Cardiology in the Young

cambridge.org/cty

CardioFusion-XAI: a robust multi-scale and explainable framework for cardiac MRI-based disease classification

Shwetambari Borade, Saraswati Mishra, Rupali Vairagade and Pranali Pawar

Cyber Security, Shah & Anchor Kutchhi Engineering College, India,

Original Article

Cite this article: Borade S, Mishra S, Vairagade R, and Pawar P (2026) CardioFusion-XAI: a robust multi-scale and explainable framework for cardiac MRI-based disease classification. *Cardiology in the Young* **36**: 1795–1802. doi: 10.1017/S1047951126123671

Received: 25 February 2026 UTC
Revised: 19 June 2026 UTC
Accepted: 27 June 2026 UTC
First published online: 3 September 2026 UTC

Keywords:
Cardiac MRI; cardiovascular disease classification; Grad-CAM; explainable artificial intelligence; hybrid deep learning; vision transformer

Corresponding author:
Shwetambari Borade;
Email: shwetambari.borade@sakec.ac.in

Abstract

Background: Cardiovascular diseases remain a major global health burden, and early diagnosis using cardiac magnetic resonance imaging (MRI) is essential. However, manual interpretation is subjective, while existing deep learning models often show inconsistent preprocessing, limited multi-scale feature learning, and poor interpretability. *Methods:* CardioFusion-XAI, a hybrid explainable deep learning framework, was developed for four-class cardiac MRI classification. WaveCLAHE-Net integrates wavelet-based enhancement and contrast-limited adaptive histogram equalisation to improve image quality. CardioXtract Fusion Net combines ResNet-50, EfficientNet-B0, and a Vision Transformer to learn complementary local, multi-scale, and global features. An enhanced classification head performs prediction, while Grad-CAM and t-distributed stochastic neighbour embedding support model interpretability. *Results:* Using the Sunnybrook Cardiac Dataset, CardioFusion-XAI achieved 98.89% accuracy, 98.93% precision, a 98.90% F1-score, and an area under the curve of 0.9998. *Conclusions:* CardioFusion-XAI provides accurate and interpretable cardiac MRI classification. Further multicentre validation is required before clinical implementation.

Introduction

Cardiovascular disease is one of the most severe and widespread global health problems of death and disability, due to its predominance. Cardiovascular diseases encompass diverse heart and blood vessel disorders through elaborate and sometimes unsteady pathological routes.[1] If not recognised or poorly treated, cardiovascular diseases often cause acute myocardial infarction, chronic heart failure, and disabling strokes, as well as serious and life-threatening accidents.[2] Early and reliable diagnosis is vital for effective cardiovascular disease management and enhancing patient outcomes.

MRI was widely used for evaluating the non-invasive cardiac due to the superior soft tissue contrast and functional image capability.[3] Cardiac MRI offers clinical information for diagnosing cardiomyopathies, heart failure, and myocardial abnormalities.[4] However, the manual interpretation of multi-parametric MRI scans was time-consuming due to inter-observer variability and reduced diagnostic consistency in the complex clinical cases.[5]

Deep learning is a recent advancement of convolutional neural networks and vision transformers enhanced the automated cardiac MRI analysis.[6] ResNet and EfficientNet frameworks have strong classification capability, and the transformer-based models enhance global contextual learning via a self-attention mechanism.[7] The hybrid convolutional neural network-transformer framework enhances the feature representation of local spatial information with long-range anatomical dependencies.[8]

Recent progress has several challenges to the clinical adoption of deep learning for cardiac MRI classification. However, convolutional neural network-based models suffer from capturing long-range anatomical dependency, and transformer-based approaches require high computational resources and large-scale training data.[9,10] Some studies lack robust preprocessing, an explainability mechanism, and reduce clinical trust, model transparency.[11,12] To address these issues, a novel CardioFusion-XAI framework has been proposed and evaluated specifically for four-class cardiac MRI classification using the Sunnybrook Cardiac Dataset, consisting of heart failure with infarction, heart failure, hypertrophy, and normal cardiac categories. Therefore, the scope of this work is limited to automated classification within this dataset rather than comprehensive cardiovascular disease diagnosis.

Although the present study is evaluated using the Sunnybrook Cardiac Dataset, which primarily contains adult cardiac MRI cases, the proposed framework is designed as a general cardiac MRI analysis approach. The methodology may provide a foundation for future investigations involving paediatric and CHD populations, where accurate and interpretable cardiac image analysis is also clinically important. However, no paediatric or congenital cardiac datasets were included in the current study.

© The Author(s), 2026. Published by Cambridge University Press.

The contributions of the proposed CardioFusion-XAI framework are as follows:

- This study proposed a novel CardioFusion-XAI framework to enhance cardiac MRI quality via noise reduction, contrast enhancement, and structural normalisation.
- The proposed model combined with hybrid deep learning, like EfficientNet-B0, ResNet-50, and vision transformers, to capture both fine-grained local textures and long-range global anatomical dependency from cardiac MRI images.
- The proposed model used multi-scale convolutional neural network-transformer feature fusion to enhance discriminative cardiovascular disease classification performance and patient-level generalisation.
- The proposed model included explainable AI techniques like Grad-CAM and *t*-SNE visualisation to offer clinically interpretable decision support and enhance model transparency.

