European Journal of Engineering and Applied Sciences

ISSN: 2651-3412 (Print) & 2667-8454 (Online)
Journal homepage: http://dergipark.gov.tr/EJEAS
Published by Çorlu Faculty of Engineering, Tekirdağ Namık Kemal University

European J. Eng. App. Sci. 8(1), 14-23, 2025

Research Article

Detection of Different Cardiac Conditions with Machine Learning Using Wavelet Transform and GLCM Feature Fusion in ECG Images

Kadircan Karaca[1,a], Esra Sivari Resul[1,b,*], Mustafa Karhan[1,c]

[1]**Department of Computer Engineering, Engineering Faculty, Cankiri Karatekin University, Cankiri, Türkiye, 18100**
[a]kadircankaraca01@gmail.com, [b]esrasivari@karatekin.edu.tr, [c]mustafakarhan@karatekin.edu.tr

Received: 13.02.2025 Accepted: 15.04.2025 DOI: 10.55581/ejeas.1639148

Abstract: Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, accounting for 32% of global deaths. Electrocardiography (ECG) is a widely used, cost-effective, and non-invasive diagnostic tool for detecting cardiac abnormalities. However, ECG interpretation remains challenging due to noise interference, physiological variations, and the need for expert evaluation. This study proposes a machine learning-based approach for automatic classification of cardiac conditions using ECG images. The methodology involves feature extraction using Wavelet Transform (WT) and Gray-Level Co-occurrence Matrix (GLCM), followed by feature fusion to enhance classification. A total of 928 ECG images from four categories—Myocardial Infarction (MI), Abnormal Heartbeat (ABH), History of MI (HMI), and Normal—were analyzed. The extracted features were classified using XGBoost, Random Forest, Support Vector Machine, K-Nearest Neighbors, Decision Tree, and Logistic Regression. Results showed that XGBoost achieved the highest accuracy (93.55%), followed by Random Forest (93.01%), outperforming conventional methods. The findings suggest that feature fusion enhances classification and offers an interpretable, computationally efficient alternative to deep learning. This study contributes to automated cardiac diagnostics by providing a robust framework suitable for clinical applications and wearable ECG systems.

Keywords: *Cardiac detection, ECG classification, GLCM, Machine learning, Wavelet transform*

EKG Görüntülerinde Dalgacık Dönüşümü ve GLCM Özellik Füzyonu Kullanarak Farklı Kardiyak Durumların Makine Öğrenmesi ile Tespiti

Öz. Kardiyovasküler hastalıklar (KVH), küresel ölümlerin %32'sinden sorumlu olup, en yaygın ölüm nedenidir. Elektrokardiyografi (EKG), kardiyak anormalliklerin tespitinde yaygın kullanılan, düşük maliyetli ve non-invaziv bir yöntemdir. Ancak gürültü, bireysel fizyolojik farklılıklar ve uzman değerlendirme gereksinimi EKG yorumlamada zorluk yaratmaktadır. Bu çalışmada, EKG görüntüleriyle farklı kardiyak durumları otomatik sınıflandıran yeni bir makine öğrenmesi yöntemi önerilmektedir. Yöntem kapsamında Dalgacık Dönüşümü (WT) ve Gri-Seviye Eş-Oluşum Matrisi (GLCM) ile özellik çıkarımı yapılmış, özellik füzyonu gerçekleştirilmiştir. 928 EKG görüntüsü dört kategori Miyokard Enfarktüsü – (MI), Anormal Kalp Atışı – (ABH), Geçmiş MI – (HMI), Normal içinde analiz edilmiştir. Çıkarılan özellikler XGBoost, Random Forest, Destek Vektör Makineleri, K-En Yakın Komşu, Karar Ağacı ve Lojistik Regresyon ile sınıflandırılmıştır. Sonuçlar, XGBoost'un %93.55 doğrulukla en iyi performansı sergilediğini, onu %93.01 ile Random Forest modelinin takip ettiğini göstermiştir. Bulgular, önerilen özellik füzyonunun sınıflandırma başarısını artırdığını ve derin öğrenmeye kıyasla daha yorumlanabilir, hesaplama açısından verimli bir alternatif sunduğunu göstermektedir. Çalışma, otomatik kardiyak tanı sistemlerine katkıda bulunarak klinik uygulamalara ve taşınabilir EKG cihazlarına entegre edilebilir bir makine öğrenmesi çerçevesi sunmaktadır.

Anahtar kelimeler: *Kardiyak tespit, EKG sınıflandırma, GLCM, Makine öğrenmesi, Dalgacık dönüşümü*

* Corresponding author
E-mail address: esrasivari@karatekin.edu.tr (E. Sivari Resul)

1. Introduction

Cardiovascular diseases (CVDs), which represent 32% of all global deaths, are recognized as the leading cause of mortality worldwide [1]. Early diagnosis of these conditions is regarded as crucial for minimizing the risk of death and facilitating pharmacological management. Electrocardiograms (ECGs) are commonly employed as the most widespread, cost-effective, and non-invasive diagnostic method for cardiovascular diseases. However, several disadvantages are observed in ECG-based diagnoses. The ECG signal is susceptible to contamination by various unwanted artifacts—such as baseline wandering, power-line interference, and muscle noise—and is subject to variations among individuals due to different physiological states (e.g., during sleep or exercise) and demographic factors [2]. Furthermore, ensuring consistency and accuracy in ECG interpretations can be challenging, even for experts, under these constraints.

An increase in ECG recordings has been observed due to the growing prevalence of cardiovascular diseases and the emergence of wearable health technologies [3]. This expansion in data has generated a need for efficient and timely analysis, leading to a higher demand for specialists. More than three-quarters of deaths associated with cardiovascular diseases occur in low- and middle-income countries, indicating a shortage of specialized physicians worldwide, particularly in these regions [1]. An automated ECG analysis system is regarded as a pioneering solution for addressing the current lack of specialized healthcare professionals and for ensuring consistency and accuracy in ECG interpretations.

