

Explainable Lightweight Swin Transformer for ECG Classification

Kazi Nusrat Jahan Suchi^{1*}, Nusrat N. Suzana¹, Md. Maynul Islam¹

Abstract

Cardiovascular diseases are a leading cause of death worldwide and require accurate, efficient, and interpretable diagnostic tools. Although electrocardiography (ECG) is widely used for non-invasive cardiac assessment, manual interpretation can be time-consuming and subjective. In addition, many deep learning approaches provide limited interpretability and may not adequately capture long-range dependencies in ECG signals. This study proposes an explainable ECG image classification framework based on a customized Tiny Swin Transformer. Structured preprocessing enhances waveform visibility, while shifted-window self-attention captures local morphology and broader rhythm patterns. LIME provides superpixel-based attribution maps that identify image regions contributing to the model's predictions. The framework was evaluated on a public 12-lead ECG image dataset containing four classes: myocardial infarction (MI), previous myocardial infarction (PMI), arrhythmia, and normal. The model achieved 98.22% accuracy with 28.29 million parameters and strong precision, recall, and F1-score. The LIME maps highlighted regions associated with ST-segment variation and rhythm irregularity. These results indicate that the framework offers a promising balance of predictive performance, computational efficiency, and interpretability; however, external clinical validation is required before deployment.

Keywords: Swin Transformer, LIME, Myocardial Infarction, Arrhythmia Detection

1. Introduction

Cardiovascular diseases (CVDs) account for approximately 32% of global deaths (World Health Organization [WHO], 2025). Their prevalence and associated healthcare costs highlight the need for timely, accurate, and affordable diagnostic tools. Early identification of abnormal cardiac function can support appropriate

*Corresponding Author email: jahansuchinjs@gmail.com

¹Department of Information and Communication Technology, Bangladesh University of Professionals, Dhaka, Bangladesh

DOI: <https://doi.org/10.64494/JFST/v4i1/ks/2026/01/1-22>

clinical intervention and improve patient outcomes. Electrocardiography (ECG) is widely used to detect arrhythmia, myocardial infarction, ischemia, and conduction abnormalities (Rajpurkar et al., 2017).

Deep learning has advanced ECG analysis by learning discriminative features directly from data and reducing dependence on handcrafted signal-processing methods. Rajpurkar et al. (2017) showed that a convolutional neural network (CNN) could identify arrhythmias from single-lead ECG recordings with performance comparable to cardiologists on their study dataset. However, CNNs mainly aggregate information through local receptive fields, which may limit their ability to model long-range rhythm patterns and inter-lead dependencies. Transformers address this limitation through self-attention, but many deep models still provide limited clinical interpretability. Explainable artificial intelligence methods, including LIME, SHAP, and gradient-based visualization, can improve transparency by highlighting features that influence predictions.

Transformer models perform well in capturing long-range dependencies, but their clinical use in ECG analysis is limited by high computational cost. It is also unclear whether full-scale vision transformers are necessary for ECG classification. Lightweight transformer alternatives that preserve both accuracy and interpretability are still underexplored. This study addresses that gap by showing that a tailored Tiny Swin Transformer can model ECG images effectively with reduced architectural complexity.

This paper presents an explainable ECG image classification system that integrates structured preprocessing, a lightweight Tiny Swin Transformer, and LIME-based explainability. The model extracts both local and global ECG features, while preprocessing improves input quality under varying signal conditions. LIME highlights important image regions behind each prediction, supporting clinically meaningful interpretation. The key contributions are as follows:

- a. A structured ECG image preprocessing pipeline is proposed to enhance ECG signal quality, reduce acquisition artifacts, and improve robustness across four cardiac classes, achieving a macro-averaged F1-score of 0.9804.
- b. A computationally efficient Tiny Swin Transformer is developed, achieving 98.22% ECG classification accuracy with reduced model parameters.
- c. A LIME-based explainability module is integrated to provide region-level visual explanations of ECG areas that influence individual predictions.

2. Literature Review

ECG analysis is a major area of deep learning-based cardiovascular diagnosis because ECG signals are pattern-rich and increasingly available in annotated datasets. Early deep learning studies mainly used CNNs to learn discriminative morphological features from raw ECG signals or ECG images.

Rajpurkar et al. (2017) showed that a CNN-based method could achieve cardiologist-comparable performance for arrhythmia detection on a large single-lead ECG dataset. Alkegany and Ali (2024) also reported strong performance for image-based ECG classification using a vision-transformer approach. Hybrid architectures such as CAT-Net combine convolution, attention, and transformer components to capture local and global ECG characteristics (Islam et al., 2024).

Despite their strong performance, CNNs primarily build representations through local receptive fields. This can make global rhythm regularity, long-term waveform evolution, and dependencies across leads more difficult to capture without increasing network depth and computational cost. These limitations motivate the use of attention-based and transformer architectures for ECG representation.

Transformers were introduced by Vaswani et al. (2017) to model long-range dependencies through self-attention. The Vision Transformer extended this concept to image classification by processing images as sequences of patches (Dosovitskiy et al., 2021). The Swin Transformer later improved computational efficiency through shifted-window attention and hierarchical feature representations (Liu et al., 2021). ECG-specific transformer methods include ECGformer (Akan et al., 2024), ECGTransForm (El-Ghaish & Eldele, 2024), and MSW-Transformer (Cheng et al., 2023).