The organisation of the paper is structured as follows: Section 2 reviews various deep learning approaches for cardiovascular disease classification via a cardiac MRI. Section 3 explains in detail the proposed CardioFusion-XAI framework. Section 4 discuss about a result and analysis. Finally, Section 5 presents the conclusion and future work.

Literature survey

This section presents a deep learning and medical image, which enhances the automated cardiovascular disease diagnosis via a cardiac MRI. Some studies are focused on hybrid, machine learning, and transformer framework are enhanced diagnostic accuracy and reliability.

Sadr et al.[13] suggested a hybrid deep learning and machine learning framework for cardiovascular disease prediction via both the public and clinical datasets. This model integrated convolutional neural network, RNN, and classical classifiers for enhancing both the prediction accuracy and reliability. The suggested framework involves high architectural complexity but lacks validation across various demographic and acquisition settings.

Baghdadi et al.[14] developed a gradient boosting-based cardiovascular disease prediction framework for early diagnosis via clinical records. This model attained high accuracy and the F1-score for effective asymptomatic case detection. However, the framework mainly focused on asymptomatic prediction and depended on the quality and completeness of clinical features.

Ogunpola et al.[15] assessed seven machine learning models, which include convolutional neural network-based approaches for cardiovascular disease detection. The optimised gradient boosting model was achieved across various evaluation metrics for improving the diagnostic reliability. However, the suggested approach overfits risks and requires wide optimisation procedures, which lack practical reproducibility.

Liu et al.[16] presented a neural classifier that utilises self-attention operations to model long-dependency for electrocardiography signal diagnosis. The suggested model was evaluated on the public Tianchi electrocardiography dataset, which integrates lead fusing blocks, transformer-XL encoder-based encoder modules. However, their study focuses on enhancing the interpretability, robustness, and expanding the clinical applications of AI-driven electrocardiography diagnosis.

Cockrum et a.[17] advanced a deep learning algorithm that helped discriminate cardiac amyloidosis, both cine and late gadolinium enhancement of cardiac MRI. Their study, which was evaluated by the Ohio dataset, has the ability of the vision transformers model to differentiate the cardiac amyloidosis from hypertrophic cardiomyopathy. However, their study lacks the accuracy of disease detection to characterise the clinical decision support models.

Telangore et al.[18] developed an artificial intelligence model of Sudden Cardiac Death (SCD) for early prediction via the electrocardiography signals. The suggested approach was evaluated by the MITBIH SCDs Holter Dataset, which enhances early SCD prediction as a valuable tool for continuous electrocardiography monitoring in high-risk patients. However, the suggested model predicts SCD a few minutes before onset, which limits the intervention time

Mirzaee Moghaddam Kasmaee et al.[19] introduced a deep learning model called ELRLMD, which integrates ensemble learning and Reinforcement Learning for effective myocarditis diagnosis from cardiac magnetic resonance images. The suggested model was evaluated by the cardiac magnetic resonance dataset, which lacks risk convergence on local optima due to a suboptimal initial weight setting. The summary of existing cardiovascular disease classification approaches is provided in Supplementary Table S1.

Existing studies depend on a single-model framework or traditional machine learning methods with handcrafted features, which fail to capture both fine local textures and long-range contextual dependency in the complex cardiac MRI data. Moreover, some studies lack to integrates explainability mechanisms, which limits both the interpretability and clinical trust. To address these issues, this study proposes a novel framework that integrates enhanced preprocessing, hybrid convolutional neural network-transformer feature fusion, and explainable AI techniques for accurate and interpretable cardiac MRI classification.

Materials and methods

This study proposed a novel CardioFusion-XAI framework to enhance cardiac MRI quality via noise reduction, contrast enhancement, and structural normalisation. Initially, the Sunnybrook Cardiac Dataset (SBCD), which consists of fine cardiac MRIs, categorised into four like heart failure with infarction, hypertrophy, heart failure without infarction, and normal cardiac anatomy (N). The hybrid processing step includes wavelet-based denoising to remove the noise, whilst keeping the fine anatomical details and the structural boundaries intact. Next, contrast-limited adaptive histogram equalisation increases the local contrast cleverly, not exposing extremely subtle pathological features, but rather preventing the creation of any artificial artefacts.

The hybrid feature extraction step, supported by CardioXtract Fusion Net, forms the backbone of the proposed approach. The features that were extracted are fed into an intelligent classification for superior performance and generalisation. Global average pooling and dropout-regularised multilayer perceptron layers are applied for efficient classification and overfitting reduction. The properties that have been handled are then sent to a multilayer perceptron that employs dropout regularisation as a form of enhancement. Dropout is a method that works against overfitting by randomly turning off neurons during training. A detailed architectural framework and its flow are shown in Figure 1. Existing cardiac MRI classifications are dependent on convolutional neural network-only or transformer-only frameworks. In contrast, the proposed CardioFusion-XAI

https://doi.org/10.1017/S1047951126123671 Published online by Cambridge University Press