Various benefits are provided by computer-aided diagnosis applications designed to assist physicians in diagnosing diseases, including enhanced diagnostic accuracy and efficiency, reduced errors and variability among physicians, and lower healthcare costs through the prevention of unnecessary tests and procedures. The development of intelligent computer-aided diagnosis applications capable of automated analysis has been enabled by recent advances in artificial intelligence, particularly in deep learning and machine learning methods. In the literature, several artificial intelligence approaches that prioritize deep learning have been proposed for the development of computer-aided diagnosis systems intended for the automatic detection of cardiovascular diseases [4].

A deep learning-based model was employed by Oke et al. [5], to distinguish among six categories of heart disease, yielding an accuracy of 95%. An embedded system approach relying on deep learning models was proposed by Mhamdi et al. [6], to analyze ECG images and automatically classify cardiac arrhythmias, achieving an average accuracy above 98%. A lightweight CNN combined with an attention module was introduced by Sadad et al. [7], to classify real-time ECG images in an IoT environment, resulting in a 98.39% accuracy rate.

A classification approach based on the vectorization of ECG images was developed by Ashtaiwi et al. [8], with the aim of enhancing heart disease diagnosis. In that study, ECG images were subjected to feature extraction and vectorization processes, and the resulting vectors were then transferred to a machine learning-based classification model. The proposed method attained an accuracy of approximately 90% on the test data.

In research conducted by Sattar et al. [9], advanced deep learning models—such as CNNs and RNNs featuring various layers and architectures—were utilized to automatically process ECG waveforms, and their classification accuracies were compared. An accuracy level of nearly 98% was reported on the test data. A 2D CNN-based approach was employed by Aversano et al. [10], for the early detection of heart diseases through ECG images, reaching an accuracy of around 97-98%. In another study by the same authors [11], ECG signals were converted into image format and trained using a 2D CNN model, allowing the automatic extraction of discriminative features among different cardiac disorders. According to the reported findings, the CNN-2D-based model was shown to be effective and reliable for characterizing heart diseases, even when single-lead ECG data were employed.

An ensemble compression technique was developed by Mohanty et al. [12], with the aim of more efficiently storing and transmitting cardiac data in smart healthcare systems by combining various compression algorithms or machine learning methods. The proposed technique was reported to preserve decompressed data with an average accuracy exceeding 95%, indicating that sufficient quality for diagnosis and analysis can be maintained while reducing data storage and transmission costs. In a study by Venkataiah et al. [13], classical machine learning and deep learning models were compared for the classification of ECG images. A CNN-based approach was found to achieve the highest accuracy, generally outperforming classical machine learning methods (e.g., SVM, KNN) with a rate of approximately 98%. Beyond the diagnosis of heart diseases, deep learning models have also been frequently proposed for identifying COVID-19 based on ECG images [14–16].

Despite the frequent proposal of deep learning-based methods for ECG image analysis in the literature, several disadvantages of deep learning require attention. In particular, deep learning—especially models such as CNNs—often demands very large datasets. By contrast, conventional pattern recognition and classification methods can achieve reasonable performance with limited data, a critical consideration in medical fields where obtaining sufficient labeled data poses significant challenges. Furthermore, the internal filters and neuron learning mechanisms in deep learning are typically treated as a "black box," complicating the explanation of how specific pixel or gradient information drives a model's decisions. Features extracted via conventional pattern recognition methods tend to be more interpretable for medical experts. Moreover, deep learning architectures such as CNNs or RNNs involve a large number of parameters and may require substantial GPU power and memory capacity, whereas conventional methods can usually run on CPUs and often exhibit advantages in embedded systems or resource-constrained environments (e.g., IoT devices).

In the present study, an automatic and high-performance diagnosis of cardiovascular diseases was conducted using conventional pattern recognition and machine learning methods on ECG images. A review of the existing literature

indicates that no prior approach combining these methods for diagnosing cardiovascular diseases on the dataset employed in this study has been proposed thus far. To the best of our knowledge, this is the first study that integrates WT and GLCM feature fusion with ensemble machine learning classifiers for cardiac condition detection in ECG images. The suggested method addresses the limitations of deep learning and contributes to the literature by achieving high performance even in situations where data are limited.

Another key contribution of this study is the demonstration that machine learning models, particularly XGBoost and Random Forest, can achieve high classification accuracy (93.55%) without requiring large-scale datasets. Compared to deep learning-based models, the proposed method offers a more computationally efficient and interpretable alternative, making it suitable for embedded systems and clinical applications where computational resources are limited.

Additionally, our approach provides a robust, interpretable, and cost-effective solution for real-time ECG analysis, with potential applications in wearable health devices, telemedicine platforms, and automated clinical decision support systems. Given the increasing prevalence of cardiovascular diseases and the demand for automated diagnostic solutions, our work presents an innovative framework that bridges the gap between traditional handcrafted feature-based approaches and modern AI-driven methodologies.

Beyond technical accuracy, the proposed method addresses the growing need for interpretable and resource-efficient AI solutions in healthcare. Its transparent feature extraction process allows clinicians to better understand and trust the model's decisions, which is essential for real-world adoption. Moreover, the lightweight structure supports deployment on low-power devices, making it ideal for remote monitoring, mobile health systems, and underserved clinical settings. Thus, this study not only contributes a novel methodological approach but also offers a practical and scalable solution that bridges the gap between academic innovation and clinical applicability.

2. Materials and Methods

An overview of the proposed method is presented in Figure 1. ECG images were initially provided as input and were then subjected to a series of data preprocessing steps that included resizing, normalization, and filtering. Wavelet Transform (WT) and the Gray-Level Co-occurrence Matrix (GLCM) were subsequently employed to extract and fuse the most informative features. Afterward, the data was partitioned into training and test sets for classification. During the training phase, machine learning models were designed, cross-validated, and refined through hyperparameter tuning. The test set was subsequently used to evaluate the performance of these models, and finally, various cardiac conditions were predicted as the output of the classification process.

