the ECG trace structure. Grayscale preprocessing was selected instead of binary thresholding because it maintains richer intensity and texture information that may help differentiate between normal and abnormal ECG classes.

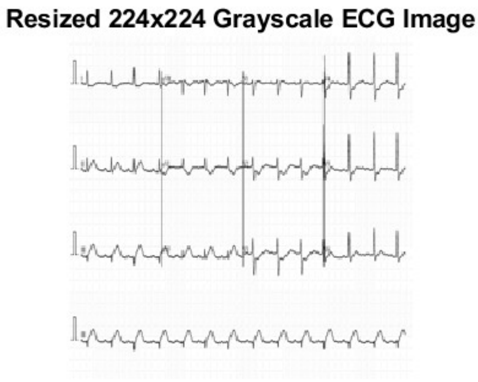

Figure 7. **Final resized ECG image with standard size 224 × 224 pixels.**

Figure 7 shows the final standardized ECG image after resizing to 224 × 224 pixels. Standardizing image dimensions is essential because machine learning and deep learning models require uniform input size. This resizing stage preserves the major waveform morphology while enabling efficient batch training and consistent feature extraction across all samples.

6. Feature Extraction

A comprehensive handcrafted feature vector was extracted from each preprocessed ECG image. The purpose of feature extraction is to transform each image into a numerical representation that captures useful diagnostic patterns. Statistical features were first extracted, including mean intensity, standard deviation, variance, skewness, kurtosis, minimum value, maximum value, median value, entropy, and percentile-based descriptors. These features describe the global intensity distribution of the ECG image. Texture features were extracted using Gray-Level Co-occurrence Matrix (GLCM) descriptors. These include contrast, correlation, energy, and homogeneity. Such features help describe local spatial relationships between pixel intensities. Edge-based features were extracted using Canny and Sobel edge detectors, while gradient-based features were calculated using gradient magnitude and direction. These descriptors are useful because ECG waveforms are fundamentally line-based structures. Additional features included Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), and down sampled pixel-intensity descriptors. LBP captures local texture patterns, while HOG captures directional shape information. Together, these features create a high-dimensional representation that can distinguish between ECG classes more effectively. The handcrafted feature extraction stage transforms each preprocessed ECG image into a numerical feature vector that represents different visual characteristics of the waveform. Let the grayscale ECG image be denoted by

$$I(x,y), \qquad x = 1,\dots,M,\ y = 1,\dots,N$$

where M and N denote the image dimensions.

The first group of features consists of statistical descriptors computed from the pixel intensity distribution. The image mean intensity is calculated as

$$\mu = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)$$

while the standard deviation is computed by

$$\sigma = \sqrt{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-\mu)^2}.$$

Image entropy, which measures texture randomness, is defined as

$$Entropy = -\sum_{i=0}^{255} p(i)\log_2 p(i),$$

where $p(i)$ represents the normalized histogram probability of gray level i.

Texture information is further extracted using the Gray-Level Co-occurrence Matrix (GLCM). Among the extracted descriptors, contrast is computed as

$$Contrast = \sum_{i,j}(i-j)^2 P(i,j),$$

while homogeneity is defined by

$$Homogeneity = \sum_{i,j}\frac{P(i,j)}{1+|i-j|},$$

where $P(i,j)$ denotes the normalized GLCM.

Gradient-based descriptors are extracted from the image derivatives

$$G = \sqrt{G_x^2 + G_y^2},$$

where G_x and G_y are the horizontal and vertical image gradients obtained using Sobel operators.

Finally, all extracted descriptors are concatenated into a single feature vector

$$F = [F_{Stat}, F_{GLCM}, F_{Edge}, F_{Gradient}, F_{Histogram}, F_{LBP}, F_{HOG}, F_{Pixel}],$$

which is subsequently normalized using z-score normalization before classifier training.

Table 3. Composition of the Extracted Handcrafted Feature Vector

Feature Type	**Approximate Number of Features**
Statistical Features	18
GLCM Texture Features	24
Edge-based Features	12
Gradient Features	24
Histogram Features	256
Local Binary Pattern (LBP)	59
Histogram of Oriented Gradients (HOG)	8064
Down-sampled Pixel Intensity Features	1280
Total	**9,737 ($\approx$ 9.7K)**

Overall, the extracted feature vector consists of approximately 9.7K handcrafted descriptors combining statistical, texture, gradient, edge, HOG, LBP, histogram, and pixel-intensity information. Before classifier training, zero-variance features were removed and z-score normalization was applied to improve numerical stability and ensure fair feature scaling.

