International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(18s) 2026 pp. 970-987
https://doi.org/10.70917/ijcisim-2026-4936
Article

An Interpretable LES-Integrated Hybrid RNN-CNN Framework for Multimodal Cardiovascular Disease Prediction

Indrapalli Swapna[1], Sasidhar Kothuru[2]

[1]Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Aziz Nagar, Hyderabad -500075, Telangana, India.
Email ID: swapna.indrapalli1012@gmail.com
[2]Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Aziz Nagar, Hyderabad -500075, Telangana, India. Email ID: sasidhar@klh.edu.in

Abstract: Coronary artery disease is one of the leading causes of death worldwide. There is a need for an efficient diagnostic system to detect CAD that can lead to effective clinical therapy. Existing deep learning models are mostly designed for either spatial imaging or temporal physiologic signal learning, but they have limited fusion ability and do not extract lesion-aware features. In response to these challenges, this study proposes an interpretable Local Energy Shape (LES)- integrated hybrid Recurrent Neural Network and Convolutional Neural Network (LES-RCNN) framework for cardiovascular disease prediction and classification using multimodal image and physiological datasets. The proposed framework implements LES-based feature extraction combined with recurrent learning and convolutional feature modeling to extract spatial, temporal, and lesion-based cardiovascular characteristics from CCTA, ECG image data, and structured CSV physiological records. Lesion localization is greatly enhanced by the structural representation provided by LES transformation. Also, the model's spatial and temporal feature learning is improved with a hybrid RNN-CNN architecture. Experimental assessment was carried out utilizing both CVD Atlas image datasets and UCI physiological datasets in binary and multi-label classification. The proposed 17-layer LES-RCNN architecture obtained 98.5% accuracy, 98.2% precision, 98.7% recall, and 98.4% F1-score. This was better than existing state-of-the-art cardiovascular disease prediction models. Through comparisons and convergence studies, it was confirmed that the learning was more stable, lost less, and had better generalization capacity. The results of the study reaffirm that the LES-based multimodal framework can provide an accurate, interpretable & computationally efficient diagnosis of early CVD & clinical decision support for clinical decision-making.
Keywords: cardiovascular disease, coronary artery disease, Local Energy Shape, LES-RCNN, deep learning, recurrent neural network, convolutional neural network, multimodal learning, ECG analysis, coronary CT angiography, lesion-aware feature extraction, cardiovascular disease prediction

1. Introduction

Cardiovascular diseases (CVDs), especially coronary artery disease (CAD), arrhythmia, myocardial infarction, and congestive heart failure, remain amongst the foremost causes of death and long-term disability worldwide [1]. The latest statistics on global health show that there are millions of deaths occurring every year due to delayed diagnosis and a lack of early-stage prediction of cardiovascular disorders [2]. The precise diagnosis of CAD is of great importance, as a clinical intervention at an early stage can reduce mortality and improve survival. Standard diagnostic modalities are Electrocardiography (ECG), Echocardiography, Stress testing, and Coronary CT Angiography (CCTA). They are quite popular in clinical practice. However, they are often highly expert-dependent and subjective. Furthermore, results are contaminated with noise, have limited sensitivity, and are subject to time delays for action [3]. The identification of big medical imaging datasets and physiological records is facilitating the use of artificial intelligence (AI) and deep learning to accomplish automatic diagnosis of cardiovascular disease [4].

In recent years, architectures like CNN, RNN, LSTM, Bi-LSTM, and transformers have shown outstanding performance in a variety of cardiovascular health care applications using deep learning [5]. CNN-based models can effectively extract spatial features from ECG images and CCTA scans. On the other hand, physiological signals can be efficiently captured by RNN and LSTM architectures with sequential and temporal dependencies [6]. Accurate disease classification using CNNs and recurrent architecture-tied patented deep learning frameworks is further improved by jointly learning spatial and temporal dependencies. Numerous studies have been conducted on explainable AI (XAI). Similarly, SHAP and attention techniques are used to improve cardiovascular prediction systems to gain clinical trust and transparency.

When you merge ECG, CCTA, biological CSN data, and clinical records, multimodal learning strategies improve reliability and robustness, according to the results in [7]. The success of transfer learning and transformer models improved feature generalization and multilabel classification. Despite the above advances, existing methods still face several challenges. These include high computational complexity, low interpretability, poor lesion-aware feature extraction, class imbalance, noisy medical data, and low model generalization capability across heterogeneous clinical datasets. Existing methods use either spatial imaging analysis or temporal ECG signal learning, but none effectively integrates both [8].