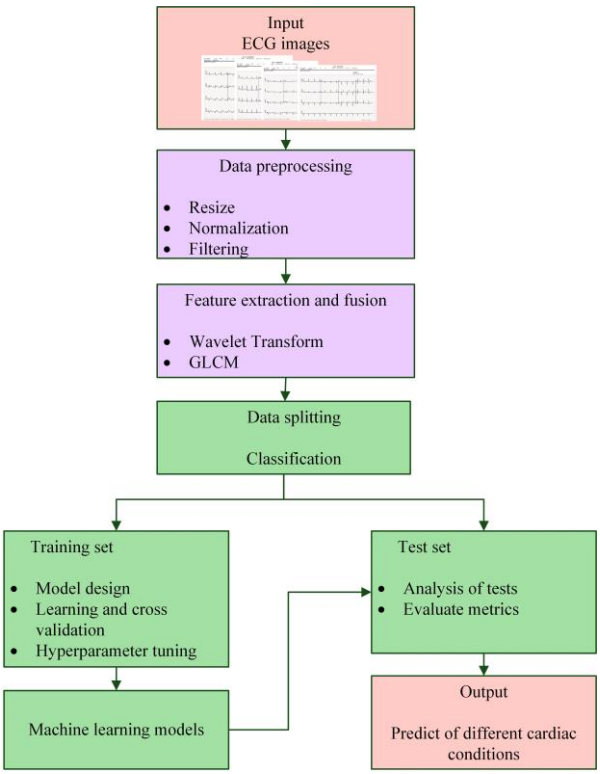

Figure 1. *Overview of the proposed method*

2.1. Dataset

The ECG Images dataset of Cardiac Patients, published in 2021 by Khan et al. [17], contained four different cardiac conditions:

- MI (Myocardial Infarction Patients) with 239 images
- ABH (Abnormal Heartbeat) with 233 images
- HMI (History of MI) with 172 images
- Normal (Normal Person) with 284 images

A total of 928 ECG images were included in the dataset, which were split into training (742 images) and test (186 images) sets using the Stratified Shuffle Split method with an %80-20 ratio. Validation data were then partitioned in real time from the training set by applying five-fold cross-validation. Samples from the dataset were illustrated in Figure 2.

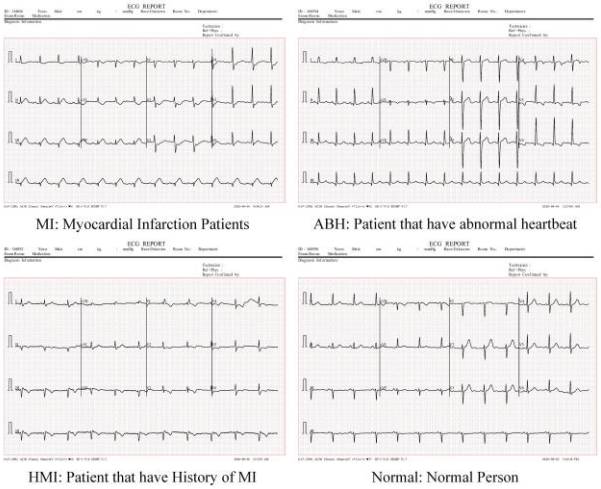

Figure 2. *Samples from the dataset*

2.2. Data Preprocessing

All images in the dataset were resized to 224 × 224 using the Bilinear interpolation algorithm and were utilized in RGB (red, green, blue) mode. The pixel values were then rescaled from 0–255 to 0–1 for normalization.

A wavelet-based preprocessing method was applied to reduce noise components in ECG images. The observed ECG data, denoted by $I_{obs}(i,j)$, is modeled as the sum of the underlying clean ECG image $I_{true}(i,j)$ and additive white Gaussian noise $N(i,j)$, i.e.,

$$I_{obs}(i,j) = I_{true}(i,j) + N(i,j) \quad (1)$$

By applying the two-dimensional Discrete Wavelet Transform (2D DWT), the image is decomposed into lower-frequency (LL) and higher-frequency (LH, HL, HH) subbands. Since noise is typically more pronounced in higher-frequency components, a thresholding operation is performed on those subbands. In particular, Donoho and Johnstone's [18] universal threshold $\lambda = \sigma\sqrt{2\ln(N)}$ is often adopted, where σ is an estimate of the noise standard deviation and *N* represents the total number of pixels. The thresholded wavelet coefficients are then reconstructed via the inverse wavelet transform (IDWT) along with the unaltered low-frequency subband, yielding a denoised ECG image in which critical morphological features, such as the P and T waves, are preserved.

This wavelet-based denoising process significantly enhances the clarity of ECG signals, thereby improving the reliability of both machine learning and classical methods used in subsequent clinical or research-oriented tasks. In particular, high-frequency artifacts that could distort vital waveform characteristics are attenuated, ensuring more accurate detection and measurement of ECG features.