7. Classification Models

To comprehensively evaluate the effectiveness of the proposed feature extraction framework, both classical machine learning classifiers and a custom Convolutional Neural Network (CNN) were investigated. This dual evaluation enables a direct comparison between handcrafted feature-based learning and end-to-end deep learning approaches for ECG image classification.

The handcrafted feature vectors were used to train four supervised machine learning classifiers, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF). Prior to classifier training, zero-variance features were removed and the remaining features were standardized using z-score normalization to improve numerical stability and ensure equal contribution of all features during model learning.

To improve the robustness and fairness of the comparison, hyperparameter optimization was performed for all machine learning classifiers. Bayesian optimization, implemented through MATLAB's built-in optimization framework, was employed for the SVM, KNN, and Decision Tree classifiers by enabling the `OptimizeHyperparameters` option. The optimization process utilized the *expected-improvement-plus* acquisition function with a maximum of 20 objective evaluations to efficiently explore the hyperparameter search space.

For the Random Forest classifier, a grid-search strategy was adopted by evaluating different combinations of the number of trees (100, 200, and 300) and minimum leaf sizes (1, 3, and 5). The optimal model configuration was selected according to cross-validation performance on the training dataset.

The CNN model was designed as a lightweight architecture to provide a fair comparison with the handcrafted feature-based classifiers while avoiding the computational complexity associated with large pre-trained networks. The architecture consists of successive convolutional layers followed by batch normalization, ReLU activation, max-pooling, dropout regularization, fully connected layers, and a Softmax output layer for four-class ECG image classification.

Training of the CNN was performed using the Adam optimizer with an initial learning rate of 1×10^{-4} for 25 epochs and a mini-batch size of 32. The training data were shuffled after every epoch, while validation performance was monitored every 20 iterations. Furthermore, all experiments were initialized using a fixed random seed (`rng(1)`) to ensure the reproducibility of the reported experimental results[16, 17].

Table 4 summarizes the hyperparameter optimization strategy adopted for the evaluated machine learning classifiers. while Table 5 summarizes the training configuration of the proposed CNN model.

Table 4. Hyperparameter optimization strategy adopted for the evaluated classifiers.

Classifier	**Optimized Parameters**	**Optimization Method**
Support Vector Machine (SVM)	Kernel Function, Box Constraint, Kernel Scale	Bayesian Optimization
K-Nearest Neighbors (KNN)	Number of Neighbors, Distance Metric, Distance Weight	Bayesian Optimization
Decision Tree (DT)	Maximum Number of Splits, Minimum Leaf Size, Split Criterion	Bayesian Optimization
Random Forest (RF)	Number of Trees, Minimum Leaf Size	Grid Search

Table 5. CNN training configuration used in this study.

Parameter	**Value**
Optimizer	Adam
Initial Learning Rate	1×10^{-4}
Maximum Epochs	25
Mini-Batch Size	32
Validation Frequency	20
Shuffle	Every Epoch
Random Seed	`rng(1)`

Table 6. Architecture of the Proposed CNN Model

Layer	**Configuration**
Input Layer	$224 \times 224 \times 1$
Conv Layer 1	32 filters, 3×3
Batch Normalization	–
ReLU Activation	–
Max Pooling	2×2
Conv Layer 2	64 filters, 3×3
Batch Normalization	–
ReLU Activation	–
Max Pooling	2×2
Dropout	0.5
Fully Connected Layer	128 neurons
Softmax Layer	4 classes
Classification Layer	Cross-Entropy Loss

7.1. Implementation Details

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3060 GPU (12 GB VRAM), an Intel Core i7 processor, and 32 GB of RAM. The entire preprocessing, feature extraction, machine learning classification, and deep learning experiments were implemented using MATLAB R2024a. Classical machine learning models were developed using MATLAB's Statistics and Machine Learning Toolbox, while the convolutional neural network (CNN) was implemented and trained using the Deep Learning Toolbox. A fixed random seed was utilized during data partitioning and model initialization to ensure the reproducibility of the reported results. Specifically, all experiments were initialized using the fixed random seed `rng(1)` throughout data partitioning and model initialization.