Recent ECG transformer studies use several strategies to capture multi-scale and time-frequency information. MTECG combines continuous wavelet transform features with a multi-branch transformer (Qiu et al., 2024), while Chen et al. (2024) proposed a Swin Transformer using time-frequency characteristics. Other approaches include MSW-Transformer (Cheng et al., 2023), Cardio Pattern Former (Uğraş et al., 2025), and ECGMamba (Qiang et al., 2024), which use multi-scale windows, pattern-guided attention, or state-space modeling.

Explainability is important for the responsible clinical use of automated ECG systems. LIME explains individual predictions by fitting a locally interpretable surrogate model (Ribeiro et al., 2016), whereas SHAP assigns feature contributions using a unified additive framework (Lundberg & Lee, 2017). Grad-CAM provides coarse localization maps for CNN-based image models (Selvaraju et al., 2017). Explainable ECG studies have also used local attribution methods to

identify waveform regions associated with cardiac disorders (Anand et al., 2022). More recent transformer-based models, including CardioPatternFormer, have emphasized interpretable attention mechanisms (Uğraş et al., 2025).

Recent studies further demonstrate the value of machine learning and explainability in clinical decision support. Nova et al. (2025) optimized machine-learning models for CVD prediction, while Shafi and Mirza (2025) combined pretrained models with LIME-based explainable artificial intelligence for interpretable brain-tumour classification. These findings reinforce the need to balance predictive performance, computational efficiency, and clinical interpretability in automated diagnosis.

However, many transformer-based ECG studies prioritize predictive performance over explanation and deployment efficiency. Full-scale ViT and Swin architectures can require substantial memory and computation, which may limit their use in point-of-care and resource-constrained settings (Alkegany & Ali, 2024; Chen et al., 2024). Therefore, a lightweight and explainable transformer framework for ECG classification remains an important research direction.

2.1 Research Gap and Research Questions

Existing ECG transformer models often achieve strong performance but require high computational resources and provide limited interpretability. Moreover, the effect of reducing model size on different ECG classes remains unclear. This study addresses these gaps by evaluating a Tiny Swin Transformer with LIME-based explanations.

The study is guided by the following research questions:

- a. Can a lightweight Tiny Swin Transformer achieve cardiologist-level ECG classification accuracy with lower complexity than full-scale vision transformers?
- b. Can LIME provide clinically meaningful attribution maps for a transformer-based ECG classification framework?
- c. Does model-size reduction affect performance across MI, PMI, Arrhythmia, and Normal ECG classes?

3. Transformer-Based Explainable ECG Classification

The proposed framework combines a lightweight Tiny Swin Transformer with LIME to build a high-performance and clinically transparent multi-class ECG image classifier. The design is motivated by three factors: the ability of Swin Transformers to learn fine waveform patterns and long-range rhythm structures, the efficiency of shifted-window self-attention compared with global self-

attention, and the need for explainable clinical decisions. The model classifies standardized ECG images into MI, PMI, Arrhythmia, and Normal.

The Tiny Swin Transformer was selected to reduce computational cost while preserving the hierarchical shifted-window attention mechanism. Standard Swin-Base and Swin-Large configurations contain approximately 88 million and 197 million parameters, respectively, whereas Swin-Tiny contains approximately 28 million parameters (Liu et al., 2021). This smaller parameter budget may be suitable for ECG images, which contain structured waveform patterns rather than the broad visual diversity found in natural-image datasets.

The preprocessing pipeline contains four steps that standardize ECG images and improve input quality for the transformer backbone.

Background Removal: Threshold-based processing suppresses the white ECG paper grid and other non-informative background regions, isolating the waveform signal from acquisition noise.

Region of Interest Cropping: ROI-based cropping removes redundant borders, demographic headers, and footer annotations so that only the diagnostically relevant waveform area remains.

Resizing: Cropped images are resized to 224×224 pixels using bilinear interpolation, preserving waveform continuity while matching the Tiny Swin Transformer input size.

Normalization: Images are converted to tensors and normalized using ImageNet mean (0.485, 0.456, 0.406) and standard deviation (0.229, 0.224, 0.225), supporting stable optimization and transfer learning.

After preprocessing, each ECG image becomes a clean tensor compatible with the Swin Transformer, with reduced noise and preserved structural waveform information.

The customized Tiny Swin Transformer backbone, shown in Figure 1, performs hierarchical multi-scale feature extraction. Each 224×224 ECG image is divided into non-overlapping 4×4 patches, linearly embedded, and processed through transformer stages that progressively reduce spatial resolution while increasing feature dimensionality. Shifted-window self-attention captures local waveform morphology and broader rhythm dependencies efficiently (Liu et al., 2021). A task-specific fully connected layer produces four logits corresponding to MI,

PMI, arrhythmia, and normal classes. ImageNet-pretrained weights are used for transfer learning. The model is optimized with AdamW using a learning rate of 1×10^{-4} and weight decay of 0.05, with cosine annealing and cross-entropy loss. Data augmentation should be limited to transformations that do not alter clinically meaningful ECG morphology.