2.3. Feature Extraction and Feature-Level Fusion

The Gray-Level Co-occurrence Matrix (GLCM) [19] is a second-order statistical method used to model the spatial relationship between pairs of pixel intensity values in an image. Let *I* be a grayscale image with intensity values in the range *{0,1,…,L-1}*. For a specified displacement vector δ (defined by distance and orientation, e.g., $\delta = (d, 0)$ for horizontal adjacency), the GLCM, denoted as $P_s(i,j)$, is constructed such that each entry represents the probability (or frequency) of observing intensity i at a reference pixel and intensity j at the neighbor pixel offset by δ. The co-occurrence matrix is given in Eq. (2) where N_s is a normalization factor.

$$P_s(i,j) = \frac{1}{N_s}\sum_{(x,y)} \begin{matrix} 1, if\ I(x,y) = i\ and\ (x+\delta_x, y+\delta_y) = j, \\ 0, otherwise, \end{matrix} \quad (2)$$

Among the various descriptors derivable from the GLCM, four fundamental measures typically employed to characterize textural patterns are contrast, correlation, energy, and homogeneity. Contrast quantifies the spatial frequency of intensity differences, essentially capturing how sharply or smoothly intensities change within the image. In Eq. (3), *L* is the number of gray levels and $P_s(i,j)$ denotes the probability of observing the pixel-intensity pair (i,j) under the specified offset δ.

$$Contrast = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2\, P_s(i,j) \quad (3)$$

Correlation, on the other hand, measures the linear dependency between intensities in neighboring pixels. In Eq. (4), μ_i and μ_j are the means, and σ_i and σ_j are the standard deviations of the row and column sums of P_s, respectively. A higher correlation value indicates a strong linear relationship among neighboring intensities.

$$Correlation = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i\sigma_j} P_s(i,j) \quad (4)$$

Energy corresponds to the sum of squared entries in the GLCM, reflecting uniformity within the image and is calculated as given in Eq. (5). Higher energy values indicate more frequent occurrence of certain density pairs and therefore less textural variation.

$$Energy = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}[P_s(i,j)]^2 \quad (5)$$

Finally, homogeneity emphasizes the distribution of elements near the diagonal of the co-occurrence matrix, providing a measure of local similarity or smoothness and is calculated as given in Eq. (6). Images exhibiting consistent or slowly varying intensities tend to yield higher homogeneity scores.

$$Homogeneity = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\frac{P_s(i,j)}{1+(i-j)^2} \quad (6)$$

Taken together, these four descriptors allow for a robust characterization of the spatial relationships in ECG images, capturing pertinent texture cues that can significantly aid in subsequent classification tasks. To complement the spatial descriptors obtained from GLCM, wavelet-based features are extracted from ECG images to encode multi-resolution and frequency-domain features. Specifically, the 2D DWT [20] is applied to each ECG image *I(x,y)*, decomposing it into subbands at multiple scales. At each decomposition level *s*, four subband coefficient matrices are generated: LL (low–low), LH (low–high), HL (high–low), and HH (high–high). Denoting the wavelet coefficients at level *s* and subband *b* by $W_{s,b}(x,y)$, a common approach is to calculate the subband-specific energies as given in Eq. (7). In Eq. (7), *(x,y)* spans all coefficients in the subband *b* at level *s*.

$$E_{s,b} = \sum_{(x,y)} |W_{s,b}(x,y)|^2 \quad (7)$$

In addition to energy, other statistical measures—such as mean, variance, or entropy of the wavelet coefficients—can also be extracted. These wavelet-based descriptors provide insight into local frequencies and transient behaviors in the ECG image, capturing edge-like features and subtle amplitude variations that may be diagnostic in distinguishing specific cardiac conditions.

A combined feature vector (*F*) is created by combining the sets of GLCM and WT features. Specifically, let $g = [g_1, g_2, \ldots]$ denote the GLCM feature vector and $w = [w_1, w_2, \ldots]$ represent the wavelet-based feature vector. The resulting combined feature vector was constructed as $F = [g, w]$ thereby integrating both the textural relationships captured by GLCM descriptors and the multi-resolution information derived from the wavelet coefficients. This comprehensive representation facilitated more robust classification of the ECG images.

2.3. Classification

In the classification phase, features were classified using machine learning techniques and tools. During the model design stage, the Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), XGBoost (XGB), and Decision Tree (DT) algorithms were used. In the modeling phase of the machine learning classifiers, the GridSearchCV approach was employed to carry out hyperparameter optimization. This method systematically evaluates all possible hyperparameter value combinations, thereby identifying the configurations that yield the highest performance metrics.

Table 1 presents the Scikit-learn [21] classification algorithms used in the application, along with the specific hyperparameter ranges explored during optimization. The selection of appropriate hyperparameters is crucial for optimizing the performance of machine learning models. In this study, various classifiers were employed, each with a set of predefined hyperparameters to fine-tune their learning behavior. LR was configured with L2 regularization (C = 1.0) to prevent overfitting, while the Newton-Conjugate Gradient (newton-cg) solver was chosen for its efficiency in handling multinomial LR problems. The maximum number of iterations (max_iter = 100) ensures convergence, and one-vs-rest (ovr) multi-class strategy was used to facilitate multi-class classification. For the SVM classifier, a linear kernel was utilized, given its suitability for high-dimensional data, while C = 0.1 was selected to balance the trade-off between margin maximization and misclassification cost.

For ensemble-based methods, XGB and RF classifiers were optimized using hyperparameters tailored to their specific learning mechanisms. In XGB, the number of boosting rounds was set to 100 (n_estimators = 100), while a learning rate of 0.1 ensured a gradual update of weights, preventing overfitting. The maximum tree depth (max_depth = 3) was constrained to control model complexity. Similarly, RF was configured with 20 decision trees (n_estimators = 20), while max_depth = 20 and min_samples_split = 5 were set to balance depth and data splitting criteria. DT utilized the Gini impurity criterion for split evaluation, with a maximum depth of 10 and a minimum sample split threshold of 2 to avoid excessive fragmentation. Lastly, KNN was set with 5 neighbors (n_neighbors = 5), using the Minkowski distance metric (p = 2), which corresponds to the Euclidean distance, ensuring an effective similarity measure for classification.

During the training phase, a five-fold cross-validation strategy was implemented for these classifiers. In each iteration, the training set was randomly partitioned into five distinct subsets, from which four were used for training and one was designated as the validation subset. Fig. 3 illustrates the learning curves of the machine learning classifiers derived during the training process. In these curves, the x-axis corresponds to the algorithms' accuracy, while the y-axis indicates the amount of training data. Within each learning curve in Fig. 3, the red line represents the training accuracy, and the green line denotes the cross-validation accuracy obtained via the five-fold cross-validation procedure.