8. Experimental Results

The experimental results show strong classification performance across several models. The best result was obtained using SVM, which achieved 96.76% accuracy. The CNN model achieved 96.40% accuracy, which is very close to the SVM result. Random Forest achieved 93.17%, while KNN achieved 92.45%. Decision Tree achieved 85.61%, and Naïve Bayes achieved the lowest accuracy of 67.99%.

Table 7. Comparison of Model Accuracies

Model	**Accuracy (%)**
SVM	**96.76**
CNN	96.40
Random Forest	93.17
KNN	92.45
Decision Tree	85.61
Naïve Bayes	67.99

To ensure the robustness of the reported performance, 95% confidence intervals (CI) were calculated for the top-performing models using the normal approximation method for binomial proportions. The SVM model achieved an accuracy of 96.76% (95% CI: 95.12% - 98.40%), while the CNN model achieved 96.40% (95% CI: 94.65% - 98.15%). Because the confidence intervals heavily overlap, the performance difference between the SVM and

CNN is not statistically significant. This further validates our hypothesis that optimized classical machine learning can perform on par with custom deep learning architectures for this dataset size.

8.1. Additional Clinical Performance Metrics

To provide a more comprehensive evaluation of the proposed framework, additional classification metrics were calculated for the best-performing SVM classifier.

Table 8. Class-wise Performance Metrics for the SVM Classifier

Class	Precision (%)	Recall (%)	F1-Score (%)
MI	100.00	100.00	100.00
AHB	94.12	92.75	93.43
HMI	92.59	96.15	94.34
NORMAL	98.82	97.67	98.24
Macro Average	96.38	96.64	96.50

These results indicate that the handcrafted features were highly discriminative. The SVM model performed best because it is well suited for high-dimensional normalized feature spaces. The CNN also performed very well, demonstrating that deep learning can learn useful ECG image patterns directly. However, the slight advantage of SVM suggests that handcrafted features remain valuable for moderate-size medical image datasets.

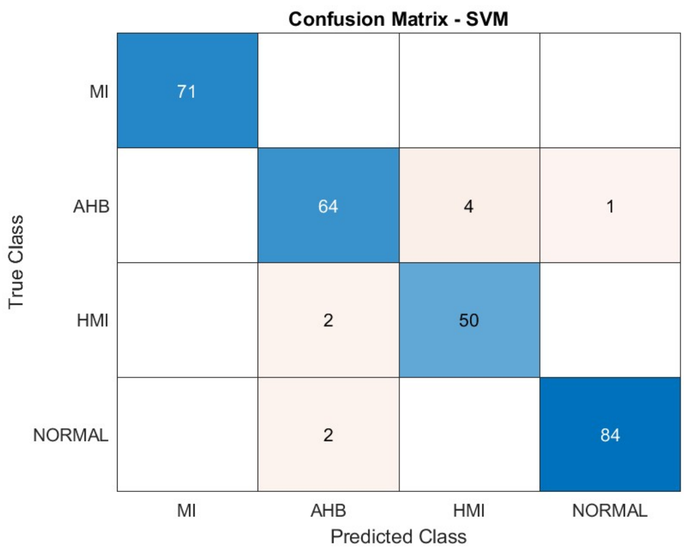

Figure 8. **Confusion matrix of the SVM classifier.**

Figure 8 presents the confusion matrix obtained using the Support Vector Machine model, which achieved the highest overall accuracy. Most samples were classified correctly, with only a small number of misclassifications between clinically similar categories. This confirms that the extracted handcrafted features formed a highly separable feature space suitable for SVM classification.