To overcome these limitations, the proposed study presents a Local Energy Shape-integrated hybrid RNN-CNN architecture for cardiovascular disease prediction and classification. The proposed framework benefits from large-eddy simulation of local structural and geometric features in medical images through recurrent and convolutional layers. The combination of explainable learning techniques and multi-modal fusion will improve classification performance, lesion localization ability, robustness, and clinical interpretability. As a proof of concept, this paper describes a reliable and efficient AI-driven model capable of diagnosing cardiovascular diseases early. It uses manually designed lesion-aware descriptors that leverage deep hierarchical feature learning.

1.1.Research Objectives

The main aim of this study is to build an explainable LES-integrated hybrid RNN-CNN framework for multimodal cardiovascular disease prediction. The proposed models cogently combine LES-based feature extraction with recurrent and convolutional network learning models and leverage cardiovascular image data and structured physiological CSV data for disease diagnosis. The objective of this study is to enhance feature learning, classification robustness, and lesion localization. In addition, the proposed framework will be very helpful for both binary and multi-label CVD classification with better accuracy.

1.2.Contributions

The proposed model presents a hybrid RNN-CNN framework for multimodal prediction and classification of cardiovascular disease, integrated with LES. The proposed model incorporates tapered estimators that account for lesions derived from LES, and a model combining recurrent and convolutional learning that captures both spatial and temporal features of the cardiovascular system. The framework unifies clinical image-based and physiological datasets in a joint multimodal learning architecture. Experimental analysis shows better performance, achieving 98.5% classification accuracy over existing state-of-the-art cardiovascular disease prediction models. The proposed LES-RCNN framework enhances lesion localization performance and convergence stability for intelligent cardiovascular lesion detection.

2. Related Works

Cardiovascular disease (CVD) is becoming one of the leading health problems. Examples include coronary artery disease (CAD), arrhythmia, and myocardial infarction. Recently, deep learning, along with various neural networks, has strengthened automated disease diagnosis and prediction from ECG, CCTA, and multimodal clinical datasets, further supported by artificial intelligence (AI) technologies over the past decade [9], in cardiovascular healthcare applications. The use of convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory (LSTM), transformers, and explainable artificial intelligence (XAI) techniques has increased significantly [10].

Al-Bairmani et al. (2026) [11] devised a SHAP-based joint CNN-Bi-LSTM-transformer strategy for multi-label arrhythmia detection using 12-lead ECG. Their architecture efficiently handled spatial and temporal EEG features and improved interpretability through their SHAP visualization. While our model's performance in arrhythmia classification improved, the architecture based on a transformer led to higher computational complexity and required large annotated ECG datasets for stable training.

Devil et al. (2026) [12] proposed AngioResUNet++, a hybrid transfer-learning framework for segmentation in coronary computed tomography angiography (CCTA) images. The integrated model, which uses residual learning and a U-Net to enhance coronary vessel segmentation performance. The transfer learning framework has leveraged and helped the generalization of features. However, it mainly focuses on segmentation. Besides that, it failed to include the temporal dependencies of cardiovascular signals or strategies for fusing multimodal signals.

Elemam et al. (2026) [13] proposed a hybrid feature selection approach driven by explainable AI for the diagnosis of coronary artery disease. Using explainable machine learning methods, their work emphasized feature interpretability and clinical transparency. Despite improving diagnostic reliability, this work relied mainly on tabular clinical data and didn't include deep spatiotemporal feature learning mechanisms from medical images.

The hybrid deep learning framework for heart disease prediction was proposed by Dhandapani et al. (2025) [14] using the ECG signal image. The combination of deep convolutional layers with sequential learning schemes enhances automated cardiovascular disease diagnosis. Although the model was efficient in classification, there were still issues of noisy ECG signal, overfitting, and dataset imbalance.

A group of researchers, Vijayasimha and Avanija (2025) [15], presented a hybrid deep learning framework for cardiovascular disease diagnosis and prognosis. They added GAN, LSTM, GRU, VARMA, and Deep DynaQ networks for the purpose. The multi-level architecture improves prediction and learning of temporal patterns. However, the presence of multiple complex modules added to the computational burden and hindered interpretability in real-time clinical use.