Overall, the learning curves indicate that RF, SVM, and XGB achieve the highest validation accuracies, approaching approximately 90-95% as the number of training samples increases. LR also exhibits strong performance, converging to a validation accuracy close to 85-88%. In contrast, KNN maintains the lowest validation accuracy throughout the training process, and the DT model settles at an intermediate level of performance around 75-80%. These outcomes suggest that ensemble methods (RF and XGB) and SVM are particularly well-suited to the problem setup, whereas KNN's simpler, distance-based approach is comparatively less effective.

From an overfitting perspective, both DT and RF display near-perfect training scores accompanied by comparatively lower validation accuracies, indicating a degree of overfitting in those models (particularly in the early training stages). However, the gap between training and validation scores narrows as the sample size increases, suggesting that additional data helps mitigate overfitting. SVM, XGB and LR similarly achieve high training accuracy but ultimately generalize well, evidenced by validation scores that steadily rise and closely track the training curves at higher sample sizes.

Table 1 The values of the hyperparameters

Classifier Hyperparameter	Value
Logistic regression	
C	1.0
solver	newton-cg
max_iter	100
multi_class	ovr
Support vector machine	
C	0.1
kernel	linear
XGBoost	
n_estimators	100
learning_rate	0.1
max_depth	3
K-nearest neighbors	
n_neighbors	5
p	2
metric	Minkowski
Decision Tree	
criterion	Gini
max_depth	10
min_samples_split	2
Random Forest	
max_depth	20
min_samples_split	5
n_estimators	20

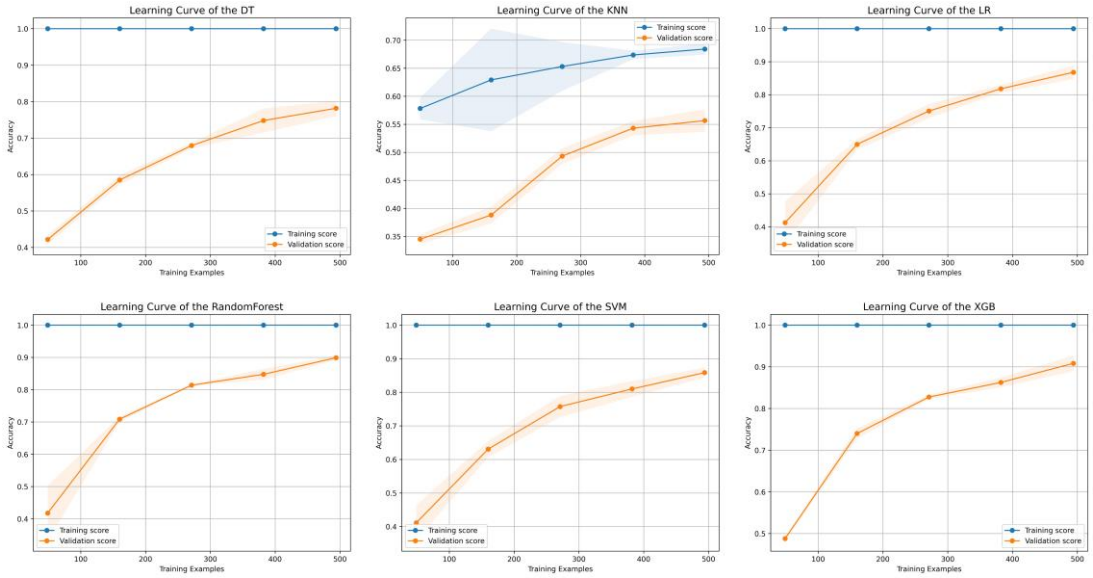

Figure 1. *Learning curves of machine learning algorithms*

3. Results and Discussion

3.1. Performance Metrics

The confusion matrix provides a tabular representation of the classification outcomes, comparing predicted labels to true labels. In binary classification, the matrix commonly reports four types of results: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). By examining these outcomes, one can diagnose specific error patterns, such as how frequently the model inaccurately identifies positive samples as negative and vice versa.

Several metrics can be derived from the confusion matrix, and their formulas are presented in Table 2. Accuracy captures the proportion of correct predictions among all predictions. Precision quantifies how many of the predicted positive cases are truly positive, while recall indicates the fraction of actual positive cases correctly identified. The F1 score is a harmonic mean of precision and recall, thus balancing both metrics in a single measure. ROC AUC (Area Under the Receiver Operating Characteristic Curve) assesses the model's ability to distinguish between classes over varying decision thresholds, and higher values suggest better overall performance. Lastly, the Jaccard similarity coefficient (JSC) measures the similarity between the predicted and actual positive sets, reflecting how much overlap exists between these two sets.

These evaluation metrics, derived from the confusion matrix, provide a multifaceted view of model performance beyond overall accuracy. While accuracy gives a general sense of correctness, it may be misleading in imbalanced datasets, making precision, recall, and F1 score crucial for a more nuanced understanding. Precision is particularly important in medical diagnostics where false positives may lead to unnecessary stress or treatment, whereas recall is vital when missing true cases—such as undiagnosed cardiovascular conditions—poses significant risk. The F1 score balances these two concerns, offering a more reliable indicator when class distributions are uneven. Additionally, ROC AUC serves as a threshold-independent metric, revealing how well the model distinguishes between classes across various sensitivity levels. Jaccard similarity further complements these measures by quantifying the overlap between predicted and actual positive cases, providing insight into the consistency of model predictions.