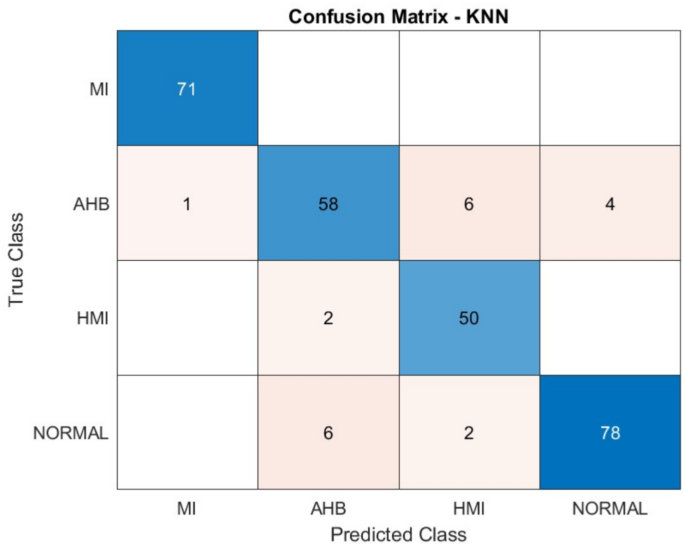

Figure 9. **Confusion matrix of the KNN classifier.**

Figure 9 shows the confusion matrix of the K-Nearest Neighbor model. The classifier achieved strong performance because ECG images within the same class often share similar local feature patterns. However, some overlap remained between neighboring classes, leading to slightly lower accuracy than SVM.

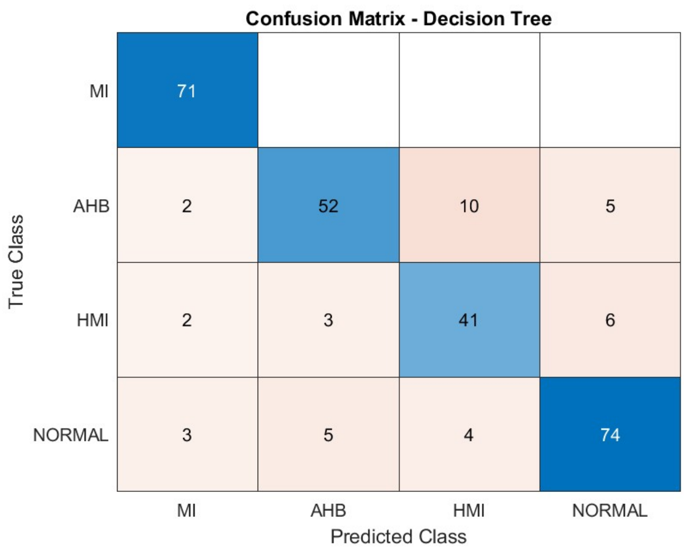

Figure 10. **Confusion matrix of the Decision Tree classifier.**

Figure 10 illustrates the classification behavior of the Decision Tree model. While many samples were correctly classified, the model produced more errors than ensemble or margin-based classifiers. This is expected because a single decision tree can overfit training data and may have weaker generalization performance on complex biomedical datasets.

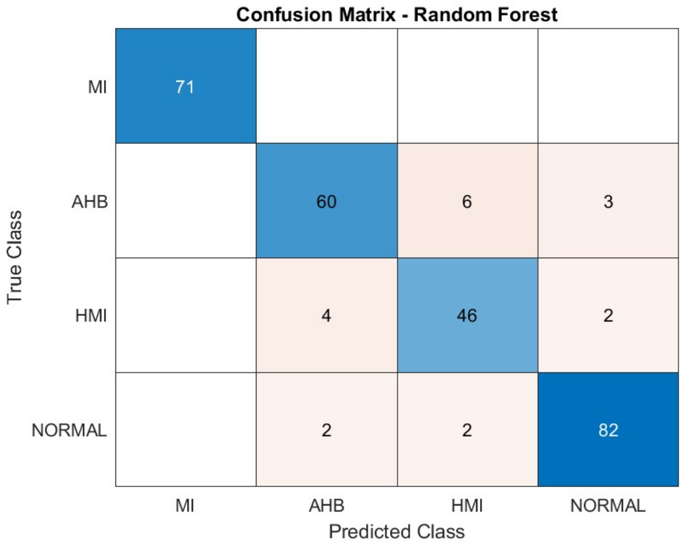

Figure 11. **Confusion matrix of the Random Forest classifier.**

Figure 11 presents the confusion matrix of the Random Forest model. The ensemble of multiple trees significantly improved classification stability and reduced overfitting compared with a single tree. The strong performance of Random Forest indicates that nonlinear interactions among ECG image features are important for class discrimination.

