A Hybrid Deep Learning Heart Disease Prediction Framework Utilizing Multi-Modal Medical Imaging and Novel Feature Fusion Techniques

P. Archana
Department of Computer Science and Engineering, BGS Institute of Technology, Adichunchanagiri University, BG Nagara, Mandya, India
archanahavishsurya@gmail.com (corresponding author)

S. V. Shashikala
Department of Computer Science and Engineering, BGS Institute of Technology, Adichunchanagiri University, BG Nagara, Mandya, India
shashisv7@gmail.com

Received: 6 July 2025 | Revised: 16 August 2025 and 23 August 2025 | Accepted: 2 September 2025

 DOI: https://doi.org/10.48084/etasr.13204

ABSTRACT

Heart diseases require advanced diagnostic techniques for early and accurate detection. This paper combines multi-modal data sources, such as CT images, MRI scans, and ECG signals to provide a hybrid deep learning architecture for accurate cardiac disease identification. The system uses specific feature extraction methods, such as 3D-UNet for 3D MRI and CT images and Temporal Convolutional Graph Neural Networks (TC-GNN) for ECG, and then uses genetic algorithms to optimize the features. Autoencoders, which are 1D for ECG and 3D for MRI and CT, are employed for non-linear dimensionality reduction in order to handle the high dimensionality of fused information. A Convolutional Neural Network (CNN) processes the fused compact features for the final classification. The proposed model achieved a 97.1% accuracy, outperforming known models. Accuracy, recall, F1-score, and ROC-AUC scores support its generalizability and robustness. This multi-modal and feature-aware approach significantly increases classification accuracy, reduces false positives and false negatives, and provides a scalable clinical decision support solution for cardiovascular diagnostics.

Keywords-hybrid deep learning; multi-modal imaging; feature fusion; heart disease diagnosis; CNN

I. INTRODUCTION

According to estimates from the World Health Organization, cardiovascular diseases (CVDs) are a global leading cause of death [1]. Early and accurate diagnosis is essential for the management and treatment of heart diseases, and it frequently results in better patient outcomes and reduced medical costs. Early and precise diagnosis is essential since CVDs account for 31% of total yearly deaths [2]. Traditional diagnostic methods include blood biomarker analysis, MRI, CT, and ECGs. However, these techniques sometimes require manual interpretation by trained medical personnel, which can be time-consuming and prone to inter-observer variability. The advances in processing power and Machine Learning (ML) techniques [3] along with the growing availability of large-scale medical datasets have enabled improvements in medical diagnostics [4].

ML models are trained on statistical and clinical datasets for heart disease prediction. Several ML classifiers, including Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Trees (DTs), have demonstrated dependable performance [5, 6]. The application of Deep Learning (DL) in medical imaging has seen substantial growth in recent years. Authors in [7] provided a comprehensive survey of DL techniques in medical image analysis, highlighting the dominance of Convolutional Neural Networks (CNNs) in tasks such as classification, detection, and segmentation. In cardiology, CNNs have been successfully used to detect arrhythmias from ECG signals, identify abnormalities in echocardiograms, and segment heart structures in MRI scans. LSTM networks, on the other hand, have shown promise in analyzing sequential medical data. Authors in [8] used LSTM models to detect cardiac arrhythmias from single-lead ECG signals, achieving performance comparable to expert

cardiologists. Despite their success, both CNN and LSTM models are often limited to single-modality inputs.

Authors in [9] showed that the accuracy of integrating ECG with clinical and demographic variables in a CNN-based framework was higher than 90%. CNNs have shown remarkable performance in the field of MRI-based cardiac prediction. A CNN model that detected heart problems using MRI with 95% accuracy was published in [10]. In order to quickly segment 3D MRI volumes, authors in [11] optimized a fully CNN with batch normalization, accomplishing segmentation in less than 23 s. CNNs with ADAM activation were used in [12] to measure left ventricular volume. Authors in [13] achieved over 92% accuracy by fusing MRI and ECG utilizing 3D-U-Net and sequential CNN with attention. Authors in [14] showcased a model that deep-dived into an ECG dataset. Authors in [15] achieved up to 96% accuracy without explicit dimensionality reduction. These findings demonstrate the increasing applicability of multi-modal frameworks powered by DL for extremely precise cardiac disease prediction. Nevertheless, the majority of the current models only consider unimodal data, which restricts their capacity to adequately represent the variety of heart disease presentations. This paper suggests a hybrid DL architecture that uses a novel feature fusion technique to combine multi-modal data, including cardiac MRI images, CT scan images, and ECG signals.