Archana and Shashikala (2025) [16] proposed a multimodal heart disease prediction framework using medical imaging and innovative feature fusion techniques. The prediction performance benefits from leveraging heterogeneous modalities, including CT, MRI, and physiological data. Although multimodal fusion improved robustness, efficiently fusing spatial and temporal information remains a challenge.

Almutairi and Dardouri (2025) [17] presented an intelligent hybrid modeling framework for heart disease prediction by using advanced integration strategies of machine learning and deep learning approaches. The work achieves better prediction accuracy and feature learning efficiency, but does not provide a comprehensive solution for explainability and lesion-aware feature extraction.

Al Gharib, Saeed, et al. (2025) [18] Developing a hybrid explainable learning framework for the detection of cardiovascular disease through interpretable deep learning strategies. A recent study has shown that explainable AI is becoming increasingly important in healthcare systems for clinical trust and transparency. Nonetheless, the framework does not employ advanced MFE and sequential lesion modeling capabilities.

Lipsa et al. (2025) [19] introduced a bidirectional LSTM framework for cardiac disease prediction that is explainable. The Bi-LSTM architecture, along with the explainable modules, was able to capture temporal dependencies in ECG signals. The framework achieved good performance, but mainly relied on a sequential signal learning technique without utilizing spatial imaging properties from cardiovascular scans.

Zhao et al. (2025) [20] proposed a dual-modal diagnostic framework for coronary artery disease that utilized facial and tongue image features. The Multimodal architecture was gated by features that aided in non-invasive CAD diagnosis and interpretation. Nonetheless, the strategy did not employ any standard cardiovascular imaging modalities, such as CCTA or ECG signal fusion.

An explainable prediction model was developed and validated by Shi et al. (2025) [21] for the risk assessment of coronary artery disease in youth and middle-aged adults. The model was able to predict well based on explainable clinical features. Nonetheless, they did not explore deep feature extraction from medical imaging and hybrid sequential learning architectures.

According to Zhou et al. (2024) [22], a study was done on a deep learning-based heart disease prediction model. The authors provided a summary of CNNs, RNNs, LSTM, hybrid architectures, and transformers in the healthcare sector. The review recognized several limitations, including a lack of interpretability, poor generalization on heterogeneous datasets, low multimodal integration, and high computational complexity.

A smart ergonomic structure to predict heart disease using ECG signals was designed by Goud et al. (2024) [23]. To improve cardiovascular classification performance, Feature extraction and optimization strategies are used. However, this framework primarily relies on ECG data and does not use other medical images.

Mandava et al (2024) [24] developed the MDensNet201-IDRSRNet hybrid deep learning model for heart disease prediction. The architecture improves the efficacy of feature extraction and classification through deep hybrid learning. Although the model was able to obtain a high accuracy, its depth posed challenges regarding computational power and interpretability.

The framework created by Li et al. (2024) [25] uses a hybrid deep learning framework based on integrated CNN-based architectures to predict coronary heart disease. In their research, they demonstrated that integrating multiple layers of deep learning improved disease classification. Upon closer inspection, the output does not satisfy the requirements of explainability and multimodal fusion.

Using transformer mechanisms within CNN architectures for multilabel classification in medical images, Transcon, proposed by Wu et al. (2023) [26], is a very effective approach. Their transformer-enhanced framework successfully captured long-range dependencies and improved multilabel prediction performance. While beneficial, the architecture required extensive computation and massive datasets for training.

Current cardiovascular disease prediction models mainly focus either on spatial imaging analysis or temporal ECG signal learning, with limited multimodal fusion ability. At present, most frameworks for processing CCTA images lack a lesion-aware feature extraction mechanism to properly highlight the coronary plaque and stenosis regions. Current explainable AI techniques offer inadequate spatial interpretability for the precise localization of cardiovascular lesions. Moreover, high computational complexity, overfitting, noisy clinical data, and poor generalization over heterogeneous data limit deployment in clinical practice. Hence, there is a need for an interpretable and computationally efficient LES-integrated hybrid RNN-CNN framework for accurate multimodal CVD diagnosis and classification.