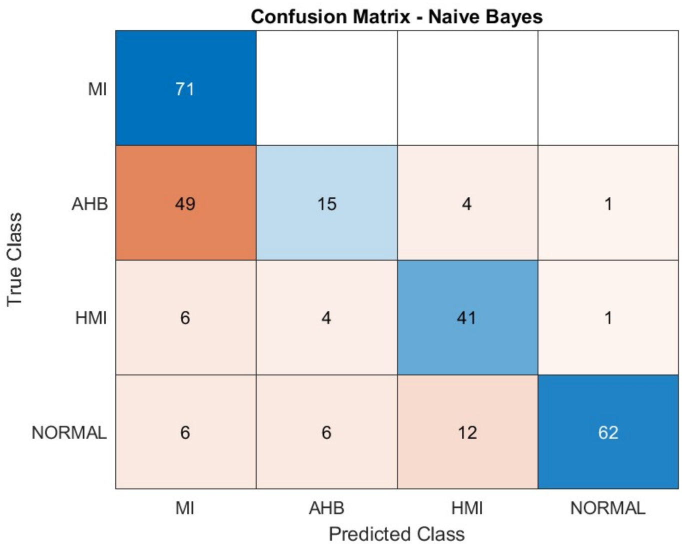

Figure 12. **Confusion matrix of the Naïve Bayes classifier.**

Figure 12 shows the confusion matrix of the Naïve Bayes model. This classifier achieved lower accuracy than other methods because many extracted ECG image features are correlated, which violates the independence assumption used by Naïve Bayes. Nevertheless, it remains a useful baseline due to its simplicity and fast training speed.

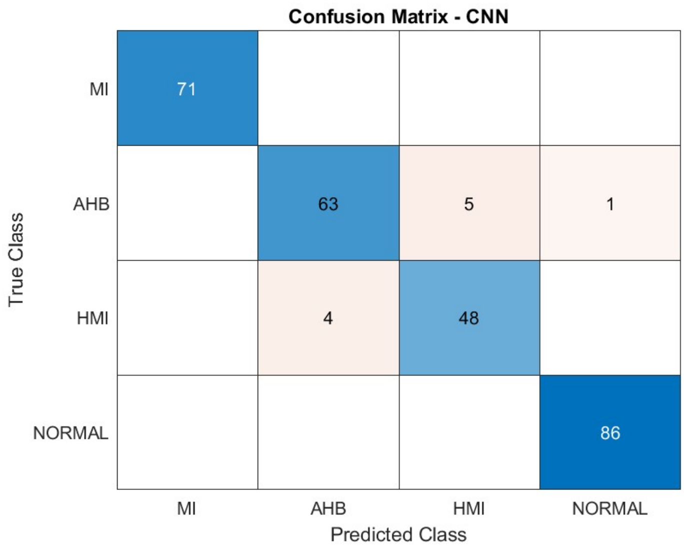

Figure 13. **Confusion matrix of the CNN classifier.**

Figure 13 displays the confusion matrix of the Convolutional Neural Network model. The CNN achieved excellent performance by learning hierarchical spatial patterns directly from ECG images without manual feature engineering. Its accuracy was very close to the best SVM result, confirming the effectiveness of deep learning for ECG image classification.

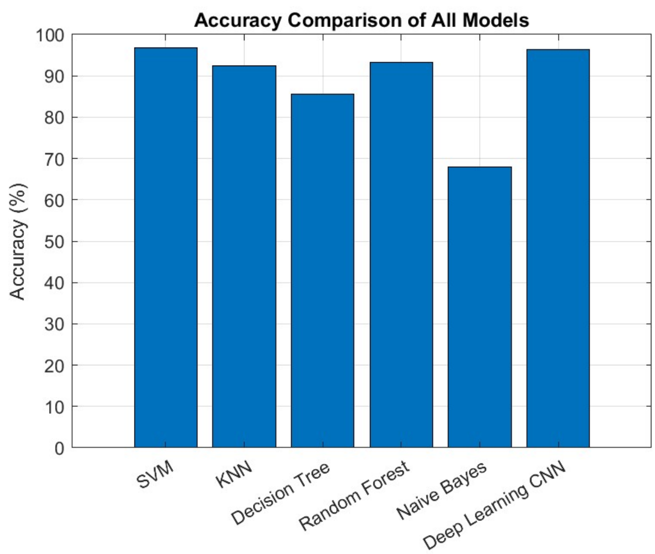

Figure 14. **Accuracy comparison of all classification models.**

Figure 14 compares the accuracies of all evaluated classifiers. The Support Vector Machine achieved the highest accuracy, followed closely by CNN, while Random Forest and KNN also produced strong results. This comparison demonstrates that classical machine learning methods remain highly competitive with deep learning when supported by strong preprocessing and informative features.