A. Problem Statement

The complexity of combining various data, including ECG, MRI, and CT, makes the accurate diagnosis of heart diseases a difficult task. Current models typically exhibit poor generalization and diagnostic accuracy in real-world clinical scenarios due to their lack of resilience, disregard for multi-modal fusion, or failure to preserve important temporal and spatial information.

B. Paper Contributions

The main contribution of the current work are:

- Multi-Modal Data Integration: To improve diagnostic performance, a single diagnostic framework was developed that combines ECG, MRI, and CT data to capture a variety of cardiac problems.
- Dimensionality Reduction Using Autoencoders: Non-linear dimensionality reduction was carried out while maintaining discriminative information by using modality-specific autoencoders, such as 1D for ECG and 3D for MRI/CT.
- Improved Feature Extraction: 3D-UNet with Genetic Algorithms for spatial feature selection in MRI/CT and TC-GNN for temporal ECG patterns were integrated to improve feature quality.
- Superior Classification Performance: Achieved state-of-the-art accuracy (97.1%) with CNN-based classification of fused features. The model was validated using cross-validation, ROC analysis, and comparison benchmarks.

II. THE PROPOSED MODEL

This section offers a thorough description of the proposed hybrid DL model along with functional architecture schematic illustrations. Figure 1 depicts the general layout and operation of the suggested system.

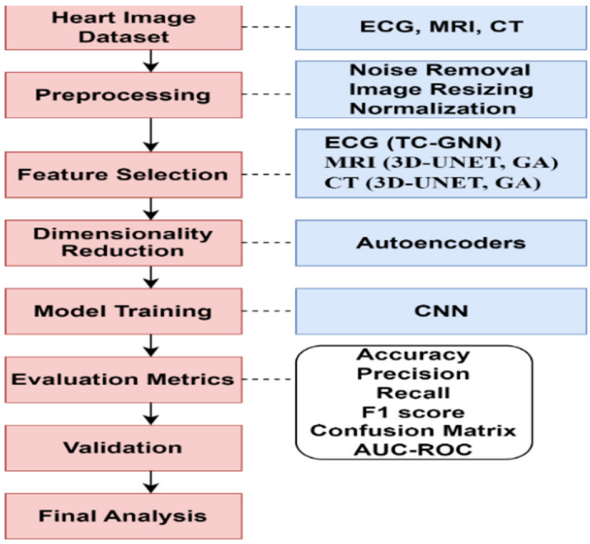

Fig. 1. Workflow of the proposed model.

A. Dataset Description

This work employs a hybrid multi-modal dataset that comprises CT images [16], MRI scans [17], and ECG signals [18] to record different facets of heart structure and function. The ECG dataset consists of 3,951 samples that demonstrate the electrical activity of the heart and are crucial for detecting arrhythmias and myocardial infarctions. The MRI dataset, which has 63,400 samples, provides thorough structural and functional imaging to identify abnormalities such as hypertrophy and valve dysfunctions. The CT scan collection, which includes 5,896 samples, offers cross-sectional imaging of the anatomy of the heart, highlighting conditions such as coronary artery disease and calcifications.

B. Data Preprocessing and Balancing

1) ECG Dataset Preprocessing and Balancing

Bandpass filters and wavelet transforms were used for noise removal in raw ECG signals to remove artifacts and improve signal quality. Using the transfer function, the bandpass filter eliminates noise outside the frequency range of interest. Wavelet transformations break down the signal into details and approximations, eliminating high-frequency noise while keeping the important information. The ECG signals were converted into spectrogram or waveform images and resized to 224×224 pixels for compatibility with the CNN models. This uniformity enables the use of pre-trained architectures for heart attack detection. The pixel values were scaled in the range of [0,1] using min-max normalization. This stabilizes the training process by reducing the impact of varying pixel intensities, ensuring faster convergence and improved performance. Finally, a Synthetic Minority Oversampling Technique

(SMOTE) is employed to address class imbalances, generating synthetic samples for underrepresented categories such as rare heart attack events in ECG data.

2) MRI Dataset Preprocessing and Balancing

The MRI volumes undergo the following transformations: The volumes are resized to a uniform dimension of 128×128 to maintain consistency for batch training and an anisotropic diffusion filtering is applied to reduce noise while preserving edges, governed by the Perona-Malik equation:

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(\| \nabla I \|)\nabla I) \quad (1)$$

where $c(\cdot)$ is the diffusion coefficient controlling smoothing.