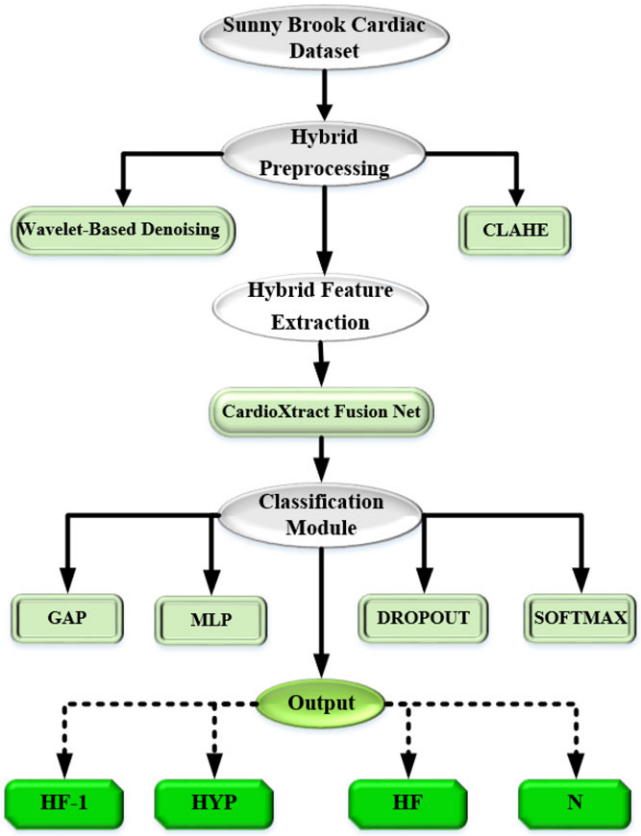


Figure 1. Architectural framework for cardiac disease classification.

Table 1. Patient-level class distribution of the Sunnybrook Cardiac Dataset

Class	Total patients
Heart failure with infarction	12
Heart failure	12
Hypertrophy	12
N	9
Total	45

model integrates Wavelet-contrast-limited adaptive histogram equalisation preprocessing, complementary convolutional neural network feature extraction, transformer-based contextual fusion, and explainable AI mechanisms for robust and interpretable patient-level cardiac MRI classification.

Image acquisition

The proposed model was evaluated by an SBCD, which was publicly available as the 2009 Cardiac MR Left Ventricle Segmentation Challenge dataset. It contains 45 cine-MRI studies classified into four classes are shown in Table 1. It also includes 233 cine-MRI sequences, approximately 51,515 MRI slices with expert-annotated contours for reliable ground-truth validation. Five-fold cross-validation was performed using patient-level splitting to prevent inter-slice data leakage. All MRI slices belonging to a single patient were assigned exclusively to either the training or validation subset within a given fold. During evaluation, slice-level predictions generated by the proposed framework were aggregated into a single patient-level prediction using majority voting across all slices belonging to the same patient. Thus, the dataset was publicly available, fully anonymised, and no additional ethical approval was required.

Preprocessing

Pre-processing plays a vital role in the hybrid WaveCLAHE Net to enhance the cardiac MRI quality before the model training. Initially, all the MRI images are resized into 256×256 pixels for input standardisation. Then, wavelet-based denoising was applied to suppress noise and maintain anatomical structures. Thus, the denoised image is reconstructed by inverse discrete wavelet transform after the soft-thresholding[20] of wavelet coefficients was expressed as Equation (1):

$$\tilde{w}_{j,k} = \text{sign}(w_{j,k}) \cdot \max(|w_{j,k}| - T_j, 0) \quad (1)$$

where the wavelet coefficients are denoted by $w_{j,k}$, At j, the threshold values are expressed by T_j. Then contrast-limited adaptive histogram equalisation[21] was applied to improve both the local contrast and normalise intensity variations in cardiac MRI images. The histogram equalisation mapping function is expressed as Equation (2):

$$S_k = T(r_k) = \frac{L-1}{MN} \sum_{j=0}^{k} n_j \quad (2)$$

where the total number of pixels is denoted by MN, r_j denotes the intensity of the number of pixels expressed by n_j, and the number of intensity levels was represented by L. Finally, Z-score normalisation was performed to standardise intensity distributions across the dataset.

Feature extraction

After preprocessing, a hybrid feature extraction integrates EfficientNet-B0,[7] ResNet-50,[22,] and vision transformers[23] to capture both the local and global cardiac representations from MRI images. In that, EfficientNet-B0 and ResNet-50 extract discriminative spatial features, and the vision transformers performs global contextual fusion to model long-range anatomical dependency. The Vision Transformer utilised a patch size of 16×16, embedding dimension of 768, 12 transformer encoder layers, 12 attention heads, and a multilayer perceptron dimension of 3072. The detailed architecture of the proposed CardioXtract Fusion Net model was illustrated in Supplementary Figure S2.

EfficientNet-B0

The EfficientNet-B0 is used to extract both the fine-grained local features and myocardial texture representations from cardiac MRI images. From Block 7, the extracted high-level feature representations are expressed as Equation (3):

$$F = \text{EffB0}_{\text{block 7}}(I) \quad (3)$$

where the input cardiac MRI images are denoted by I, from EfficientNet-B0, the extracted high-level feature map is represented by F. The extracted feature map is partitioned into non-overlapping patches and projected into D-dimensional embeddings, which are denoted as Equation (4):

https://doi.org/10.1017/S1047951126123671 Published online by Cambridge University Press

$$P_{\text{emb}} = \text{Conv2D}(D,\ \text{kernel size} = P,\ \text{stride} = P)(F) \quad (4)$$

where the patch size is denoted by P, the embedding dimensions are expressed by D, and the patch embedding is denoted by P_{emb}. The extracted embeddings are a discriminative local cardiac information for consequent feature fusion. The EfficientNet-B0 framework was illustrated in Supplementary Figure S3.