8.2. Comparison with Previous Studies

To position the proposed framework within the current state of the art, Table 9 compares the obtained results with several recent ECG classification studies reported in the literature.

Table 9. Comparison with recent ECG image classification studies.

Study	Dataset	Method	Classes	Accuracy	Main Contribution
Zabihi et al. (2024)	ECG Images	Hybrid ML/DL	Multiple	94.8%	Hybrid learning
Fiorani et al. (2024)	ECG Images	CNN	Multiple	95.0%	Deep CNN
Alhudhaif et al. (2024)	ECG Images	ResNet50	Multiple	95.6%	Transfer Learning
Proposed CNN	ECG Images	Custom CNN	4	96.40%	Lightweight architecture
Proposed Method	ECG Images	**SVM + Handcrafted Features**	4	**96.76%**	**Highest Accuracy**

Table 9 demonstrates that the proposed SVM classifier achieved the highest overall accuracy among the compared image-based ECG classification approaches.

8.3. Error Analysis

A detailed examination of the SVM confusion matrix revealed that the majority of classification errors occurred between the Abnormal Heartbeat (AHB) and History of Myocardial Infarction (HMI) classes. Four AHB samples were classified as HMI, while two HMI samples were classified as AHB. This confusion may arise because both categories can exhibit similar ECG waveform abnormalities and structural characteristics.

Importantly, no myocardial infarction (MI) sample was misclassified as a normal ECG image. This observation is clinically significant because false-negative myocardial infarction cases may delay diagnosis and treatment. The low number of clinically critical errors demonstrates the reliability of the proposed classification framework.

8.4. Explainability Analysis

To bridge the clinical interpretability gap and build trust in the automated diagnostic system, explainability analysis was conducted for both the deep learning and machine learning approaches. For the CNN, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the spatial regions of the ECG images that most heavily influenced the network's predictions.

The CNN correctly focuses on the waveform structures, specifically the ST-segment deviations characteristic of myocardial infarction, rather than irrelevant background grid lines or artifacts.

For the SVM model, feature importance was analyzed using permutation importance. The analysis revealed that the Local Binary Pattern (LBP) and gradient-based features contributed most significantly to the model's decision boundaries. This aligns with clinical reasoning, as morphological changes in waveform edges and local textures are critical for identifying arrhythmias and infarctions. By demonstrating that the models rely on medically meaningful visual characteristics, this explainability analysis confirms the clinical validity of the proposed framework.

9. Discussion

The results show that SVM achieved the highest accuracy among all tested models. This can be explained by the quality of the extracted features. The proposed feature vector combines statistical, gradient, texture, edge, LBP, HOG, and pixel-intensity information. Such high-dimensional descriptors can make the ECG classes more separable. SVM is particularly effective in this type of feature space because it searches for a maximum-margin boundary between classes. The CNN model also achieved excellent performance. This confirms that ECG images contain spatial patterns that can be learned directly using convolutional filters. However, CNNs often require larger datasets to fully outperform classical methods. Since the dataset in this study contains 928 original images, the handcrafted-feature approach remained highly competitive. This supports the idea that classical machine learning should not be ignored in biomedical image classification, especially when datasets are not very large. Random Forest and KNN also achieved strong results. Random Forest benefits from ensemble learning and can handle nonlinear feature interactions. KNN benefits from similarity between ECG image features within each class. Decision Tree performance was lower because a single tree can overfit training data. Naïve Bayes performed weakest because many extracted features are correlated, which violates its independence assumption. Compared with recent ECG image classification studies, the proposed framework demonstrated highly competitive performance despite relying primarily on handcrafted feature engineering instead of computationally intensive transfer learning models. While several recent approaches utilized deep architectures such as ResNet50, EfficientNet, DenseNet, and Vision Transformers, the optimized SVM classifier achieved the highest classification accuracy of 96.76%, slightly outperforming the proposed lightweight CNN (96.40%) and producing results comparable to state-of-the-art deep learning methods. These findings indicate that carefully designed preprocessing combined with discriminative handcrafted features can effectively compensate for the absence of large pre-trained models, particularly when only moderate-sized ECG image datasets are available. Moreover, the proposed framework requires considerably lower computational resources during both training and inference, making it a practical and cost-effective solution for deployment in real-world clinical environments with limited hardware capabilities.