3) CT Scan Dataset Preprocessing and Balancing

CT scans are pre-processed to detect coronary blockages or structural changes associated with heart attacks. A low-pass filter is applied to suppress high-frequency noise. Also, anisotropic diffusion is applied that enhances edges while reducing noise:

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(x,y,t)\nabla I) \quad (2)$$

where $c(x,y,t)$ controls diffusion. The images are resized to 224×224 size. All the pixel values are normalized to [0,1] or [−1,1] to stabilize training. A Generative Adversarial Network (GAN) is used to synthesize additional CT images of minority classes, such as rare patterns of coronary artery occlusion associated with heart attacks.

C. Feature Selection and Segmentation

Feature selection is the process of identifying the most relevant features from the dataset to improve model accuracy and efficiency.

1) ECG Feature Selection

Rich temporal information is extracted from ECG signals using a Temporal Convolutional Graph Neural Network (TCGNN), which captures both local and long-range relationships. After extracting spatial representations using 1D convolutional layers, the model models inter-time-step dependencies using temporal graph convolution layers. Finally, informative time steps are weighed using an attention method. This makes the ECG data more discriminative for classification tasks by effectively reducing duplicate characteristics while maintaining crucial temporal information. The process is highlighted in Algorithm 1.

Algorithm 1:	ECG feature extraction using TCGNN
Input:	$X_{ECG} \in R^{(T*F)}$: ECG data with T time steps and F features
Output:	Z_{ECG}: Reduced ECG feature vector
Steps:	
1	Initialize TCGNN with weights θ
2	$H_{ECG} \leftarrow X_{ECG}$
3	Apply 1D Conv: $H_{ECG} \leftarrow Conv1D(H_{ECG}, \theta_{conv})$ $H_{ECG} \leftarrow ReLU(H_{ECG})$ $H_{ECG} \leftarrow MaxPool1D(H_{ECG})$
4	Construct Temporal Graph G
5	Apply GraphConv: $H_G \leftarrow GraphConv(H_{ECG}, G, \theta_{gcn})$ $H_G \leftarrow ReLU(H_G)$
6	Apply Attention: $A \leftarrow Attention(H_G)$ $Z_{ECG} \leftarrow \sum_{i=1}^{n} A_i . H_G[i]$

2) MRI and CT Segmentation Using 3D-UNet and Feature Selection

For volumetric imaging modalities like CT and MRI, Regions of Interest (ROIs) are extracted and spatial segmentation is carried out using 3D-UNet. This eliminates extraneous information and background noise, allowing the model to concentrate on structures with therapeutic value. 3D-UNet's encoder-decoder architecture efficiently records spatial hierarchies, allowing for the accurate localization of data required for cardiac condition categorization. Algorithm 2 shows the procedure.

Algorithm 2:	MRI/CT Segmentation using 3D-UNet
Input:	$X_{MRI}, X_{CT} \in R^{(H*W*D*C)}$: Volumetric data
Output:	ROI_{MRI}, ROI_{CT}: Segmented regions of interest
Steps:	
1	$S_{MRI} \leftarrow 3D - UNet(X_{MRI})$
2	$S_{CT} \leftarrow 3D - UNet(X_{CT})$
3	$ROI_{MRI} \leftarrow Segment(S_{MRI})$
4	$ROI_{CT} \leftarrow Segment(S_{CT})$

The most relevant information is extracted from segmented MRI and CT images using a Genetic Algorithm (GA). The GA optimizes feature subsets based on the classification accuracy of a validation set. It identifies feature sets that enhance model performance while reducing dimensionality and redundancy by simulating natural evolution through crossover, mutation, and selection. The process is highlighted in Algorithm 3.

Algorithm 3:	Feature Selection using GA
Input:	ROI_{MRI}, ROI_{CT}: Features from Segmented regions of interest
Output:	Z_{MRI}, Z_{CT}: Optimized features
Steps:	
1	Encode features as chromosomes
2	Initialize population
3	Evaluate fitness (classification accuracy)
4	Apply selection, crossover, mutation
5	$Z_{MRI} \leftarrow SelectedFeatures(ROI_{MRI})$
6	$Z_{CT} \leftarrow SelectedFeatures(ROI_{CT})$

D. Dimensionality Reduction Using Autoencoders

Reducing dimensionality is essential for enhancing model effectiveness and preventing overfitting. While 3D autoencoders are selected for MRI and CT due to their volumetric character, a 1D autoencoder is applied to the ECG data. While reconstructing input data, these architectures preserve important spatial and temporal features that are necessary for precise classification by learning compact, useful latent representations. The process is highlighted in Algorithm 4.