Table 1: Model Architecture and Training Configuration

Model Component	Configuration
Backbone Network	Tiny Swin Transformer
Input Resolution	224×224 RGB ECG Images
Number of Classes	4 (MI, PMI, Arrhythmia, Normal)
Learning Rate	1×10^{-4}
Optimizer	AdamW
Batch Size	32
Scheduler	Cosine Annealing
Loss Function	Cross-entropy loss
Epochs	30

The number of training epochs is set to 30 based on empirical observation of the training and validation loss curves. The model consistently converged within 25-30 epochs, after which no significant improvement in validation accuracy was observed. Training beyond 30 epochs showed signs of overfitting, with validation loss beginning to increase while training accuracy continued to rise. Table 2 presents the model's validation accuracy at different epoch values, confirming that 30 epochs provides the optimal trade-off between convergence and generalization.

Table 2: Validation Accuracy Across Training Epochs

Epochs	Validation Accuracy
10	72.50
20	89.16
30	98.22
40	99.15
50	99.98

LIME was integrated to improve the interpretability of the ECG classification framework (Ribeiro et al., 2016). For each ECG image, LIME divides the input into superpixels, perturbs selected regions, and observes the resulting change in

predicted class probability. It then fits a sparse local surrogate model to estimate the contribution of each region. The resulting region-level attribution maps are overlaid on the ECG image to show which waveform areas support or oppose the prediction. These maps can be inspected for alignment with potentially relevant patterns, such as ST-segment variation, rhythm irregularity, and waveform regularity; however, clinical meaningfulness should be confirmed through expert evaluation.

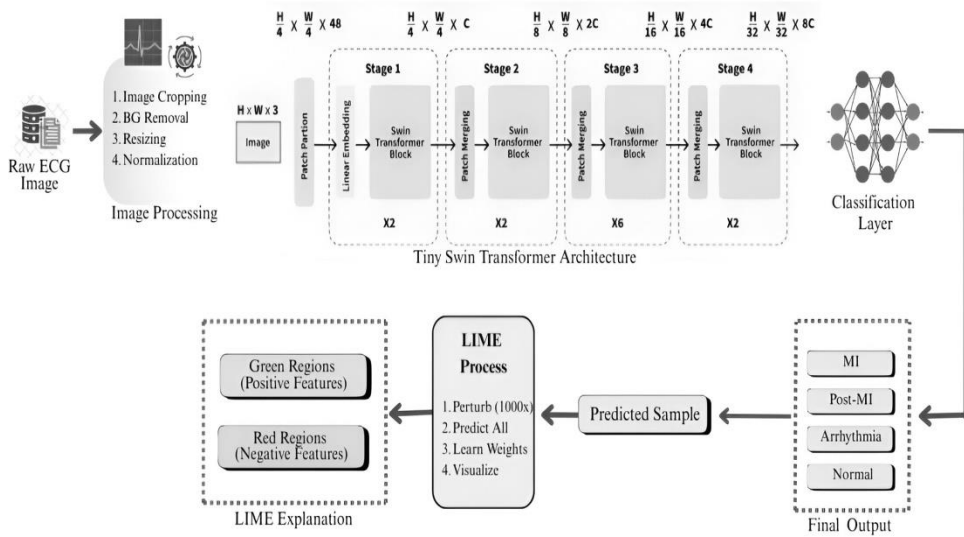


Figure 1: Framework of the proposed approach

4. Results and Discussion

The proposed framework was implemented in Python and evaluated using Google Colab. For reproducibility, the final manuscript should report the exact Colab accelerator, software-library versions, training duration, and inference-measurement procedure. The experiments assess classification performance, computational characteristics, and explanation quality across the four ECG categories.

4.1 ECG Dataset

This study uses the publicly available ECG Images Dataset of Cardiac Patients by Khan and Hussain (2021). The dataset contains 928 high-resolution 12-lead ECG report images collected at Ch. Pervaiz Elahi Institute of Cardiology, Pakistan. The images include lead-wise waveforms and rhythm patterns and are grouped into four diagnostic categories: normal, MI, PMI, and arrhythmia. The

class distribution is presented in Table 3. The final manuscript should also report the exact training, validation, and test split, the random seed, and whether patient-level separation was enforced.

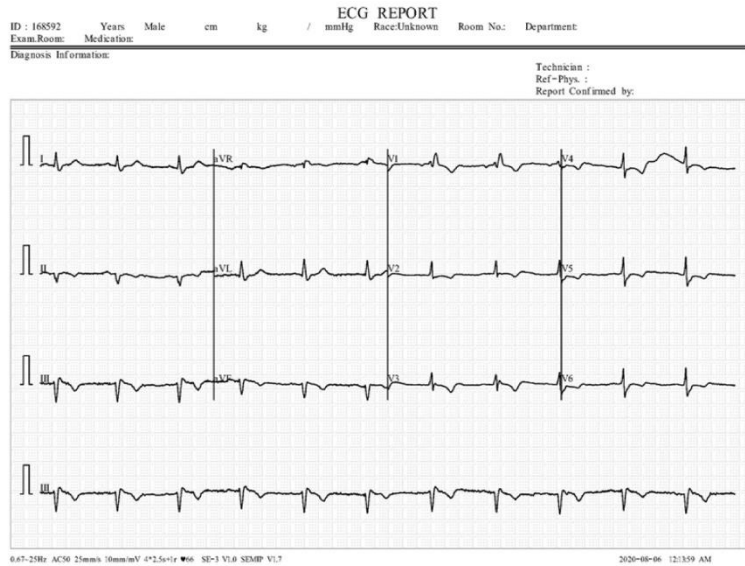
Table 3: Distribution of ECG Images by Category