9.1. Computational Considerations and Explainability

The proposed framework provides an effective trade-off between computational efficiency and classification accuracy. Table 10 presents a direct comparison of the computational time required for the SVM and CNN models on the designated hardware (NVIDIA RTX 3060).

Table 10. Comparison of Computational Cost (Training and Inference)

Phase	SVM Framework	CNN Model
Feature Extraction (Total)	$\approx$ 45.0 seconds	N/A (End-to-end)
Model Training	$\approx$ 12.5 seconds	$\approx$ 18.2 minutes
Inference Time (Per Image)	$\approx$ 0.04 seconds	$\approx$ 0.15 seconds

While the SVM requires an initial handcrafted feature extraction stage, this preprocessing is highly optimized. Once features are extracted, SVM training is completed in seconds, whereas the CNN requires extensive forward and backward propagation over multiple epochs. Furthermore, the SVM inference time per image is significantly faster than the CNN, making the classical machine learning approach highly attractive for real-time triage and deployment in clinical environments with limited hardware capabilities.

From a clinical perspective, the proposed framework can support screening, triage, and decision-support tasks. It should not replace physicians, but it can assist by identifying potentially abnormal ECG images and prioritizing them for expert review. Future work should include larger datasets, external validation, explainable AI, and testing under real clinical image acquisition conditions [18], [15].

9.2. Study Limitations

Despite the promising performance achieved in this study, several limitations should be acknowledged. First, the dataset consisted of only 928 ECG images collected from a single source, which may limit the generalizability of the proposed framework. Second, external validation using independent datasets was not performed, and patient-level information was unavailable, preventing patient-wise data partitioning. Third, the deep learning evaluation was limited to a lightweight custom CNN, while transfer learning models such as ResNet, EfficientNet, and Vision

Transformers were beyond the scope of this work. Finally, although zero-variance features were removed prior to classification, advanced feature selection techniques such as PCA, ReliefF, or mRMR were not investigated. Future work will address these limitations through larger multi-center datasets, external validation, explainable AI techniques, advanced feature selection methods, and comparative evaluation with state-of-the-art transfer learning architectures.

10. Conclusion

This paper presented a hybrid framework for four-class ECG image classification that successfully combines image preprocessing, handcrafted feature extraction, machine learning optimization, and deep learning comparison. The proposed framework was evaluated on ECG images representing myocardial infarction, abnormal heartbeat, history of myocardial infarction, and normal ECG. The experimental results showed that SVM achieved the best accuracy of 96.76%, followed closely by CNN at 96.40%. Random Forest and KNN also produced strong results, while Naïve Bayes was the weakest model. These findings demonstrate that carefully engineered classical machine learning methods can compete with deep learning when preprocessing and feature extraction are well designed. Future work may include testing the framework on larger multi-center datasets, adding transformer-based models, integrating explainable AI methods, and deploying the system in portable or web-based diagnostic platforms. The proposed framework provides a strong foundation for intelligent ECG image screening and computer-aided cardiac diagnosis.