Algorithm 4:	Autoencoder-Based Dimensionality Reduction
Input:	Z_{ECG}, Z_{MRI}, Z_{CT}: Selected features
Output:	$Z'_{ECG}, Z'_{MRI}, Z'_{CT}$: Reduced features
Steps:	
1	$Z'_{ECG} \leftarrow 1D - Autoencoder(Z_{ECG})$
2	$Z'_{MRI} \leftarrow 3D - Autoencoder(Z_{MRI})$
3	$Z'_{CT} \leftarrow 3D - Autoencoder(Z_{CT})$

E. Fusion and Classification Using CNNs

The reduced characteristics from ECG, MRI, and CT are concatenated and run through a CNN for final classification in order to take advantage of the complementary nature of various modalities. Through the integration of temporal and spatial inputs, this fusion method improves decision-making. Dense layers are utilized to produce class probabilities after convolutional layers further isolate significant patterns from the fused feature map. This stage, which is the foundation of the classification pipeline, guarantees precise cardiac disease prediction using multimodal data. The process is highlighted in Algorithm 5.

Algorithm 5:	Fusion and Classification using CNN
Input:	$Z'_{ECG}, Z'_{MRI}, Z'_{CT}$: Reduced features
Output:	Y: Class probabilities
Steps:	
1	$Z \leftarrow Concat(Z'_{ECG}, Z'_{MRI}, Z'_{CT})$
2	$H \leftarrow Conv2D(Z, \theta_{conv2D})$
3	$H \leftarrow ReLU(H)$
4	$H \leftarrow MaxPool2D(H)$
5	$Flatten\ H$
6	$Y \leftarrow Dense(H, \theta_{dense})$

F. Model Evaluation and Validation

Evaluation of classification models was carried out by using the prediction scores such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The evaluation metrics are defined by:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{Total Samples}} * 100 \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} * 100 \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} * 100 \quad (5)$$

$$\text{F1} - \text{score} = \frac{\text{2.Precision} \times \text{Recall}}{\text{Precision+Recall}} * 100 \quad (6)$$

$$\text{Dice Score} = \frac{\text{2.TP}}{\text{2.TP+FP+FN}} * 100 \quad (7)$$

$$\text{True Positive Rate} = \frac{\text{TP}}{\text{TP+FN}} * 100 \quad (8)$$

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{FP+TN}} * 100 \quad (9)$$

The working of model was evaluated by k-fold cross validation with K = 10.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The model was implemented on a system with 3080 GPU, 16GB RAM, and an Intel Core i9 CPU. The dataset was split into training (70%), validation (15%), and test (15%) sets. The hyperparameter values were: Epochs: 50, Batch Size: 32, Learning Rate: 0.0001, Optimizer: Adam, Loss Function: Binary Cross-Entropy

A. Impact of Autoencoder-Basumased Dimensionality Reduction

Applying autoencoder-based dimensionality reduction greatly improves classification metrics, according to the performance evaluation. Accuracy for ECG data increased from 92.13% to 96.41%, while precision, recall, F1 score also increased in tandem along with a significant decrease in false positive rate. Similar patterns were seen in the MRI and CT modalities, where autoencoders improved overall performance by 3–4%. Results were further enhanced by the fusion of modalities, which achieved the greatest accuracy of 97.13% with autoencoder, demonstrating its efficacy in maintaining discriminative characteristics. The overall performance of the proposed work is outlined in Table I.