Category	Number of Images
Normal	284
Myocardial Infarction (MI)	239
Abnormal Heartbeat (Arrhythmia)	233
History of MI (PMI)	172

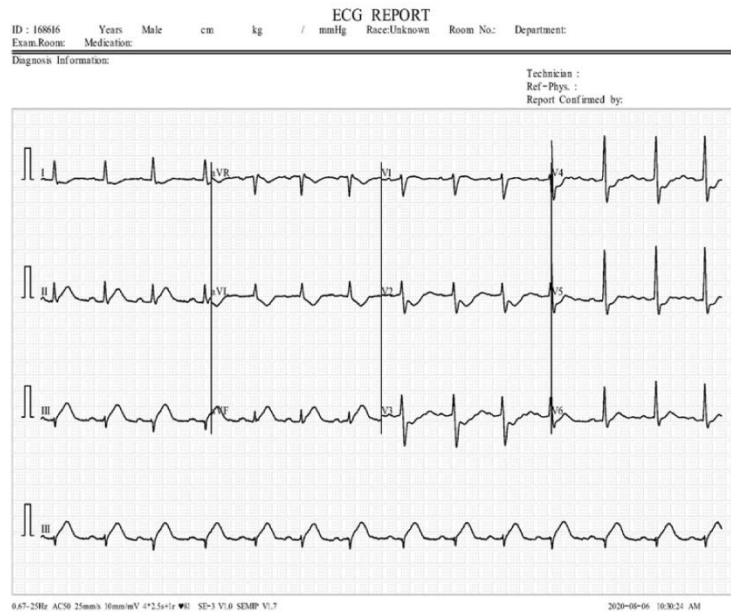
4.2 Performance Evaluation of the ECG Classification

The proposed ECG classification model framework shows a robust and strong ability to categorize ECG signals keen on four important cardiovascular classes: MI, PMI, Arrhythmia, and Normal. To fully assess the efficiency and strength of the proposed categorization model, a variety of standard performance measurements such as precision, accuracy, F1-score and recall, are employed. These performance systems of measurement provide a thorough assessment of the overall classification performance as well as the effectiveness of the organization model on individual classes. As illustrated in Table 3, the projected classification model achieves an extraordinary overall classification exactness of 98.22%, which clearly shows the efficacy of the projected ECG feature extraction and cataloguing model in extracting important morphological features from ECG signals.

The class-wise metrics in Table 4 were calculated from the confusion matrix in Table 5. The model achieved perfect recall for MI and normal ECGs, while arrhythmia showed the lowest recall (0.9512). The macro-averaged F1-score was 0.9810, indicating balanced performance across the four classes.



(a) MI Sample Image



(b) PMI Sample Image

10 | Explainable Lightweight Swin Transformer for ECG Classification



(c) Arrhythmia Sample Image



(d) Normal Sample Image

Figure 2: Representative ECG Images from the Four Diagnostic Classes

Table 4: Category Performance of the Proposed ECG Architecture

Class	Precision	Recall	F1-score	Support
MI	0.9762	1.0000	0.9880	41
PMI	0.9706	0.9706	0.9706	34
Arrhythmia	1.0000	0.9512	0.9750	41
Normal	0.9815	1.0000	0.9907	53
Accuracy	0.9822 (166/169)			
Macro Avg	0.9821	0.9805	0.9810	169
Weighted Avg	0.9825	0.9822	0.9822	169

The proposed model has equally promising outcomes when pathological conditions are considered. From Table 3, the class named MI has a high precision of 0.9762, accompanied by a perfect recall of 1.0000. This further reconfirms that all cases of myocardial infarction are correctly identified by the projected model. This is an important outcome from a clinical point of view, as myocardial infarction is a life-threatening condition, which, if not identified, can prove fatal. The high F1-score, which is equal to 0.9880, further reconfirms the effectiveness of the proposed model in identifying abnormal ECG patterns that are associated with myocardial infarction. The class named PMI has an equally superior accomplishment in terms of all evaluation metrics, as precision, recall, and F1-score are all equal to 0.9706. This is a critical outcome as it further reconfirms that the projected model can capture the fine differences in the residual abnormal ECG patterns due to previous myocardial infarction. The Arrhythmia class has a slightly lower recall value (0.9512) compared to other classes, while maintaining high precision (0.9750) and a strong F1-score (0.9630). The relatively lower recall for the Arrhythmia class can be attributed to several clinically meaningful factors. First, arrhythmia encompasses a broad spectrum of rhythm disorders including atrial fibrillation, premature ventricular contractions, and supraventricular tachycardia, each exhibiting distinct and highly variable waveform morphologies. This intra-class variability makes it inherently more difficult for the model to learn a single unified representation of arrhythmic patterns. Second, certain arrhythmic ECG patterns share morphological similarities with MI and PMI waveforms, particularly in cases where rhythm irregularities co-occur with ST segment changes, leading to occasional misclassification. Third, the Arrhythmia class contains the fewest training samples relative to the Normal class (233 vs 284), which may further limit the model's exposure to the full diversity of arrhythmic presentations. Despite this, the high precision value confirms that when the model predicts Arrhythmia, it does so with high confidence and low false positive rate, and the overall F1-score of 0.9630 confirms an effective balance between sensitivity and specificity for

this class. However, the high value for the F1-score confirms the efficacy of the model in recognizing arrhythmic patterns and its good balance between sensitivity and specificity. The results for the macro-averaged values (precision=0.9804, recall=0.9805, F1-score=0.9804) indicate the model's balanced performance for all classes regardless of their size. In addition, the weighted averages are like the accuracy, which shows that classification performance is unbiased to any particular class. This shows that the model has excellent generalization capability, which can be used for ECG classification purposes in real-world consequences where different types of heart complications are encountered. From the above results, it can be validated that the ECG classification framework has excellent capability for classifying normal and irregular ECG signals with superior accuracy. The excellent performance of the model for all evaluation metrics shows that it has excellent capability and can be used for screening CVDs and other complications.