ResNet-50

The proposed model integrates ResNet-50 to learn both the hierarchical spatial and anatomical representations from cardiac MRI images. The extracted deep features from the final residual stage were formulated as Equation (5):

$$F = \text{ResNet50}_{\text{block 5}}(I) \quad (5)$$

where the input cardiac MRI images are expressed by I, via ResNet-50, the high-level feature maps are denoted by F. The extracted feature maps are converted into structured patch embeddings to maintain regional cardiac information for fusion learning. The architecture of ResNet-50 is shown in Supplementary Figure S4.

Vision transformer-based feature fusion in the cardioXtract fusion net

From the EfficientNet-B0 and ResNet-50, the features are extracted, and the vision transformers is used to perform a global contextual fusion. The synchronised convolutional neural network feature maps are represented as Equation (6):

$$F = [F_{\text{ResNet50}} \| F_{\text{EffB0}}] \quad (6)$$

where the ResNet-50 and EfficientNet-B0 extract features are denoted by $F_{ResNet50}$ and F_{EffB0}. The fused feature map is partitioned into non-overlapping patches and projected into D-dimensional embeddings expressed as Equation (7):

$$z_i = W_e \cdot \text{Flatten}(P_i) + b_e \quad (7)$$

where the i^{th} The image patch was denoted by P_i, learnable projection parameters are expressed by W_e,b_e, and the embedded token is represented by z_i. Positional embeddings are added to maintain a spatial relationship among patches as indicated in Equation (8):

$$Z_0 = [z_1 + E_1, z_2 + E_2, \ldots, z_N + E_N] \quad (8)$$

where the i^{th} patch tokens are associated with the positional embedding denoted by E_i, and for the transformer encoder, the input token sequence is represented by Z_0. The multi-head self-attention[23] mechanism is applied to model long-range feature dependency, denoted by Equation (9):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

where the query, key, and value matrices are denoted by Q, K, V, and the key dimensions are represented by d_k. The fused global representation is denoted as Equation (10):

$$F_{\text{ViT}} = \text{MSA}(Z_0) \quad (10)$$

where from the vision transformers, the globally contextualised features are denoted by F_{ViT}. It enhances global contextual learning and discriminative cardiac disease representation.

Classification

After the CardioXtract Fusion Net model, the fused features are forwarded into classification for cardiovascular disease prediction. Initially, the global average pooling layer reduces spatial dimensions of the fused feature maps and maintains the most discriminative global representations. Then, the pooled features are fed into the multilayer perceptron with ReLU activation and dropout regularisation to enhance nonlinear feature learning and reduce overfitting. A dropout rate of 0.5 was employed within the multilayer perceptron classification head to reduce overfitting. Finally, the Softmax layer generates normalised probability scores for multi-class cardiac disease classification, such as heart failure, heart failure without infarction, hypertrophy, and normal classes. The detailed classification process is mathematically formulated in Supplementary Section S3.4.1–S3.4.4.

Visual analysis using Grad-CAM and t-SNE

This study employed Grad-CAM as a visual analysis tool to highlight image regions that contributed strongly to convolutional neural network-based feature extraction. Although Grad-CAM provides useful qualitative insights into model attention, it does not fully explain the final decision-making process of the integrated convolutional neural network–transformer architecture.[24] to improve clinical interpretability and generate visual heatmaps that highlight diagnostically relevant regions in cardiac MRI images. The generated activation maps help to identify anatomical structures, including myocardial walls and ventricular boundaries, which contribute to disease classification. Additionally, t-SNE was utilised to visualise the distribution and separability of learned feature representations across cardiac disease categories. These visualisations support qualitative assessment of feature discrimination but should not be interpreted as evidence of clinical explainability.[25]

Result and analysis

The proposed CardioFusion-XAI framework was evaluated by five-fold cross-validation for cardiovascular disease classification from cardiac MRI images. It was evaluated by several performance metrics like accuracy, precision, recall, specificity, F1-score, Cohen's Kappa, and AUC metrics. Additionally, Grad-CAM visualisations and t-SNE feature projections are used to analyse the model interpretability and feature separability. The detailed preprocessing outcomes are provided in Supplementary Figure S5.

Performance metrics

The proposed CardioFusion-XAI framework was evaluated by five-fold level cross-validation and several metrics to ensure a reliable generalisation.

Implementation details and experimental configuration

The proposed CardioFusion-XAI framework was implemented on Python, TensorFlow, and Keras. ImageNet-pretrained EfficientNet-B0 and ResNet-50 models are integrated by the vision transformers

https://doi.org/10.1017/S1047951126123671 Published online by Cambridge University Press

fusion module. The model was trained by the Adam optimiser. Hyperparameters, including learning rate (0.0001), batch size (16), dropout rate, and training epochs, were selected based on preliminary validation experiments and prior literature on cardiac MRI classification. Early stopping and learning-rate scheduling were employed to reduce overfitting and improve model generalisation. Early stopping and learning rate reduction are applied to enhance convergence and reduce overfitting. Data augmentation included random rotation (±15°), horizontal flipping, zooming (0.9–1.1), width shifting (±10%), and brightness adjustment (0.8–1.2). Several data augmentations include random rotation, horizontal flipping, zooming, width shifting, and brightness adjustment. For all experiments, a fixed random speed of 42 was used for reproducibility. Thus, the framework was trained on an NVIDIA RTX 3080 GPU with 32 GB RAM and approximately 34.8 million trainable parameters. End-diastolic and end-systolic MRI phases are selected for the left ventricle for analysis. Model optimisation was performed using categorical cross-entropy loss and the Adam optimiser. Early stopping monitored validation loss with a patience value of 10 epochs, restoring the best-performing model weights. The learning rate was reduced by a factor of 0.5 when validation loss failed to improve for five consecutive epochs.