TABLE I. PERFORMANCE ANALYSIS OF THE PROPOSED MODEL

Mod.	DimRed	Acc.	Prec.	Rec.	F1 Score	Dice Score	TPR	FPR
ECG	No	92.1	92.0	92.2	92.1	92.1	92.2	7.9
ECG	Yes	96.4	96.0	96.7	96.3	96.3	96.7	3.9
MRI	No	88.9	88.9	88.7	88.8	88.8	88.7	11.0
MRI	Yes	92.5	92.9	92.1	92.5	92.5	92.1	7.1
CT	No	90.4	90.4	90.3	90.3	90.3	90.3	9.5
CT	Yes	93.6	93.5	93.6	93.6	93.6	93.6	6.4
Fus.	No	94.3	94.5	95.0	94.8	94.8	95.0	5.4
Fus.	Yes	97.1	97.3	97.6	97.4	97.4	97.6	2.7

Comparison of model accuracy with and without autoencoder-based dimensionality reduction in ECG, MRI, CT, and fused modalities is shown in Figure 2. The findings clearly demonstrate that the use of autoencoders increases accuracy across all modalities. ECG, MRI, and CT confusion matrices with and without dimensionality reduction based on autoencoders are shown in Figure 3. It is evident that autoencoder integration lowers false negatives as well as false positives. For example, false positives decreased from 153 to 63 and false negatives decreased from 158 to 79 in the ECG modality. Similar patterns were also seen in CT and MRI, with CT false negatives declining from 282 to 189 and MRI false

positives decreasing from 3,507 to 2,239. These findings demonstrate that by improving feature representation prior to classification, autoencoders considerably increase model sensitivity and specificity.

ROC curves with and without autoencoder-based dimensionality reduction for the diagnosis of cardiac disease in ECG, MRI, CT, and fusion modalities is presented in Figure 4. With autoencoder integration, the AUC increased for all modalities. Fusion from 0.96 to 0.98, MRI from 0.90 to 0.94, CT from 0.92 to 0.95, and ECG from 0.93 to 0.97, all showed improvements. These findings demonstrate the improved discriminatory capacity and decreased misclassification. Fusion with the autoencoder produced the highest AUC, indicating its superior ability to distinguish between favorable and unfavorable circumstances. The clear separation from the diagonal (random guess) line further validates the model's robustness and prediction accuracy.

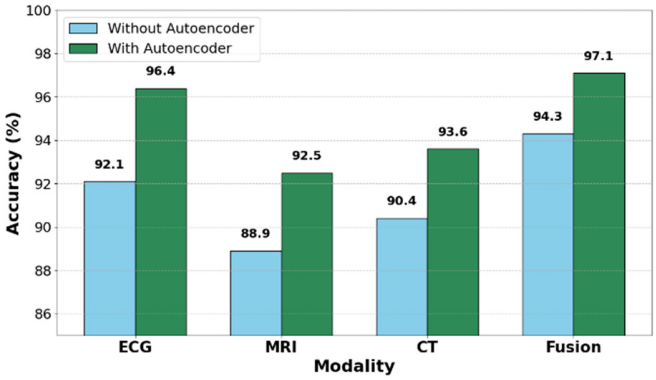

Fig. 2. Accuracy with and without dimensionality reduction.