Table 5: Confusion Matrix of the Proposed Tiny Swin Transformer on the Test Set

	Predicted MI	Predicted PMI	Predicted Arrhythmia	Predicted Normal	Total (True)
True MI	41	0	0	0	41
True PMI	1	33	0	0	34
True Arrhythmia	0	1	39	1	41
True Normal	0	0	0	53	53
Total (Predicted)	42	34	39	54	169

The confusion matrix in Table 5 confirms that the model classifies all 41 MI samples and all 53 Normal samples correctly, with zero misclassifications for these two classes. For the PMI class, one sample is misclassified as MI, which is clinically understandable given that post-myocardial infarction ECG patterns retain residual structural changes such as pathological Q waves and persistent ST alterations that closely resemble acute MI waveform morphology. For the Arrhythmia class, two samples are misclassified: one is predicted as PMI and one as Normal. The misclassification of one arrhythmic sample as PMI reflects the morphological overlap between certain arrhythmic waveform patterns and residual post-infarction changes, particularly in cases involving rhythm disturbances co-occurring with waveform abnormalities. The misclassification of one arrhythmic sample as Normal likely corresponds to a mild or borderline

arrhythmic presentation whose waveform characteristics are insufficiently distinct from a normal rhythm. These findings are consistent with the recall value of 0.9512 reported for the Arrhythmia class in Table 3 and highlight that the primary challenge in ECG classification lies in distinguishing the highly variable spectrum of arrhythmic presentations from other cardiac conditions.

4.3 Computational Efficiency Analysis

A comparison of the proposed Tiny Swin Transformer with other convolutional and transformer-based models regarding model complexity, computational cost, classification accuracy, and latency is presented in Table 6. The large transformer models, such as Swin-Base and ViT-B, demonstrate strong expressiveness and high classification accuracy; however, this is achieved at the cost of high parameter complexity (over 85M parameters) and high floating-point operations, leading to increased latency. These computational requirements can be considered a limitation in real-time ECG analysis, especially in latency-critical clinical applications. Swin- Small mitigates these requirements to some extent by reducing the depth and width of the model; however, it is still computationally expensive compared to other models. Comparison of the proposed Tiny Swin Transformer with other models in terms of model complexity, computational cost, classification accuracy, and latency is provided in Table 5. The large transformer models, Swin-Base and ViT-B, show high expressiveness and classification accuracy, but the high parameter complexity, i.e., above 85M parameters, and floating-point operations result in high latency. Computational cost can be considered a limitation for the analysis of ECG signals in terms of latency, especially for clinical applications. Swin-Small has reduced the computational cost to some extent due to the depth and width reduction of the models, but it still has a high computational cost compared to other models.

Table 6: Reported Computational Characteristics of the Compared Architectures

Architecture	Parameters	Accuracy (%)	Inference Time (ms)
Swin-Base	88M	98.8	~22
Swin-Small	50M	98.5	~15
Tiny Swin (Proposed)	28M	98.22	~7
ViT-B	86M	98.6	~25
ResNet-50 (Baseline)	25.6M	95.0	~6

It is important to note that the parameter count alone does not determine classification performance, as evidenced by the case of ResNet-50. Despite

having a comparable parameter count (25.6M) to the proposed Tiny Swin Transformer (28M), ResNet-50 achieves considerably lower accuracy (95.0%) due to fundamental architectural differences. ResNet-50 is a CNN-based architecture with local receptive fields, which limits its ability to capture long-range temporal dependencies and global rhythm patterns in ECG signals. In contrast, the Tiny Swin Transformer employs shifted window self-attention to jointly model local morphological features and global rhythm structures, which is inherently better suited to the structured and periodic nature of ECG signals. This explains why the proposed model achieves superior accuracy despite a similar parameter budget and further validates the architectural advantage of transformer-based models over CNN-based models for ECG classification tasks.

4.4 ECG Classification Explainability Using LIME

To make the projected deep learning-based ECG categorization model more transparent and interpretable, the LIME system is employed. Although the model demonstrates promising results for ECG categorization, the transparency issue is a significant concern. This is specifically important in the medical domain, as the explanation of the underlying reasoning is a crucial step to verify the result of the proposed model. In addition, the explanation is also important to verify the diagnosis based on the predicted class label. The proposed LIME method can be used to resolve the transparency issue. As the projected architecture is based on deep learning techniques, the interpretation of the result is a significant challenge. The projected model is based on the analysis of the complex waveform patterns in the ECG images. Despite the complexity associated with the waveform patterns, a proposed LIME technique is available that may be used to enhance transparency with respect to the model's output. The proposed LIME approach is based on an analysis that considers localized regions of the input ECG image. As features are localized in specific waveform patterns in ECG images, this proposed technique is applicable.