Table 2 Performance metrics

Metric	Formula
Accuracy	$\frac{TP + \mathrm{T}N}{TP + \mathrm{F}P + \mathrm{F}N + \mathrm{T}N}$
Precision	$\frac{TP}{TP + \mathrm{F}P}$
Recall	$\frac{TP}{TP + \mathrm{F}N}$
F_1 Score	$2\ X\ \frac{Precision\ X\ Recall}{Precision +\ Recall}$
Jaccard similarity coefficient	$\frac{TP}{TP + \mathrm{F}N + FP}$

3.2. Test Results

During the testing phase, a total of 186 images were employed, consisting of 57 Normal, 47 ABH, 34 HMI, and 48 MI. In Fig. 4, the confusion matrices of the classifiers are presented.

Overall, the confusion matrices revealed that the MI class was predicted with near-perfect accuracy by all models, indicating that minimal misclassification was observed for this specific category. In contrast, the ABH and HMI classes were misclassified to varying degrees, suggesting that these categories posed more challenges for certain algorithms. In particular, the KNN model was found to struggle considerably in distinguishing ABH, leading to a large portion of ABH instances being misclassified as other classes.

By comparison, the ensemble methods (RF and XGB) were observed to achieve near-flawless performance, as almost all samples were assigned to the correct classes. Similarly, SVM was found to yield strong overall accuracy, although

occasional confusion was noted between ABH and HMI. While the DT and LR models also performed favorably, slightly higher misclassification rates were exhibited, especially at the boundary between ABH and HMI. Taken together, these findings suggest that ensemble approaches offered the most robust classification performance, whereas KNN emerged as the least effective within this particular context.

In Table 3, the performance metrics for all six classification algorithms were presented. It was observed that XGB achieved the highest accuracy (93.55%), precision (93.70%), recall (93.55%), F1 score (93.45%), and JSC (87.25%), thereby establishing itself as the most robust method among the models considered. Although RF registered a slightly lower accuracy (93.01%), it was found to surpass all other algorithms in terms of ROC-AUC (99.48%), indicating strong discriminative power. KNN, on the other hand, exhibited the weakest outcomes, as evidenced by its notably low accuracy (61.29%) and comparatively modest values for precision, recall, and F1 score.

These discrepancies in performance were likely attributable to the inherent strengths of ensemble methods, which combine multiple decision boundaries to reduce variance and enhance generalization. Specifically, XGB was observed to benefit from gradient boosting strategies that fine-tune weak learners, while RF leveraged bootstrap aggregation to mitigate overfitting. By contrast, the distance-based approach of KNN, which relies on local similarity, appeared inadequate in capturing the nuances of the feature space without additional optimization or feature engineering. Meanwhile, the DT and LR classifiers maintained respectable accuracy levels of 81.72% and 83.87%, respectively, yet they were outperformed by ensemble methods in all other key metrics, particularly in ROC-AUC and JSC.

In conclusion, the superiority of ensemble classifiers was consistently highlighted by their elevated performance across multiple evaluation criteria. Although SVM also demonstrated commendable results, achieving an accuracy of 86.02% and a high ROC-AUC (98.14%), the methods based on boosted or bagged decision trees proved more adept at optimizing both sensitivity and specificity. Consequently, these findings suggest that, for this specific classification task, XGB and RF present the most effective solutions, whereas KNN's performance may be improved through further feature engineering or hyperparameter tuning.

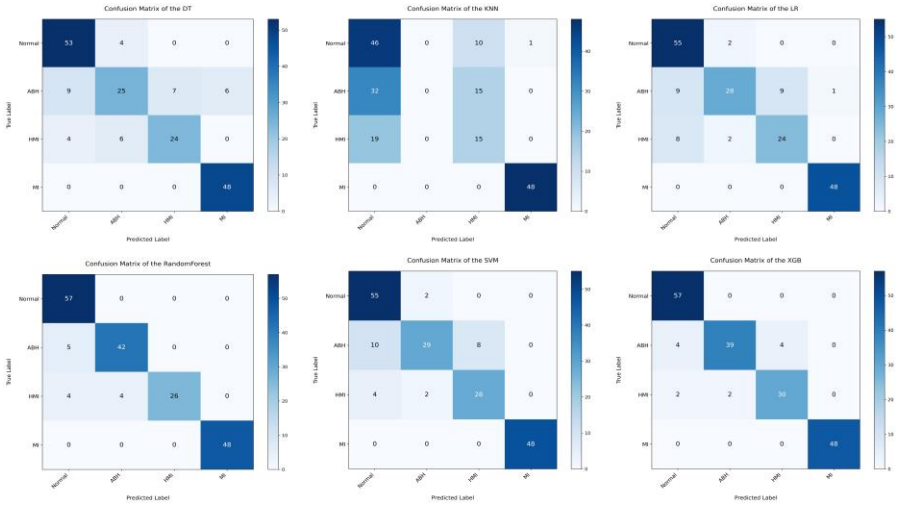

Figure 2. *Confusion matrices of machine learning algorithm*

Table 3 Test results

Classifier	ACC (%)	Precision (%)	Recall (%)	F_1 Score (%)	ROC-AUC (%)	JSC (%)
RF	93.01	93.62	93.01	92.88	99.48	86.30
SVM	86.02	86.66	86.02	85.50	98.14	75.25
KNN	61.29	73.64	61.29	53.57	79.95	43.13
DT	81.72	81.82	81.72	80.72	88.05	68.45
LR	83.87	84.35	83.87	83.24	96.10	71.27
XGB	93.55	93.70	93.55	93.45	98.93	87.25

4. Discussion

The findings of this study highlight the effectiveness of combining WT and GLCM feature fusion with machine learning algorithms for classifying different cardiac conditions based on ECG images. The results indicate that ensemble-based classifiers, particularly XGB and RF, outperform other models by achieving high classification accuracy and robust generalization. However, several aspects require further discussion regarding the limitations, advantages, disadvantages, and clinical implications of the proposed method.