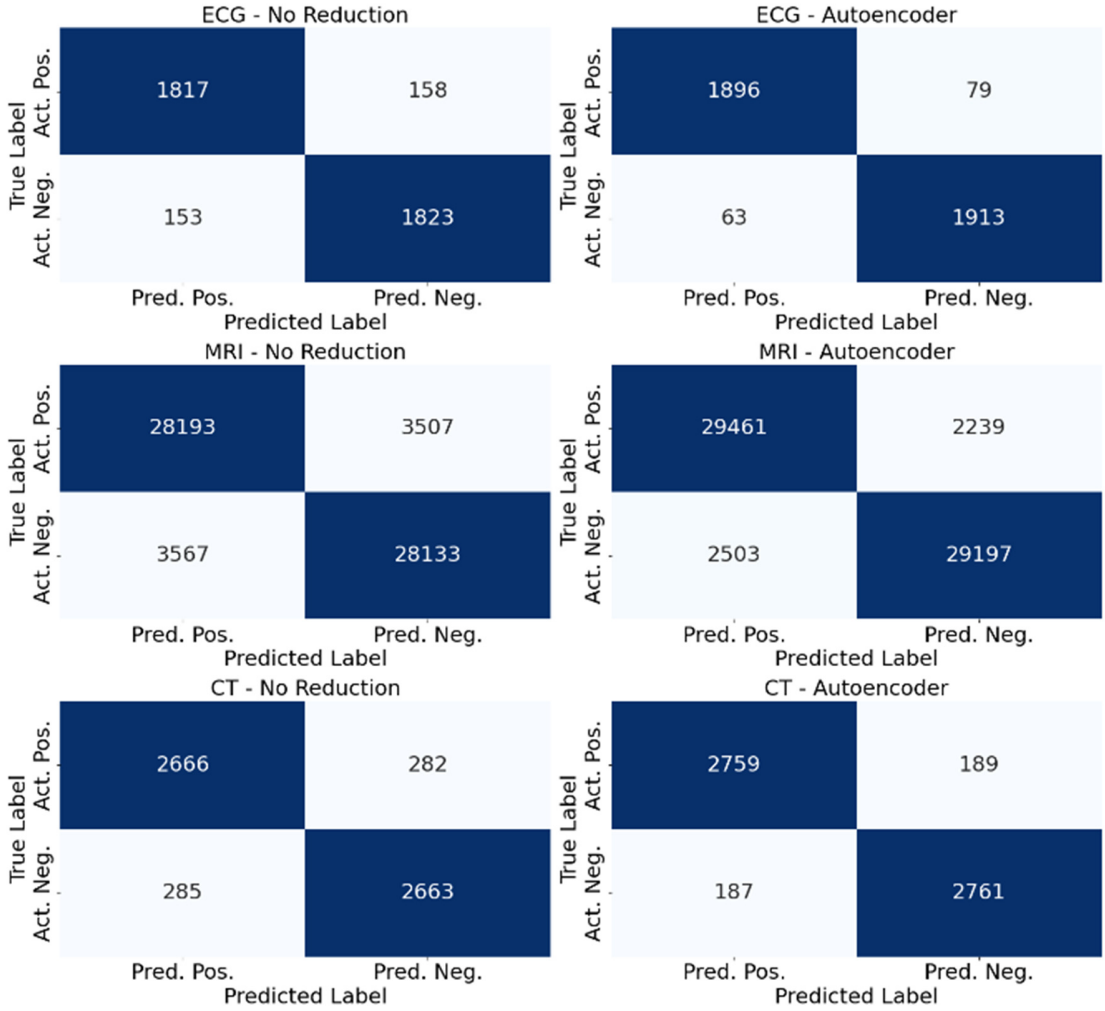

Fig. 3. Confusion matrices of all data modalities.

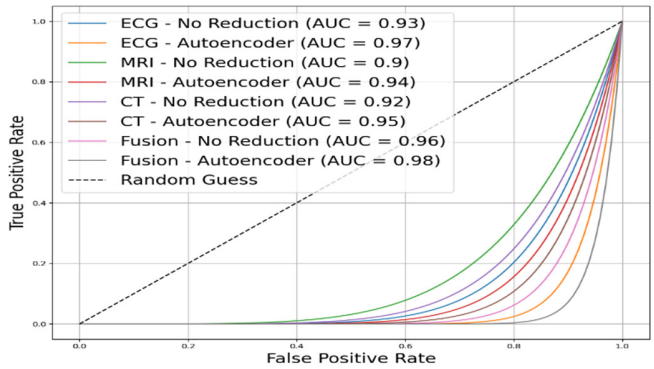

Fig. 4. Representation of ROC-AUC of all data modalities.

B. Cross-Validation Analysis

The proposed model was evaluated using 10-fold cross-validation spanning the ECG, MRI, and CT modalities as well as their fusion. The data presented in Table II, shows the consistent performance, with ECG having the greatest individual modality mean accuracy of 95.90% (±0.77), followed by CT (92.84% ±0.83) and MRI (91.92% ±0.82). The fusion technique outperformed individual modalities with a mean accuracy of 96.86% and a reduced standard deviation (±0.60), indicating consistent and dependable categorization, enhanced generalization, and reduced variability between folds.

TABLE II. CROSS VALIDATION ACCURACY

Fold	ECG acc.	MRI acc.	CT acc.	Fusion acc.
Fold 1	94.10	90.30	91.20	95.50
Fold 2	95.20	91.00	91.90	96.30
Fold 3	96.00	92.10	93.10	96.90
Fold 4	96.30	91.50	92.60	97.20
Fold 5	95.70	92.40	93.20	96.70
Fold 6	96.50	93.00	93.90	97.40
Fold 7	96.10	91.90	92.80	97.10
Fold 8	96.80	92.70	93.60	97.60
Fold 9	95.60	91.80	92.50	96.80
Fold 10	96.40	92.50	93.60	97.10
Mean	95.90	91.92	92.84	96.86
Std Dev	0.77	0.82	0.83	0.60

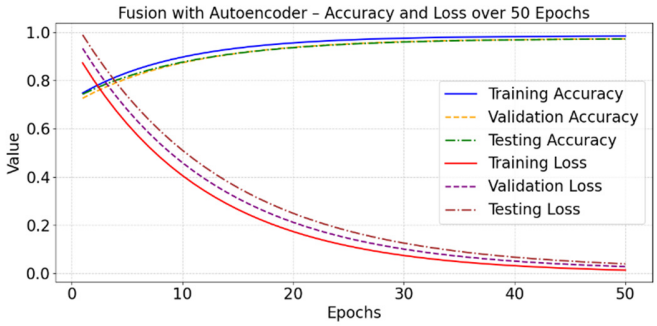

Fig. 5. Representation of training, validation, and testing accuracy and loss of fusion for 50 epochs.