Quantitative evaluation metrics

The proposed CardioFusion-XAI framework was evaluated by several metrics, which were evaluated as follows:

- Accuracy represents the overall proportion of correctly classified instances among all evaluated samples and is computed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

- Precision indicates the reliability of positive predictions by measuring the ratio of correctly predicted positive cases to all predicted positives, expressed as

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

- Sensitivity (Recall) assesses the model's ability to correctly identify true positive cases from all actual positives, formulated as

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

- Specificity reflects the capability of the model to recognise true negative cases among all actual negatives correctly and is calculated as

$$Specificity = \frac{TN}{TN + FP} \tag{19}$$

The proposed CardioFusion-XAI framework achieved high classification performance across all five cross-validation folds. To assess the stability of the proposed framework, performance metrics are reported as mean ± standard deviation across the five validation folds. The relatively low variability observed among folds indicates consistent model behaviour despite the limited dataset size. Nevertheless, the small number of available patients may lead to optimistic performance estimates, and therefore, the reported results should be interpreted with appropriate caution. The average metrics attained an accuracy of (0.9889 ± 0.0061), Precision (0.9893 ± 0.0058), Recall (0.9889 ± 0.0061), and F1-score (0.9890 ± 0.0060), showing a strong predictive capability with low variability across folds. Then, the model achieved high Specificity (0.9963 ± 0.0023), Cohen's Kappa (0.9848 ± 0.0089), and AUC (0.9998 ± 0.0003), which represent excellent class separability and robust generalisation performance. The detailed fold-wise quantitative metrics are provided in Supplementary Table S3.

ROC and precision-recall curve analysis

The Receiver Operating Characteristic curve (ROC) and precision-recall analyses showed a strong discriminative performance across various cardiovascular diseases via a five-fold cross-validation. The detailed ROC curve is provided in Supplementary Figure S6.

Patient-level evaluation strategy

The class receiving the highest number of slice-level predictions was assigned as the final patient-level diagnosis. All quantitative metrics reported in this study were computed using patient-level predictions obtained from five-fold cross-validation.

Training and validation performance

The proposed CardioFusion-XAI framework has stable convergence and effective learning behaviour across all five folds with minimal overfitting. Both the training and validation curves have strong generalisation capability. Detailed training-validation accuracy and loss curves are provided in Supplementary Figures S7 and S8.

Confusion matrix analysis

The confusion matrix analysis has highly accurate classification across various cardiovascular diseases with less inter-class misclassification in five-fold cross-validation. Detailed confusion matrices are provided in Supplementary Figure S9. The confusion matrices presented in Supplementary Figure S9 correspond to patient-level predictions obtained after aggregation of slice-level outputs using majority voting. Each matrix, therefore, represents the final diagnostic classification of independent patients rather than individual MRI slices.

Interpretability analysis using Grad-CAM and t-SNE

The proposed framework employed Grad-CAM and *t*-SNE as supporting visual analysis techniques. Grad-CAM highlighted image regions that contributed strongly to convolutional neural network feature extraction, whereas *t*-SNE illustrated the separability of learned feature representations across disease categories. While these visualisations provide useful qualitative insights into model behaviour, they should not be interpreted as comprehensive explanations of the final convolutional neural network–transformer fused decision process or as definitive evidence of clinical interpretability. Detailed Grad-CAM visualisations are presented in Supplementary Figure S10.

https://doi.org/10.1017/S1047951126123671 Published online by Cambridge University Press

Table 2. Ablation study of the proposed CardioFusion-XAI framework

Model variant	Accuracy	Precision	Recall	F1 score	AUC
Without vision transformers	0.9554	0.9561	0.9554	0.9555	0.9691
Without ResNet-50 branch	0.9812	0.9820	0.9812	0.9810	0.9918
Without EfficientNet-B0 branch	0.9835	0.9842	0.9833	0.9837	0.9929
Without Hybrid preprocessing (WaveCLAHE Net)	0.9678	0.9685	0.9674	0.9675	0.9787
Without CLAHE	0.9845	0.9851	0.9849	0.9852	0.9934
Without wavelet	0.9817	0.9822	0.9825	0.9821	0.9918
Full CardioFusion-XAI	0.9889	0.9893	0.9889	0.9890	0.9998

Ablation study

Table 2 displays the ablation study of individual modules and the proposed model. Standalone backbone architectures, including ResNet-50, EfficientNet-B0, and Vision Transformer, were evaluated under identical experimental settings. The results demonstrate that the proposed hybrid architecture consistently outperformed individual backbone models, confirming the complementary benefits of multi-scale convolutional neural network feature extraction and transformer-based contextual fusion. The proposed CardioFusion-XAI model attained 98.89% accuracy and 0.9998 AUC. Removing vision transformers, convolutional neural network branches, or preprocessing modules confirms that hybrid feature fusion and image enhancement are crucial for accurate cardiac MRI classification.