Despite its promising results, the proposed method is not without limitations. First, the dataset size (928 images) remains relatively small, which may impact the generalizability of the model to broader clinical applications. While machine learning models, particularly ensemble

methods, demonstrated strong performance, deep learning-based approaches might outperform them if larger datasets were available. Additionally, the reliance on handcrafted features (WT and GLCM), although advantageous for interpretability, may not capture all latent patterns present in ECG images as efficiently as deep learning-based feature extraction methods.

Another limitation pertains to class imbalance issues, particularly in the HMI and ABH categories, which exhibited higher misclassification rates. While stratified cross-validation and ensemble techniques mitigated this issue to some extent, further optimization, such as using synthetic data augmentation techniques, could improve classification performance. Furthermore, the study only focuses on ECG images, whereas integrating raw ECG signals or multimodal physiological data (e.g., heart rate variability, patient demographics) could enhance diagnostic accuracy.

One of the most notable advantages of the proposed methodology is its efficacy in scenarios with limited data availability. Unlike deep learning methods, which typically require thousands of labeled samples, the machine learning approach combined with feature engineering provides high performance without excessive computational demands. This makes the method highly suitable for resource-constrained environments, such as embedded systems in portable and wearable ECG monitoring devices.

Moreover, wavelet-based denoising significantly enhances the clarity of ECG images, which is a crucial advantage in medical image analysis. The feature-level fusion of GLCM and WT captures both spatial and frequency-based characteristics, leading to more robust feature representations. Additionally, the use of interpretable handcrafted features improves model explainability, a critical factor in medical applications where decisions must be justified to clinicians.

A comprehensive comparison of the proposed method with relevant studies in the literature is presented in Table 4. While most existing studies adopt deep learning-based approaches—particularly CNNs—for ECG image classification and report high accuracy levels (e.g., Sadad et al., [7]: 98.39%, Mhamdi et al., [6]: 95%, Aversano et al., [11]: 97%), these models often require large-scale datasets, GPU-based hardware, and entail limited explainability due to their black-box nature. In contrast, our study employs a handcrafted feature fusion method based on WT and GLCM, followed by conventional machine learning classifiers. Despite using the same dataset size (928 ECG images), our approach achieved a high classification accuracy of 93.55% with the XGB model, while maintaining low computational complexity and high interpretability—features that are especially beneficial for real-world clinical applications and deployment in embedded or mobile systems.

Several studies have emphasized model deployment on edge devices; for instance, Mhamdi et al., [6] utilized MobileNetV2 and VGG16 models optimized for Raspberry Pi, achieving 92–94% accuracy on-device. While impressive, such deep models still necessitate model compression and fine-tuning to function on limited hardware. Similarly, Aversano et al., [11] developed two CNN-based architectures with accuracy 97% in multi-lead ECG analysis. However, these models demand significant computational resources and training time, and their internal operations are less interpretable to clinicians. In contrast, our method provides transparency through handcrafted features and operates effectively on CPUs without specialized hardware. Moreover, unlike deep learning models that often overfit small datasets or require augmentation, our feature fusion strategy performs competitively without such dependencies.

Although our method may not outperform the most advanced deep models in raw accuracy, it offers significant advantages in terms of deployment feasibility, explainability, and efficiency—critical factors in clinical decision support systems, telemedicine, and wearable health monitoring technologies. Furthermore, unlike black-box CNNs, our approach produces interpretable features that align with clinical expectations. While manual feature engineering may limit scalability to some extent, future research could explore hybrid models combining shallow CNN embeddings or transfer learning with handcrafted features to enhance performance without compromising transparency. Overall, the proposed method contributes a valuable alternative to the literature by offering a reliable, interpretable, and computationally lightweight solution for ECG image classification, particularly suited for settings with limited data and hardware resources.

The proposed approach has substantial potential for real-world medical applications. With the increasing prevalence of wearable health technologies, automated ECG classification models could be integrated into portable monitoring devices, telemedicine applications, and clinical decision support systems. Given that more than three-quarters of cardiovascular disease-related deaths occur in low- and middle-income countries, where access to specialized cardiologists is limited, deploying efficient, lightweight, and explainable artificial intelligence models could significantly improve early diagnosis and patient management.

Furthermore, early and accurate identification of abnormal ECG patterns can facilitate timely medical intervention, reducing the risk of severe cardiac events. The integration of machine learning-based ECG analysis with electronic health records (EHR) could enhance personalized treatment strategies, enabling clinicians to tailor interventions based on automated insights derived from ECG patterns.

Table 4 Overview of studies classifying ECG images

Author, Year, Reference	Diagnosis	Data Size	Method	Result (ACC %)
Oke et al., 2025, [5]	Normal, MI, ABH, Atrial fibrillation, Ischemic heart disease, Sinus bradycardia	2848	VGG16 + SVM + Random Forest	95.00

Mhamdi et al., 2022, [6]	Normal, MI, ABH, HMI	928	Fine-Tuning MobileNetV2	95.00
Aversano et al., 2024, [11]	Normal, MI, ABH, HMI	928	CNN-2D	97.00
Sadad et al., 2023, [7]	Normal, MI, ABH, HMI	928	Lightweight CNN + Attention Module	98.39
Ashtaiwi et al., 2024, [8]	Normal, MI, ABH, HMI	928	Image Vectorization + ANN	89.58
This Study	Normal, MI, ABH, HMI	928	WT + GLCM + XGB	93.55

5. Conclusions and Future Directions

This study aimed to develop an effective and interpretable machine learning-based approach for classifying different cardiac conditions using ECG images, leveraging WT and GLCM feature fusion. The experimental results demonstrated that ensemble classifiers, particularly XGB and RF, achieved the highest classification performance, with XGB attaining an accuracy of 93.55%, highlighting the effectiveness of the proposed methodology in ECG-based diagnosis.

This study demonstrates that WT and GLCM-based feature fusion, combined with ensemble machine learning classifiers, can achieve high performance in ECG-based cardiac condition classification, even with limited data availability. The proposed approach offers a computationally efficient, interpretable, and clinically relevant alternative to deep learning-based methods. While certain limitations remain, this work contributes to the growing body of research on AI-driven cardiac diagnostics and highlights the potential of machine learning in real-world medical applications.