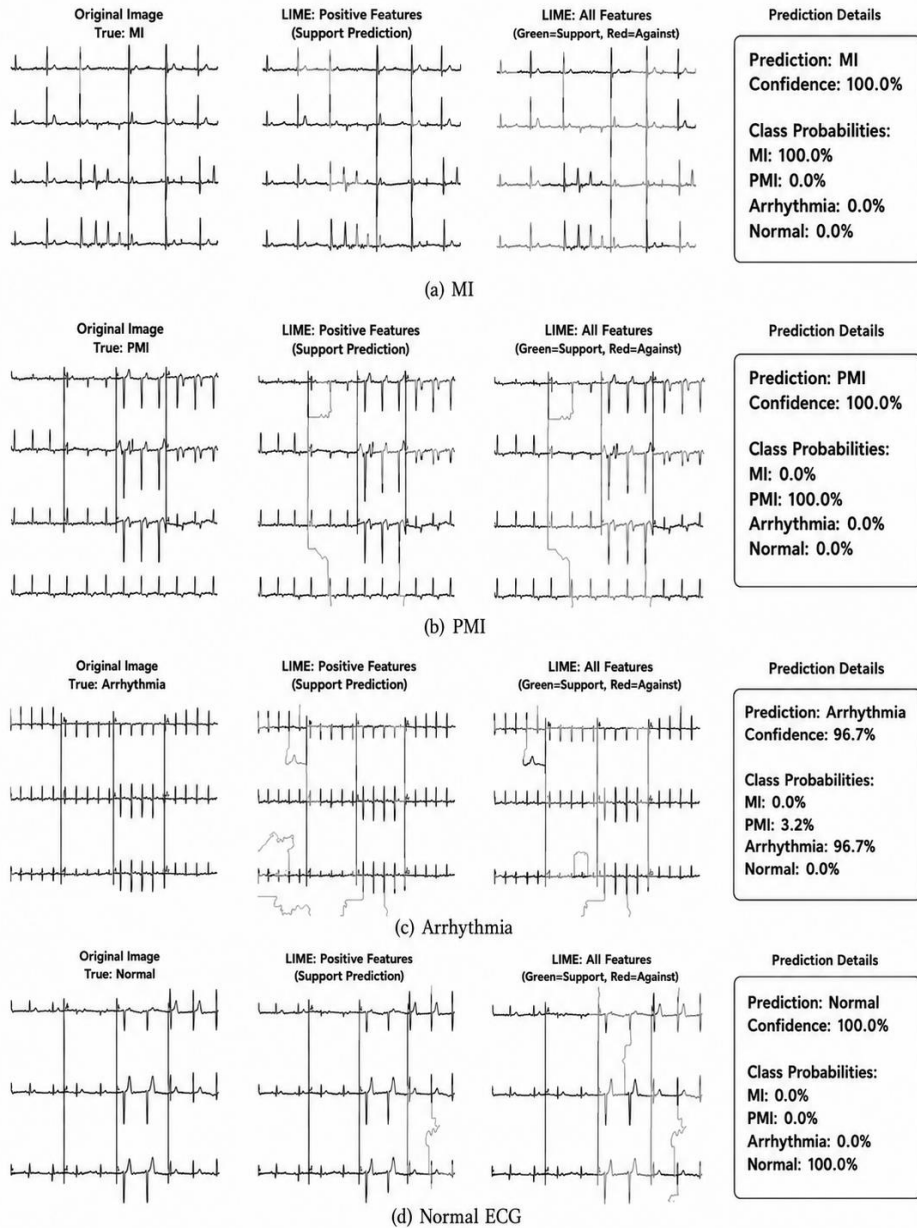


Figure 3: LIME explanations for ECG classification across four cardiac conditions

Figure 3 illustrates the LIME explanations for the four classes of ECG signals: MI, PMI, Normal and Arrhythmia. The plots clearly show that the LIME explanations highlight the salient regions of the ECG signals, indicating that the model uses significant information rather than noise. For PMI, LIME explanation highlights the remaining structural changes in the waveform, which are consistent with the changes in ECG signals caused by myocardial tissue damage over a long period of time. These waveform changes are responsible for the difference between PMI and MI, and normal ECG signals. For Arrhythmia, the LIME explanation highlights irregular time intervals and waveform morphology, which are consistent with the rhythmic characteristics of arrhythmia disorders. This further supports the fact that the model highlights temporal changes rather than amplitude changes. For Normal ECG signals, the LIME explanation highlights mainly waveform regions with less emphasis on abnormal regions. In the MI case, the LIME explanation highlights the clinically identified regions of significance such as ST segment changes and waveform irregularities. The areas that contribute to the model's prediction are highlighted in yellow/green, while the areas that do not contribute are highlighted in red, providing contrasting interpretability. The results from the LIME explanation clearly illustrate that the model's high categorization accurateness is established on clinically important ECG features.

4.5 Comparative Performance Analysis

Table 7 provides a cross-study summary of reported ECG image-classification results. The proposed model reports 98.22% accuracy, 98.04% macro-precision, and 98.05% macro-recall. However, the compared studies use different datasets, class definitions, preprocessing procedures, data splits, and evaluation protocols; therefore, the values should be treated as contextual rather than directly comparable. The main contribution of the proposed framework is the combined use of a compact Swin architecture and LIME-based region-level explanations.