Comparative analysis

The proposed framework was compared with several existing approaches. For fair comparison, all baseline models were reimplemented and evaluated using the same Sunnybrook Cardiac Dataset, patient-level five-fold cross-validation protocol, preprocessing pipeline, data augmentation strategy, optimiser settings, batch size, learning rate schedule, and early-stopping criteria employed by the proposed framework. This ensured that performance differences were attributable to model architecture rather than variations in experimental configuration, like convolutional neural network + LSTM,[13] VGG16,[26,] and conventional 2D convolutional neural network,[27] via the same dataset, preprocessing strategy, and patient-level data splitting for fair evaluation.

Figure 2 demonstrates the performance comparison of the proposed and other existing approaches across various evaluation metrics. The proposed model attained high accuracy (98.89%), F1-score (98.90%), and specificity (99.63%), showing superior discriminative capability for cardiac MRI classification. This enhancement confirms the effectiveness of hybrid convolutional neural network-transformer feature fusion and enhanced preprocessing in learning complex cardiac representations. External multi-centre validation wasn't performed due to the limited publicly available annotated cardiac MRI datasets. Table 3 represents the patients distribution across five cross-validation folds

Figure 2 presents the comparative performance analysis of the proposed CardioFusion-XAI framework against convolutional neural network + LSTM, VGG16, and conventional 2D convolutional neural network models. The proposed framework achieved the highest performance across all evaluation metrics, attaining an accuracy of 98.89%, sensitivity of 98.89%, precision of 98.93%, F1-score of 98.90%, and specificity of 99.63%. Among the baseline methods, VGG16 demonstrated the strongest performance with an accuracy of 96.4% and specificity of 97.7%, whereas the conventional 2D convolutional neural network achieved the lowest overall performance. The superior results of

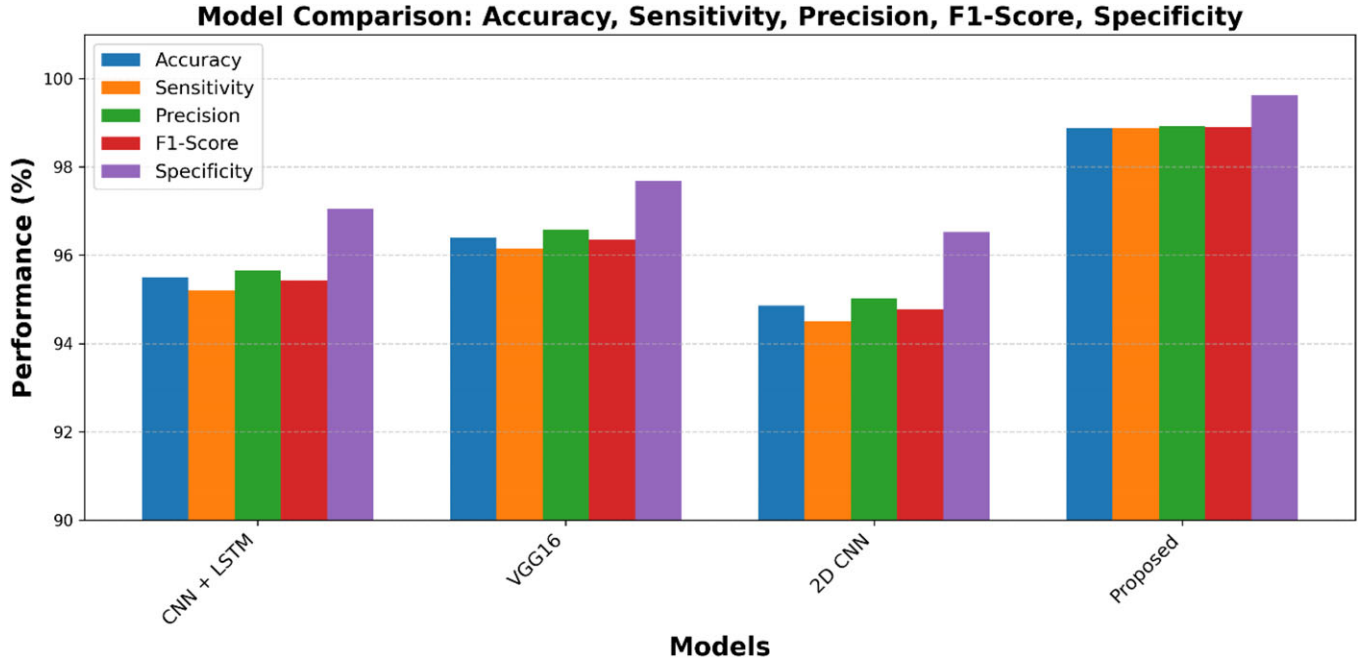


Figure 2. Comparative performance analysis of baseline models and proposed method.

https://doi.org/10.1017/S1047951126123671 Published online by Cambridge University Press

Table 3. Patient distribution across five cross-validation folds

Fold	Training patients	Validation patients
Fold 1	36	9
Fold 2	36	9
Fold 3	36	9
Fold 4	36	9
Fold 5	36	9

CardioFusion-XAI can be attributed to the integration of WaveCLAHE preprocessing, complementary convolutional neural network-based feature extraction using EfficientNet-B0 and ResNet-50, and Vision Transformer-based feature fusion, which effectively capture both local anatomical patterns and global contextual information. These findings confirm the effectiveness of the proposed hybrid architecture for accurate and robust cardiac MRI classification. Table 4 presents the Comparison Table for proposed work.