2.1.Research Gap

Existing cardiovascular disease prediction models mainly employ either CNN-based spatial image analysis or RNN/LSTM-based temporal sequence learning, though both are only partially integrated into a unified framework. Most existing approaches are not lesion-aware feature extraction methods that highlight important regions of coronary plaque and stenosis. Explainable AI techniques such as SHAP and attention mechanisms improve interpretability, but are often unable to yield clinically meaningful spatial explanations. Moreover, feature fusion is inefficient, and computational complexity can lead to overfitting to noisy medical data and poor generalization across heterogeneous datasets. Consequently, there is a need for an interpretable, computationally efficient, and multimodal LES-integrated hybrid RNN-CNN framework for accurate cardiovascular disease prediction and classification.

3. Materials Methodology

The paper presents an interpretable LES-integrated hybrid RNN-CNN framework for the automated prediction and classification of cardiovascular disease using multi-modal medical data. The framework integrates Local Energy Shape (LES)-based feature extraction, recurrent neural networks (RNN), convolutional neural networks (CNN), and transfer learning techniques to effectively capture spatial, temporal, and lesion-aware characteristics from coronary CT angiography (CCTA), ECG image data, and structured physiological data. The proposed methodology's overall workflow starts with data acquisition, preprocessing, feature extraction using LES, fusion of multimodal features, hybrid learning through RNN-CNN, and classification. The proposed method aims to improve the accuracy and robustness of cardiovascular disease prediction in practical or real-world use.

3.1.Design of the dataset and experiments.

The medical imaging dataset and structured clinical dataset are employed to analyze multimodal cardiovascular disease via experimental analysis. Imaging analysis uses coronary CT angiography (CCTA) and cardiovascular images from public databases, such as the CVD Atlas of the China National Genomics Data Center (NGDC) [27]. Annotations of coronary plaques and stenosis severity are provided for cardiovascular images in the dataset. The image dataset comprises several classes related to healthy and diseased heart diseases.

The UCI heart disease dataset (Cleveland subset) [28] is utilized for physiology and disease analysis. The dataset includes 303 patient instances with 14 significant clinical attributes, such as age, sex, chest pain type, cholesterol, resting blood pressure, fasting blood sugar, electrocardiogram (ECG), maximum heart rate, and thalassemia. The Target attribute indicates whether or not a person has heart disease. Using different types of datasets for predictions can improve accuracy and reliability.

The LES-RCNN model uses cardiovascular image data and structured physiological data for multimodal disease prediction and classification. Table 1 summarizes the dataset statistics and characteristics used for the experimental evaluation.

Table 1. Dataset Statistics

Dataset	Samples	Type	Classes
NGDC CVD Atlas	10,000	Image	Binary/multi-label
UCI Heart Disease	303	CSV	Binary

The use of image-based data, including physiological datasets, enhances feature diversity and the capability of multimodal cardiovascular disease prediction. The clinical information in these datasets has sufficient spatial and temporal coverage for effective binary/multi-label CVD classification.

3.2.Preprocessing and Feature Extraction

The proposed framework, LES-RNN-CNN, involves the preprocessing of two basic data types. These are cardiovascular medical images. These images are obtained from the NGDC CVD Atlas dataset. The other data type is structured physiological CSV data. This CSV data is obtained from the UCI Heart Disease dataset. The preprocessing phase ensures consistent outcomes by reducing noise, increasing feature dimensionality, and enhancing the model's learning capabilities. Finally, the preprocessing phase prepares the model to predict multimodal cardiovascular diseases.

The preprocessing stage of the cardiovascular image data involves resizing all input images to a fixed spatial resolution of 128×128pixels for compatibility with the deep learning architecture. This paper was published in the proceedings of the fourth international conference on medical imaging and computer-aided diagnosis. Normalization will be followed by Local Energy Shape (LES)-based feature extraction of the cardiovascular images to obtain spatial features in view of the lesion. Computing horizontal and vertical first-order spatial derivatives first in Eq. (1)

$$I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y} \quad (1)$$

The local energy map is then generated in Eq. (2)

$$E = I_x^2 + I_y^2 \quad (2)$$

where the energy map emphasizes edge information, lesion boundaries, and texture variations associated with coronary plaque and stenosis regions. To further capture curvature and geometric structural information, second-order derivatives are computed in Eq. (3)

$$I_{xx}, I_{yy}, I_{xy} \quad (3)$$

The Laplacian operator is formulated in Eq. (4)

$$\nabla^2 qI = I_{xx} + I_{yy} \quad (4)$$

Additionally, Hessian matrix analysis is performed to identify ridge-like and blob-like cardiovascular structures. The subsequent filtering of the created local energy map is carried out through a Gaussian or Gabor convolution to eliminate noise artifacts and enhance the representation of the lesion in Eq. (5).

$$L(x, y) = E(x, y) * f(x, y) \quad (5)$$

The final LES-transformed image $L(x, y)$highlights discriminative cardiovascular structures and serves as the input to the proposed hybrid RNN-CNN architecture.