Future research should focus on expanding the dataset size and diversity to enhance model generalizability. Additionally, combining deep learning feature extraction with conventional machine learning classifiers could offer the best of both worlds-leveraging deep feature representations while maintaining model interpretability. Exploring ensemble models that incorporate both handcrafted and deep features could lead to further performance improvements. Finally, validating the model on real-world clinical ECG data and testing its integration with portable ECG devices will be crucial steps toward practical deployment.

Author Contribution

Data curation -Kadircan Karaca (KK); Formal analysis – Esra Sivari Resul (ESR); Investigation - KK; Experimental Performance - KK; Data Collection - KK; Literature review - ESR; Writing - ESR; Review and editing – Mustafa Karhan (MK); Conceptualization - ESR and MK; Methodology - KK and MK; Validation - ESR and MK; Visualization - ESR and KK; Supervision - ESR and MK. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

The authors declared no conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Timmis, A., Group on behalf of the AW, Vardas, P., et al. (2022). European Society of Cardiology: cardiovascular disease statistics 2021. Eur Heart J., 43(8), 716–799. https://doi.org/10.1093/EURHEARTJ/EHAB892

[2] Zanchi, B., Monachino, G., Fiorillo, L., et al. (2025). Synthetic ECG signals generation: A scoping review. Comput Biol Med, 184, 109453. https://doi.org/10.1016/J.COMPBIOMED.2024.109453

[3] Kaplan Berkaya, S., Uysal, A.K., Sora Gunal, E., et al. (2018). A survey on ECG analysis. Biomed Signal Process Control, 43, 216–235. https://doi.org/10.1016/J.BSPC.2018.03.003

[4] Lopez-Jimenez, F., Attia, Z., Arruda-Olson, A.M., et al. (2020). Artificial Intelligence in Cardiology: Present and Future. Mayo Clin Proc, 95(5), 1015–1039. https://doi.org/10.1016/J.MAYOCP.2020.01.038

[5] Oke, O.A., Cavus, N. (2025). Electrocardiogram image classification for six classes of heart diseases. Iran Journal of Computer Science, 2025, 1–21. https://doi.org/10.1007/S42044-025-00227-X

[6] Mhamdi, L., Dammak, O., Cottin, F., Dhaou, I. (2022). Artificial Intelligence for Cardiac Diseases Diagnosis and Prediction Using ECG Images on Embedded Systems. Biomedicines, 10, 2013. https://doi.org/10.3390/BIOMEDICINES10082013

[7] Sadad, T., Safran, M., Khan, I., et al. (2023). Efficient Classification of ECG Images Using a Lightweight CNN with Attention Module and IoT. Sensors, 23, 7697. https://doi.org/10.3390/S23187697

[8] Ashtaiwi, A.A., Khalifa, T., Alirr, O. (2024). Enhancing heart disease diagnosis through ECG image vectorization-based classification. Heliyon, 10(18), e37574. https://doi.org/10.1016/j.heliyon.2024.e37574

[9] Sattar, S., Mumtaz, R., Qadir, M., et al. (2024). Cardiac Arrhythmia Classification Using Advanced Deep Learning Techniques on Digitized ECG Datasets. Sensors, 24, 2484. https://doi.org/10.3390/S24082484

[10] Aversano, L., Bernardi, M.L., Cimitile, M., et al. (2023). Early Diagnosis of Cardiac Diseases using ECG Images and CNN-2D. Procedia Comput Sci, 225, 2866–2875. https://doi.org/10.1016/J.PROCS.2023.10.279

[11] Aversano, L., Bernardi, M.L., Cimitile, M., et al. (2024). Characterization of Heart Diseases per Single Lead Using ECG Images and CNN-2D. Sensors, 24(11), 3485. https://doi.org/10.3390/S24113485

[12] Mohanty, M.N., Baliarsingh, S., Panda, P.K. (2025). An Ensemble Technique for Cardiac Data Compression in Smart Healthcare System. SN Comput Sci, 6(1), 78. https://doi.org/10.1007/s42979-024-03605-7

[13] Venkataiah, Dr.V., Mamatha, B. (2024). Detection of Cardiovascular Diseases in ECG Images Using Machine

Learning and Deep Learning Methods. International Journal of Information Technology and Computer Engineering, 12(3), 478–494.

[14] Attallah, O. (2022). An Intelligent ECG-Based Tool for Diagnosing COVID-19 via Ensemble Deep Learning Techniques. Biosensors (Basel), 12(5), 299. https://doi.org/10.3390/BIOS12050299/S1

[15] Rahman, T., Akinbi, A., Chowdhury, M.E.H., et al. (2022). COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network. Health Inf Sci Syst, 10(1), 1–16. https://doi.org/10.1007/S13755-021-00169-1/FIGURES/12

[16] Khan, A.H., Hussain, M., Malik, M.K. (2021). ECG Images dataset of Cardiac and COVID-19 Patients. Data Brief, 34, 106762. https://doi.org/10.1016/J.DIB.2021.106762

[17] Khan, A.H., Hussain, M. (2021). ECG Images dataset of Cardiac Patients. 2. https://doi.org/10.17632/GWBZ3FSGP8.2

[18] Donoho, D.L., Johnstone, J.M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81(3), 425–455. https://doi.org/10.1093/BIOMET/81.3.425

[19] Lee, H., Yoon, T., Yeo, C., et al. (2021). Cardiac Arrhythmia Classification Based on One-Dimensional Morphological Features. Applied Sciences, 11(20), 9460. https://doi.org/10.3390/APP11209460

[20] Mallat, S. (2008). A Wavelet Tour of Signal Processing.

[21] Pedregosa, F., Weiss, R., Brucher, M., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[22]