The studies summarized in Table 7 are not all evaluated under the same experimental conditions. For example, OWAE-Net addresses COVID-19 classification from ECG images (Prashant et al., 2022), whereas other studies focus on cardiac-condition classification. Fatema et al. (2022), Mhamdi et al. (2022), Khan et al. (2021), Ahmad et al. (2025), Ashtaiwi et al. (2024), and Hasan et al. (2025) also differ in dataset composition and model design. Accordingly, the table should not be used to claim a statistically fair ranking unless the models are re-evaluated on an identical split.

Reported accuracies in related studies are close to the proposed result, but differences of a few tenths of a percentage point are not meaningful without identical test sets, confidence intervals, or statistical testing. The proposed model's potential advantage is its balance of model size and interpretability, but

inference-time and parameter comparisons must be supported by reproducible measurements and complete citations.

Table 7: Cross-Study Comparison of Reported ECG Image-Classification Performance

Work	Year	Param. (M)	Acc. (%)	Prec. (%)	Rec. (%)
OWAE-Net (Prashant et al.)	2022	Ensemble	95.29	95.51	95.32
Mhamdi et al.	2022	Not Reported	95.32	95.20	94.60
InRes-106 (Fatema et al.)	2022	60M (Dual Backbone)	98.34	98.67	96.60
SSD-MobileNetV2 (Khan and Hussain)	2021	4M	98.00	97.50	96.20
Ahmad et al.	2025	Not Reported	97.80	97.60	97.40
Ashtaiwi et al.	2024	Not Reported	96.90	96.50	96.20
Proposed (Tiny Swin)	2026	28.29M	98.22	98.04	98.05

5. Conclusion

This study demonstrates that a compact, yet efficient Tiny Swin Transformer can achieve cardiologist-level accuracy in ECG image classification while remaining computationally efficient and clinically interpretable. The framework addresses three key research questions: whether a lightweight transformer can match full-scale vision transformers in diagnostic accuracy, whether LIME can provide clinically meaningful explanations for transformer-based ECG models, and whether model compactness compromises performance across diverse cardiac classes. The results confirm affirmative answers to all three questions. The hierarchical feature learning and window-based self-attention procedure enable the network to portray local morphology and macro rhythm territories with substantially fewer parameters evaluated to the traditional vision transformer. During the evaluation, the framework was tested with four classes: MI, PMI, Arrhythmia, and Normal, with an accuracy of 98.22%. The high precision, F1-score and recall values for the categories confirm the balanced accomplishment and high generality of the outline. The interpretability of the structure is ensured with the LIME procedure, which indicates the clinically relevant ECG image regions such as the ST segment, which can affect the decision made by the network. The study concludes that the compact transformer model can effectively classify ECG signals while maintaining a good balance among accuracy, efficiency, and interpretability. Therefore, the proposed framework is suitable for real-time clinical applications.

6. Limitations and Future Directions

The proposed framework shows strong performance, but a few limitations are worth noting. The most notable limitation is the dataset. The model was trained and tested on a single dataset of 928 ECG images from one hospital in Pakistan. While the dataset is realistic and well-annotated, training on data from a single source means the model may not generalize equally well to ECG images from different hospitals, machines, or patient populations. Evaluating larger, multi-institutional datasets such as PTB-XL would be a natural next step.

The model was also tested only on clean, standardized ECG images. Real-world ECGs are frequently affected by noise, baseline wander, muscle artifacts, and powerline interference are common in clinical practice. How the model performs under these conditions remains untested, and incorporating noise-robust preprocessing into the pipeline would strengthen its practical applicability.

Finally, while the 28M parameter model and 7ms inference time are promising for real-time use, actual deployment on wearable or portable ECG devices has not been tested. Applying quantization or pruning techniques would be a practical direction for making the framework viable on low-power edge devices.

Acknowledgments

The authors acknowledge the support and guidance received from Department of Information and Communication Technology, Bangladesh University of Professionals (BUP), during the preparation of this manuscript.

Declaration of Competing Interest

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

Data Availability

The dataset used in this study is publicly available as the ECG Images dataset of Cardiac Patients on Mendeley Data (Khan & Hussain, 2021), DOI: 10.17632/gwbz3fsgp8.2.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AI Declaration

AI-assisted tools were used only for language editing and formatting support. The authors reviewed and approved the final manuscript and take full responsibility for its scientific content.