Discussion

The proposed CardioFusion-XAI framework attained superior cardiac MRI classification performance via the integration of WaveCLAHE preprocessing, hybrid convolutional neural network-based feature extraction, vision transformers-based contextual fusion, and regularised classification. The proposed model attained 98.89% accuracy and 98.90% F1-score compared to other existing approaches. The preprocessing stage enhances image quality, local contrast, and the hybrid feature extraction strategy effectively captures both the local textures and global anatomical dependencies. The proposed model attained high performance but lacks several challenges. Although the proposed framework achieved promising performance, its evaluation was limited to the Sunnybrook Cardiac Dataset, containing four cardiac condition categories. Therefore, the findings should be interpreted as evidence of effectiveness for four-class cardiac MRI classification rather than comprehensive cardiovascular disease diagnosis. Additional validation on larger multi-centre datasets encompassing a wider range of cardiac pathologies is required before considering broader clinical applicability. The Sunnybrook dataset lacks several subjects, and doesn't perform external multi-centre validation. Additionally, the proposed framework only evaluates adult cardiac MRI datasets but does not include paediatric or congenital cardiac conditions. Future research focuses on larger multi-centre datasets, multimodal learning, and privacy-preserving collaborative training strategies.

Although Grad-CAM and *t*-SNE provided useful visual insights into feature learning and model attention, these methods offer only partial explanations of the proposed framework. Specifically, Grad-CAM primarily reflects convolutional neural network-derived activations and does not fully capture the contribution of transformer-based contextual fusion. Similarly, *t*-SNE visualises feature-space organisation but does not establish causal or clinically interpretable reasoning. Therefore, the explainability results should be viewed as supportive qualitative analyses rather than complete explanations of model decision-making.

Despite being evaluated on an adult cardiac MRI dataset, the proposed framework addresses methodological challenges that are also relevant to paediatric cardiac imaging, including robust feature extraction, multi-scale representation learning, and explainable decision support. These capabilities may be valuable for future AI-assisted analysis of paediatric and congenital heart disease imaging, where diagnostic accuracy and interpretability are particularly important. The comparative experiments indicate that the performance improvement is not solely attributable to increased model complexity. Individual convolutional neural network and transformer backbones achieved lower classification accuracy than the proposed hybrid framework, while the removal of preprocessing or feature-fusion components resulted in measurable performance degradation. These findings support the contribution of each module within the CardioFusion-XAI architecture.

Data and code availability

The Sunnybrook Cardiac Dataset used in this study is publicly available. The source code, trained model weights, and patient-level data split files will be made publicly available upon acceptance of the manuscript to facilitate reproducibility and future research.

Conclusion

This study proposed CardioFusion-XAI, an explainable hybrid deep learning framework for four-class cardiac MRI classification using the Sunnybrook Cardiac Dataset. The framework integrates WaveCLAHE preprocessing, convolutional neural network-based feature extraction, Vision Transformer-based feature fusion, and explainable AI techniques to improve classification performance and interpretability. However, the proposed framework was evaluated only on an adult cardiac MRI dataset and was not validated on paediatric or congenital heart disease populations. Therefore, its applicability to young patients remains unconfirmed and requires further investigation using clinically relevant paediatric cardiac datasets.

Experimental results demonstrated strong classification capability, achieving an accuracy of 98.89% and an F1-score of 98.90%. The integration of multi-scale feature learning and transformer-based contextual modelling contributed to improved discrimination among heart failure without infarction, heart failure, hypertrophy, and normal cardiac categories. However, these findings were obtained from a relatively small single-centre dataset, and external

Table 4. Comparison table for proposed work

Model	Accuracy (%)	Sensitivity (%)	Precision (%)	F1-Score (%)	Specificity (%)
Convolutional neural network + LSTM	95.5	95.2	95.6	95.4	97.0
VGG16	96.4	96.1	96.6	96.4	97.7
2D convolutional neural network	94.9	94.5	95.0	94.8	96.5
Proposed CardioFusion-XAI	98.89	98.89	98.93	98.90	99.63

https://doi.org/10.1017/S1047951126123671 Published online by Cambridge University Press

validation was not available. Therefore, the reported performance should be interpreted as promising rather than conclusive, and further evaluation on larger independent cohorts is necessary before broader clinical adoption can be considered.

Nevertheless, the study was evaluated on a single dataset containing four cardiac condition classes and a limited number of subjects. Therefore, the findings should not be interpreted as evidence of a comprehensive cardiovascular disease diagnosis or immediate clinical deployment. Future work will focus on validation using larger multi-centre datasets, additional cardiac pathologies, multimodal imaging data, and privacy-preserving collaborative learning strategies.

Supplementary material. The supplementary material for this article is available at https://doi.org/10.1017/S1047951126123671.