The fusion model with an autoencoder performs better than individual ECG, MRI, and CT models as can be seen in Figure 5. The training, validation, and testing accuracies show improved learning through multi-modal feature integration, progressively increasing and converging to 98.5% by the 50th epoch. As a result, the loss numbers decrease smoothly, with validation/testing losses staying closely aligned and training loss falling below 0.05. Excellent generalization and absence of overfitting are suggested by this consistent behavior across all datasets.

C. Comparison with Baseline and Existing Models

The comparison of heart disease categorization models based on various modalities and approaches is shown in Table III. The maximum accuracy of 97.1% is attained by the proposed model, which uses autoencoder-based dimensionality reduction and combines ECG, MRI, and CT data using TC-GNN and 3D-UNet with GA. The model proposed in [13], on the other hand, achieves 92.6% using 3D-U-Net for MRI and sequential CNN for ECG without dimensionality reduction. The model from [14] shows improved performance to 93.7% by using contrastive learning on ECG and MRI and masked autoencoders. Using CNN-GRU on fused ECG signals, the model from [15] obtains 96.7%, although it does not reduce dimensionality.

TABLE III. ACCURACY RESULT COMPARISON

Paper	Modality	Feature extraction	Dimensionality reduction	Accuracy (%)
[13]	MRI, ECG	3D-U-Net, CNN	NO	92.6
[14]	ECG, MRI	Contrastive Loss	Masked Autoencoder	93.7
[15]	ECG	CNN-GRU	No	96.7
Proposed	Fusion (ECG, MRI, CT)	TC-GNN, 3D-UNet + GA	Autoencoder	97.1

IV. CONCLUSION

The proposed work presents a hybrid deep learning architecture that combines CT, MRI, and ECG data for precise and trustworthy heart disease categorization. The proposed method uses TC-GNN for temporal ECG pattern modeling, autoencoder-based dimensionality reduction, and 3D-UNet with genetic algorithms for imaging feature extraction in order to fill the knowledge gap of limited multi-modal fusion approaches in cardiovascular diagnostics. This model reduces redundancy while preserving important spatial-temporal information.

In comparison to unimodal techniques and current baseline models, the proposed multi-modal fusion strategy achieved state-of-the-art accuracy of 97.1% with notable gains in precision, recall, and AUC. The framework's robustness and generalizability are demonstrated by the decrease in false positives and false negatives along with a consistent performance confirmed by 10-fold cross-validation. These results demonstrate the study's originality and contribution to the integration of complementary data modalities for better clinical decision assistance.

From the standpoint of clinical relevance, the framework has a great potential for practical implementation as a decision-support instrument in medical settings. To facilitate seamless integration into healthcare workflows, however, real-world

issues, including interoperability with electronic health record (EHR) systems, data privacy compliance (HIPAA, GDPR), and the requirement for validation across various clinical datasets, must be properly handled. To close the gap between research and practical implementation, cooperation between clinical specialists and healthcare organizations will be crucial.

To enable scalable, privacy-compliant, and clinician-friendly implementation of the framework, future research will concentrate on real-time model optimization for edge devices, wider validation using multi-center clinical datasets, and connection with hospital information systems.

REFERENCES

[1] S. C. W. Tan, B.-B. Zheng, M.-L. Tang, H. Chu, Y.-T. Zhao, and C. Weng, "Global Burden of Cardiovascular Diseases and its Risk Factors, 1990-2021: A Systematic Analysis for the Global Burden of Disease Study 2021," *QJM: monthly journal of the Association of Physicians*, vol. 118, no. 6, pp. 411–422, Jun. 2025, https://doi.org/10.1093/qjmed/hcaf022.

[2] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis," *Journal of Big Data*, vol. 10, no. 1, Sep. 2023, Art. no. 144, https://doi.org/10.1186/s40537-023-00817-1.