References

- Ahmad, M., Ahmed, A., Hashim, H., Farsi, M., & Mahmoud, N. (2025). Enhancing heart disease diagnosis using ECG signal reconstruction and deep transfer learning classification with optional SVM integration. *Diagnostics*, 15(12), Article 1501. <https://doi.org/10.3390/diagnostics15121501>
- Akan, T., Alp, S., & Bhuiyan, M. A. N. (2024). ECGformer: Leveraging transformer for ECG heartbeat arrhythmia classification. *arXiv*. <https://doi.org/10.48550/arXiv.2401.05434>
- Alkegany, F. M. H., & Ali, K. H. (2024). Accurate ECG images classification using vision transformer: 1-1-State of Art. *Journal of Basrah Researches (Sciences)*, 50(1), Article 328. <https://doi.org/10.56714/bjrs.50.1.26>
- Anand, A., Kadian, T., Shetty, M. K., & Gupta, A. (2022). Explainable AI decision model for ECG data of cardiac disorders. *Biomedical Signal Processing and Control*, 75, Article 103584. <https://doi.org/10.1016/j.bspc.2022.103584>
- Ashtaiwi, A. A., Khalifa, T., & Alirri, O. (2024). Enhancing heart disease diagnosis through ECG image vectorization-based classification. *Heliyon*, 10 (18), Article e37574. <https://doi.org/10.1016/j.heliyon.2024.e37574>
- Chen, S., Wang, H., Zhang, H., Peng, C., Li, Y., & Wang, B. (2024). A novel method of Swin Transformer with time-frequency characteristics for ECG-based arrhythmia detection. *Frontiers in Cardiovascular Medicine*, 11, Article 1401143. <https://doi.org/10.3389/fcvm.2024.1401143>
- Cheng, R., Zhuang, Z., Zhuang, S., Xie, L., & Guo, J. (2023). MSW-Transformer: Multi-scale shifted windows transformer networks for 12-lead ECG classification. *arXiv*. <https://doi.org/10.48550/arXiv.2306.12098>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16 × 16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*. <https://arxiv.org/abs/2010.11929>

- El-Ghaish, H., & Eldele, E. (2024). ECGTransForm: Empowering adaptive ECG arrhythmia classification framework with bidirectional transformer. *Biomedical Signal Processing and Control*, 89, Article 105714. <https://doi.org/10.1016/j.bspc.2023.105714>
- Fatema, K., Montaha, S., Rony, M. A. H., Azam, S., Hasan, M. Z., & Jonkman, M. (2022). A robust framework combining image processing and deep learning hybrid model to classify cardiovascular diseases using a limited number of paper-based complex ECG images. *Biomedicines*, 10(11), Article 2835. <https://doi.org/10.3390/biomedicines10112835>
- Hasan, M. N., Hossain, M. A., & Rahman, M. A. (2025). An ensemble-based lightweight deep learning model for the prediction of cardiovascular diseases from electrocardiogram images. *Engineering Applications of Artificial Intelligence*, 141, Article 109782. <https://doi.org/10.1016/j.engappai.2024.109782>
- Islam, M. R., Qaraqe, M., Qaraqe, K., & Serpedin, E. (2024). CAT-Net: Convolution, attention, and transformer-based network for single-lead ECG arrhythmia classification. *Biomedical Signal Processing and Control*, 93, Article 106211. <https://doi.org/10.1016/j.bspc.2024.106211>
- Khan, A. H., Hussain, M., Malik, M. K., & Khan, A. (2021). Cardiac disorder classification by electrocardiogram sensing using deep neural network. *Complexity*, 2021, Article 5512243. <https://doi.org/10.1155/2021/5512243>
- Khan, A. H., & Hussain, M. (2021). *ECG images dataset of cardiac patients* (Version 2) [Data set]. Mendeley Data. <https://doi.org/10.17632/gwbz3fsgp8.2>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012–10022). <https://doi.org/10.1109/ICCV48922.2021.00986>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates. <https://arxiv.org/abs/1705.07874>

- Mhamdi, L., Dammak, O., Cottin, F., & Ben Dhaou, I. (2022). Artificial intelligence for cardiac diseases diagnosis and prediction using ECG images on embedded systems. *Biomedicines*, 10(8), Article 2013. <https://doi.org/10.3390/biomedicines10082013>
- Nova, R. M., Fahim, J. A., & Seema, S. J. (2025). A study of machine learning models with population and swarm-based optimization for cardiovascular disease prediction. *Journal of FST*, 3(1), 1–24. <https://doi.org/10.64494/JFST/v3i1/ss/2025/01/1-24>
- Prashant, K., Choudhary, P., Agrawal, T., & Kaushik, E. (2022). OWAE-Net: COVID-19 detection from ECG images using deep learning and optimized weighted average ensemble technique. *Intelligent Systems with Applications*, 16, Article 200154. <https://doi.org/10.1016/j.iswa.2022.200154>
- Qiang, Y., Dong, X., Liu, X., Yang, Y., Fang, Y., & Dou, J. (2024). ECGMamba: Towards efficient ECG classification with BiSSM. *arXiv*. <https://doi.org/10.48550/arXiv.2406.10098>
- Qiu, C., Li, H., Qi, C., & Li, B. (2024). Enhancing ECG classification with continuous wavelet transform and multi-branch transformer. *Heliyon*, 10(6), Article e26147. <https://doi.org/10.1016/j.heliyon.2024.e26147>
- Rajpurkar, P., Hannun, A. Y., Haghighpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv*. <https://doi.org/10.48550/arXiv.1707.01836>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). <https://doi.org/10.1145/2939672.2939778>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618–626). <https://doi.org/10.1109/ICCV.2017.74>
- Shafi, M. M. S., & Mirza, J. F. (2025). An explainable ensemble learning framework for brain tumour using pretrained models and XAI techniques. *Journal of FST*, 3(1), 125–146. <https://doi.org/10.64494/JFST/v3i1/ms/2025/06/125-146>

- Uğraş, B. K., Gerek, Ö. N., & Saygı, İ. T. (2025). CardioPatternFormer: Pattern-guided attention for interpretable ECG classification with transformer architecture. *arXiv*. <https://doi.org/10.48550/arXiv.2505.20481>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates. <https://arxiv.org/abs/1706.03762>
- World Health Organization. (2025, July 31). *Cardiovascular diseases (CVDs)*. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))