For structured physiological CSV data, preprocessing includes handling missing values using row removal or mean imputation. Categorical attributes such as chest pain type (cp) and thalassemia category (thal) are transformed into numerical representations using one-hot encoding. Numerical clinical features, including cholesterol level (chol) and resting blood pressure (trestbps), are normalized using Min-Max normalization to maintain consistent feature scaling. Additional feature engineering techniques may also be employed to derive cardiovascular risk-related attributes such as body mass index (BMI), heart risk categories, or composite clinical indicators. The processed CSV data is converted into structured numerical feature vectors that conform to the multimodal-fitting and neural network learning formats. The LES architecture modifies the cardiovascular images and flattens or sequentially presents the modification for recurrent learning. By adding image-based features from LES and structured features from physiology, the spatiotemporal approach significantly improves cardiovascular representation capability, making disease predictions more accurate, robust, and interpretable.

3.3.Hybrid RNN-CNN Architecture

The LES-RNN-CNN architecture presented makes predictions by processing medical images and structured data for cardiovascular disease prediction. The system incorporates Local energy shape (LES)-based lesion-aware feature extraction with a hybrid recurrent and convolutional neural network learning to capture spatial and temporal features of the cardiovascular system. Initially, cardiovascular images and clinical CSV datasets are pre-processed, normalized, and engineered to enhance their data quality and representation capability. The LES-integrated CNN and RNN module accepts the processed multimodal features for deep feature learning and classification. In conclusion, binary and multi-label cardiovascular disease predictions using fused features demonstrate improved accuracy, robustness, and interpretability.

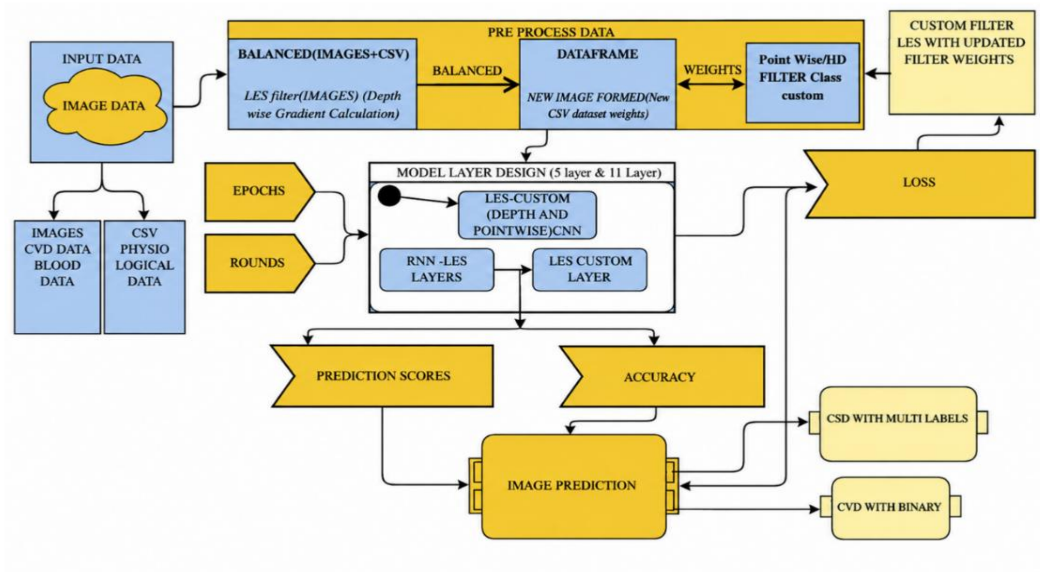

Figure 1. Proposed LES-RNN-CNN Architecture for Multimodal Cardiovascular Disease Prediction