[3] A. P and S. S. V, "Improving the Efficiency of Predicting the Heart Diseases Using Optimized Feature Selection and Ensemble Machine Learning Techniques," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 21, no. 09, pp. 27–42, Jul. 2025, https://doi.org/10.3991/ijoe.v21i09.55425.

[4] N. Bora, S. Gutta, and A. Hadaegh, "Using Machine Learning to Predict Heart Disease," *WSEAS Transactions on Biology and Biomedicine*, vol. 19, pp. 1–9, Jan. 2022, https://doi.org/10.37394/23208.2022.19.1.

[5] A. Bahuguna, M. Gupta, and R. Kumar, "Statistical Analysis and Prediction of Heart Disease Using Machine Learning," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, Jul. 2023, pp. 1–5, https://doi.org/10.1109/ICCCNT56998.2023.10308166.

[6] M. Rahardi, B. P. Asadduloh, A. Aminuddin, F. F. Abdulloh, I. Saifudin, and F. P. Kusumawijaya, "Optimizing Machine Learning Models for Class Imbalance in Heart Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23599–23604, Jun. 2025, https://doi.org/10.48084/etasr.10407.

[7] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, Jan. 2019, https://doi.org/10.1038/s41591-018-0316-z.

[8] L. Kumar, C. Anitha, V. N. Ghodke, N. Nithya, V. A. Drave, and A. Farhana, "Deep Learning Based Healthcare Method for Effective Heart Disease Prediction," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 9, Oct. 2023, https://doi.org/10.4108/eetpht.9.4283.

[9] A. Sharma, R. Kumar, and V. Jaiswal, "Classification of Heart Disease from MRI Images Using Convolutional Neural Network," in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, Solan, India, Jul. 2021, pp. 358–363, https://doi.org/10.1109/ISPCC53510.2021.9609408.

[10] R. Kannan and V. Vasanthi, "Deep Learning Framework to Predict and Diagnose the Cardiac Diseases by Image Segmentation," in *Innovations in Computer Science and Engineering: Proceedings of 7th ICICSE*, H. S. Saini, R. Sayal, R. Buyya, and G. Aliseri, Eds. Singapore: Springer, 2020, pp. 215–224, https://doi.org/10.1007/978-981-15-2043-3_26.

[11] M. Zakariah and K. AlShalfan, "Cardiovascular Disease Detection Using MRI Data with Deep Learning Approach," International Journal of Computer and Electrical Engineering, vol. 12, no. 2, pp. 72–82, 2020, https://doi.org/10.17706/ijcee.2020.12.2.72-82.

[12] S. Ahuja, D. Shrimankar, and A. Durge, "Design of an Iterative Method for Deep Multimodal Feature Fusion in Heart Disease Diagnostics Utilizing Explainable AI," in *Proceedings of the 1st International Conference on Explainable AI for Neural and Symbolic Methods*, Sep. 2025, pp. 87–95, https://doi.org/10.5220/0012899400003886.

[13] D. Alsekait *et al.*, "Heart-Net: A Multi-Modal Deep Learning Approach for Diagnosing Cardiovascular Diseases," *Computers, Materials & Continua*, vol. 80, no. 3, pp. 3967–3990, 2024, https://doi.org/10.32604/cmc.2024.054591.

[14] F. Girlanda, O. Demler, B. Menze, and N. Davoudi, "Enhancing Cardiovascular Disease Prediction through Multi-Modal Self-Supervised Learning." arXiv, Nov. 08, 2024, https://doi.org/10.48550/arXiv.2411.05900.

[15] R. Panchal, S. Tiwari, and S. Agarwal, "Multimodal image fusion on ECG signals for congestive heart failure classification," *Multimedia Tools and Applications*, vol. 84, no. 10, pp. 8247–8259, Mar. 2025, https://doi.org/10.1007/s11042-024-19052-8.

[16] N. Tomar, "CT Heart Dataset." 2021, [Online]. Available: https://www.kaggle.com/datasets/nikhilroxtomar/ct-heart-segmentation.

[17] D. Sharifrazi, "CAD Cardiac MRI Dataset," 2021. https://www.kaggle.com/datasets/danialsharifrazi/cad-cardiac-mri-dataset.

[18] E. Spiritos, "ECG Images Dataset of Cardiac Patients," 2024. https://www.kaggle.com/datasets/evilspirit05/ecg-analysis.