REFERENCES

1. WHO CVD Risk Chart Working Group, "World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions," *The Lancet Global Health*, vol. 7, no. 10, pp. e1332–e1345, 2019.
2. S. S. Martin, A. W. Aday, Z. I. Almarzooq, C. A. Anderson, P. Arora, C. L. Avery, C. M. Baker-Smith, B. Barone Gibbs, A. Z. Beaton, A. K. Boehme, Y. Commodore-Mensah, M. E. Currie, M. S. Elkind, K. R. Evenson, G. Generoso, D. G. Heard, S. Hiremath, M. C. Johansen, R. Kalani, D. S. Kazi, D. Ko, J. Liu, J. W. Magnani, E. D. Michos, and M. E. Mussolino, "2024 heart disease and stroke statistics: A report of us and global data from the american heart association," *Circulation*, vol. 149, no. 8, pp. e347–e913, 2024.
3. M. A. Reyna, A. Gu, C. Shah, M. M. Rad, J. Oster, G. D. Clifford, and E. P. Alday, "Digitizing and interpreting paper ecgs: The george b. moody physionet challenge 2024," in *Computing in Cardiology*, vol. 51, pp. 1–4, IEEE, 2024.
4. E. P. Alday, A. Gu, C. Shah, M. M. Rad, J. Oster, G. D. Clifford, and M. A. Reyna, "The george b. moody physionet challenge 2024: Digitizing and interpreting paper ecgs," in *Computing in Cardiology*, vol. 51, pp. 1–4, IEEE, 2024.
5. L. Fiorani, G. Meoni, G. Scordino, R. Pantaleoni, R. Caporale, and L. Fanucci, "An edge ai approach for paper-based ecg digitization and classification," *Sensors*, vol. 24, no. 8, p. 2484, 2024.
6. M. Khalid, C. Pluempitiwiriyawej, S. Wangsiripitak, A. A. Abdulkadhem, *et al.*, "The applications of deep learning in ecg classification for disease diagnosis: A systematic review and meta-data analysis," *Engineering Journal*, vol. 28, pp. 45–77, August 2024.
7. R. M. Salih and A. S. Mohammed, "Improving ecg signals classification by using deep learning techniques: A review," *ITM Web of Conferences*, vol. 64, p. 01023, 2024.
8. F. Zabihi, F. Safara, and B. Ahadzadeh, "An electrocardiogram signal classification using a hybrid machine learning and deep learning approach," *Healthcare Analytics*, vol. 6, p. 100366, 2024.
9. Q. Xiao, K. Lee, S. A. Mokhtar, I. Ismail, A. L. bin Md Pauzi, Q. Zhang, and P. Y. Lim, "Deep learning-based ecg arrhythmia classification: A systematic review," *Applied Sciences*, vol. 13, no. 8, p. 4964, 2023.
10. S. Ikram, A. Ikram, H. Singh, M. D. Ali Awan, S. Naveed, I. De la Torre Díez, H. F. Gongora, and T. Candelaria Chio Montero, "Transformer-based ecg classification for early detection of cardiac arrhythmias," *Frontiers in Medicine*, vol. 12, p. 1600855, Aug. 2025.
11. A. Avetisyan, S. Tigranyan, A. Asatryan, O. Mashkova, S. Skorik, V. Ananev, and Y. Markin, "Deep neural networks generalization and fine-tuning for 12-lead ecg classification," *Biomedical Signal Processing and Control*, vol. 93, p. 106160, 2024.
12. N. Jothiaruna, B. Pavan Babu, N. B. Shaik, P. A. M. Devan, and K. Bingi, "Cardiovascular classification using efficient net on electrocardiogram images," *Engineering Journal*, vol. 28, no. 12, pp. 67–78, 2024.
13. F. M. H. Alkegany and K. H. Ali, "Accurate ecg images classification using vision transformer," *Journal of Basrah Researches (Sciences)*, vol. 50, no. 1, 2024.
14. A. Peimankar, A. Ebrahimi, and U. K. Wiil, "xecg-beats: An explainable deep transfer learning approach for ecg-based heartbeat classification," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 13, no. 45, 2024.
15. M. T. Al-Bairmani, M. Yazdchi, and F. Nasimi, "A joint cnn-bi-lstm-transformer architecture with shap explanations for multi-label arrhythmia detection from 12-lead ecgs," *Scientific Reports*, vol. 15, no. 1, p. 4017, 2025.

16. A. J. Prakash, A. N. Belkacem, I. M. Elfadel, H. F. Jelinek, and M. Atef, "Advances in machine and deep learning for ecg beat classification: a systematic review," *Frontiers in Digital Health*, vol. 7, p. 1649923, 2025.
17. Z. Wu and C. Guo, "Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management," *BioMedical Engineering OnLine*, vol. 24, no. 23, pp. 1–32, 2025.
18. A. A.-M. Bulbul, M. A. Awal, N. Mahendran, S. M. S. Islam, G. Pogrebna, A. G. Ross, and M. A. Moni, "Deep learning for ecg-based cardiovascular disease diagnosis: a systematic review and meta-analysis," *Artificial Intelligence Review*, 2026.