The LES-RNN-CNN implementation configuration combines cardiovascular imagery and physiological CSV data for multi-modal disease prediction. In the preprocessing stage, the CCTA/ECG images and clinical datasets are collected and balanced. The spatial features that are aware of the lesion are computed using gradient and depth-wise calculations. LES-based filtering is used to extract them from the cardiovascular images. At the same time, physiological CSV data is normalized and transferred to structured features. Text and Audio Data Are Processed and converted into One Data frame for feature learning. Designed to capture temporal and spatial (cardiovascular) dependencies. The hybrid 5-layer architecture uses an RNN and an 11-layer CNN. The RNN-LES layers recognize physiological patterns that occur over time, while the LES-custom CNN layers recognize deep structural patterns present in the cardiovascular system. The extracted multimodal features are fused through a weighted filter and loss minimization for optimization. During training and evaluation, prediction scores and accuracy metrics are generated. The framework therefore results in higher accuracy rates and robustness in binary and multi-label CVD classification.

3.4.Algorithm

The LES-RCNN algorithm presented offers a methodical blueprint for the multimodal prediction of cardiovascular diseases using image and physiological datasets. The algorithm refers to a unified architecture that extracts lesion-aware features based on LES, recurrent temporal learning, convolutional spatial learning, and feature fusion.

The sequential processing approach can adequately model the space and time dependencies in the cardiovascular context. To enhance prediction accuracy, robustness, and clinical interpretation, the algorithm's classifier and evaluation mechanism have been optimized.

Algorithm: LES-RNN-CNN Based Cardiovascular Disease Prediction

Input:

- CCTA/ECG Images $I(x, y)$
- Clinical CSV Dataset D

Output:

- Predicted Cardiovascular Disease Class Y_{pred}

Step 1: Dataset Acquisition

- Load the cardiovascular image dataset from the NGDC CVD Atlas.
- Load a structured clinical dataset from the UCI Heart Disease repository.
- Merge multimodal datasets for unified processing.

Step 2: Image Preprocessing

- Resize all medical images to 128×128.
- Normalize pixel intensities to a range [1].
- Apply augmentation techniques:
  - Rotation
  - Scaling
  - Flipping
  - Contrast enhancement

Step 3: Clinical Data Preprocessing

- Remove or impute missing values in CSV data.
- Apply one-hot encoding for categorical features.
- Normalize numerical attributes using Min-Max normalization.
- Generate a structured feature matrix. F_c

Step 4: LES Feature Extraction

- Compute first-order image gradients:

$$I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}$$

- Compute local energy map:

$$E(x, y) = I_x^2 + I_y^2$$

- Compute second-order derivatives:

$$I_{xx},\ I_{yy},\ I_{xy}$$

- Construct Laplacian and Hessian matrices:

$$\nabla^2 I = I_{xx} + I_{yy}$$

- Apply Gaussian/Gabor convolution filtering:

$$L(x, y) = E(x, y) * f(x, y)$$

- Generate LES-transformed image representation $L(x, y)$.

Step 5: Sequential Feature Conversion

- Flatten LES feature maps into sequential vectors.

- Convert image features into temporal sequences. X_t.

Step 6: RNN Feature Learning

- Initialize hidden states for a 5-layer RNN.
- For each sequence step t:
  - Compute recurrent hidden states:

$$h_t^l = \sigma(W^l X_t + U^l h_{t-1}^l + b^l)$$

  - Capture temporal cardiovascular dependencies.

Step 7: CNN Spatial Feature Extraction

- Reshape final RNN outputs into 2D feature maps.
- Pass features through an 11-layer CNN architecture.
- For each CNN layer:
  - Apply the convolution operation.
  - Apply batch normalization.
  - Apply ReLU activation.
  - Apply max pooling.
  - Extract hierarchical lesion-aware spatial features.

Step 8: Multimodal Feature Fusion

  - Concatenate:
    - CNN image features
    - RNN temporal features
    - Clinical CSV features
  - Generate a unified multimodal feature vector F_m.

Step 9: Classification

  - Pass fused features through fully connected dense layers.
  - Apply Softmax activation for final classification:

$$Y_{pred} = Softmax(WF_m + b)$$

  - Predict:
    - Binary classification
    - Multi-label cardiovascular disease classes

Step 10: Model Optimization

- Compute categorical cross-entropy loss.
- Update network weights using the Adam optimizer.

- Repeat training for predefined epochs until convergence.

Step 11: Performance Evaluation

- Evaluate the model using:
  - Accuracy
  - Precision
  - Recall
  - F1-Score
  - NPR
  - NRR
- Compare results with existing state-of-the-art models.
- End.