References

1. Guleria P, Naga Srinivasu P, Ahmed S, Almusallam N, Alarfaj FK. XAI framework for cardiovascular disease prediction using classification techniques. Electronics 2022; 11(24): 4086.
2. Panis V, Donal E. Imaging techniques for cardiac function. Appl Sci 2021; 11(22): 10549.
3. Backhaus SJ, Schulz A, Lange T, et al. Real-time cardiovascular magnetic resonance imaging for non-invasive characterisation of heart failure with preserved ejection fraction: outcomes of the HFpEF stress trial. Clin Res Cardiol 2024; 113(3): 496–508.
4. Singh RP, Singh P, Kler TS, Sharma SM. Diagnostic value of cardiac magnetic resonance imaging (CMRI) in the evaluation of arrhythmia and cardiomyopathy. Cureus 2024; 16(10): e72675.
5. Glessgen C, Crowe LA, Wetzl J, et al. Automated vs manual cardiac MRI planning: a single-centre prospective evaluation of reliability and scan times. Eur Radiol 2025; 35(7):3927–3936.
6. Park D, Kim J, Ruan X, Chow W, Baskaran L.. CycleGAN-based image-to-image translation for synthetic contrast enhancement in non-contrast cardiac CT: a ViT-CNN hybrid deep learning approach. BMC Medical Informatics and Decision Making.2026;26:189.
7. Muthulakshmi M, Venkatesan K, Prasanalakshmi B, Bahiyah RS, Divya V. An intelligent ensemble EfficientNet prediction system for interpretations of cardiac magnetic resonance images in heart failure severity diagnosis. Intell-Based Med 2025; 11: 100218.
8. Jie X, Chen J, Lin Z. Hybrid transformer-CNN for multi-modal whole-heart segmentation with MAE pretraining. In: Proceedings of the 2025 6th International Symposium on Artificial Intelligence for Medical Sciences (ISAIMS), 2025, pp. 587–592.
9. Din S, Qaraqe M, Mourad O, Qaraqe K, Serpedin E. ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial–temporal and long-range dependency features. Artif Intell Med 2024; 150: 102818.
10. Abdollahi J, Nouri-Moghaddam B, Mohajer A, Mirzaei A, Derakhshanfard N. Bridging global semantic representation and inference efficiency via hybrid knowledge distillation for clinically deployable cardiac MRI classification. Intelligence-Based Medicine 2026; 14: 100366.
11. Bizimana PC, Zhang Z, Hounye AH, Asim M, Hammad M, El-Latif AAA. Automated heart disease prediction using an improved explainable learning-based technique. Neural Comput Appl 2024; 36(26): 16289–16318.
12. Marey A, Arjmand P, Alerab ADS, et al. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt J Radiol Nuc Med 2024; 55(1): 183.
13. Sadr H, Salari A, Ashoobi MT, Nazari M. Cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models. Eur J Med Res 2024; 29(1): 455.
14. Baghdadi NA, Farghaly Abdelaliem SM, Malki A, Gad I, Ewis A, Atlam E. Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. J Big Data 2023; 10(1): 144.
15. Ogunpola A, Saeed F, Basurra S, Albarrak AM, Qasem SN. Machine learning-based predictive models for the detection of cardiovascular diseases. Diagnostics 2024; 14(2): 144.
16. Liu X, Xu Y, Xu H, et al. Advancing interpretable cardiac disease diagnosis via a transformer-convolutional hybrid network on electrocardiograms. Eng Appl Artif Intel 2025; 152: 110675.
17. Cockrum J, Nakashima M, Ammoury C, et al. Leveraging a vision transformer model to improve diagnostic accuracy of cardiac amyloidosis with cardiac magnetic resonance. Cardiovasc Imaging 2025; 18(3): 278–290.
18. Telangore H, Azad V, Sharma M, Bhurane A, San Tan R, Acharya UR. Early prediction of sudden cardiac death using multimodal fusion of ECG features extracted from Hilbert–Huang and wavelet transforms with explainable vision transformer and CNN models. Comput Methods Programs Biomed 2024; 257: 108455.
19. Mirzaee Moghaddam Kasmaee A, Ataei A, Moravvej SV, et al. ELRL-MD: a deep learning approach for myocarditis diagnosis using cardiac magnetic resonance images with ensemble and reinforcement learning integration. Physiol Meas 2024; 45(5): 055011.
20. Algolaib N, Abdulshahed A. Exploring the potential of using the discrete wavelet transform (DWT) for heart sound analysis. Int J Eng Inform Technol (IJEIT) 2024; 12(1): 293–300.
21. Camargo SS, Figueiredo E, Coelho RC. Impact of preprocessing cardiac magnetic resonance images for enhancing the performance of deep learning to segment myocardium. Res Biomed Eng 2026; 42(2): 26.
22. Anand R, Lakshmi SV, Pandey D, Pandey BK. An enhanced ResNet-50 deep learning model for arrhythmia detection using electrocardiogram biomedical indicators. Evol Syst 2024; 15(1): 83–97.
23. Jacob AJ, Borgohain I, Chitiboi T, Sharma P, Comaniciu D, Rueckert D. Towards a cardiovascular magnetic resonance foundation model for multi-task cardiac image analysis. J Cardiov Magn Reson 2025; 27(2): 101967.
24. Gangadharaiah SH, Kavitha KS, Shafeeq A, Karanth S. Modified U-Net Framework for left ventricle and myocardium segmentation and visual interpretation using Grad-CAM. Syst Soft Comput 2025; 7: 200426.
25. Ma L, Chen R, Ge W, et al. AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women. Exp Biol Med 2025; 250: 10389.
26. Gurusubramani S, Latha B. Deep convolutional generative adversarial network for improved cardiac image classification in heart disease diagnosis. J Imaging Inform Med 2024; 38(4): 2146.
27. Mohonta SC, Motin MA, Kumar DK. Electrocardiogram based arrhythmia classification using wavelet transform with deep learning model. Sens Biosensing Res 2022; 37: 100502.

https://doi.org/10.1017/S1047951126123671 Published online by Cambridge University Press