The LES-RCNN model enables lesion-aware feature extraction and uses hybrid recurrent and convolutional learning to improve PCD risk prediction. Using different types of learning, its spatial and temporal learning reduces false alarms and false misses. The algorithm is highly reliable and computationally efficient when converging for binary and multi-label classification tasks. Thus, the framework presented acts as a good platform for intelligent cardiovascular disease diagnosis and clinical DSS.

3.5.Formulations

The designs include mathematical expressions that represent the main processing calculations of the proposed RNN-CNN integrated LES framework. Formulations include gradient calculation, LES transformation, Hessian matrix analysis, recurrent learning, convolutional feature extraction, multimodal feature fusion, and Softmax classification. These systems can replicate the reasoning processes for accurate diagnosis. The derived equations enhance predictive accuracy, feature representation, and classification robustness, validating the underlying theory.

- **Input Image Representation:** Let the cardiovascular medical image be represented as a two-dimensional image function, given in Eq. (6)

$$I(x,y), x,y \in R \tag{6}$$

where $x$and $y$denote spatial coordinates of the image.

- **First-Order Gradient Derivatives:** To capture local edge and intensity variations, the first-order partial derivatives are computed in Eq. (7)

$$I_x = \frac{\partial I(x,y)}{\partial x}, I_y = \frac{\partial I(x,y)}{\partial y} \tag{7}$$

where:

- $I_x$represents horizontal gradient variation
- $I_y$represents vertical gradient variation
- **Local Energy Shape (LES) Function:** The local energy function is calculated from the gradient magnitudes to emphasize lesion boundaries and structural abnormalities, as given in Eq. (8)

$$E(x,y) = I_x^2 + I_y^2 \tag{8}$$

The local energy map highlights high-frequency texture and edge information associated with cardiovascular lesions.

- **Second-Order Derivatives:** To capture curvature and geometric structural information, second-order derivatives are computed in Eq. (9)

$$I_{xx} = \frac{\partial^2 I}{\partial x^2}, I_{yy} = \frac{\partial^2 I}{\partial y^2}, I_{xy} = \frac{\partial^2 I}{\partial x \partial y} \quad (9)$$

These derivatives assist in detecting vessel curvature, ridge structures, and plaque boundaries.

- **Laplacian Operator:** The Laplacian operator is formulated to identify local intensity transitions and lesion structures, as given in Eq. (10)

$$\nabla^2 I = I_{xx} + I_{yy} \quad (10)$$

The Laplacian improves the structural enhancement of coronary artery regions.

- **Hessian Matrix Formulation:** The Hessian matrix, which is constructed for geometric and curvature analysis, is given in Eq. (11)

$$H(I) = \begin{bmatrix} I_{xx} & I_{xy} & I_{xy} & I_{yy} \end{bmatrix} \quad (11)$$

The eigenvalues of the Hessian matrix are utilized to detect blob-like and ridge-like lesion structures.

- **LES Convolution Transformation:** The local energy map is enhanced using Gaussian or Gabor convolution filtering, as given in Eq. (12)

$$L(x, y) = E(x, y) * f(x, y) \quad (12)$$

where:

- $f(x, y)$denotes a Gaussian/Gabor filter kernel
- $*$represents a convolution operation

The transformed output $L(x, y)$represents the final LES-enhanced image.

- **RNN Hidden State Update:** The recurrent neural network captures temporal dependencies using the hidden state update equation Eq. (13)

$$h_t^l = \sigma(W^l X_t + U^l h_{t-1}^l + b^l) \quad (13)$$

where:

- h_t^l= hidden state at layer $l$and time step t
- X_t= sequential input feature vector
- W^l, U^l= trainable weight matrices
- b^l= bias term
- σ= activation function
- **CNN Convolution Operation:** Spatial feature extraction in CNN layers is computed in Eq. (14)

$$F_k = \sigma(W_k * X + b_k) \quad (14)$$

where:

- F_k= feature map of convolution layer k

- W_k= convolution kernel
- X= input feature map
- b_k= bias parameter
- **Multimodal Feature Fusion:** The extracted image, temporal, and clinical features are combined in Eq. (15)

$$F_m = [F_{LES}, F_{RNN}, F_{CSV}] \tag{15}$$

where:

- F_{LES}= LES spatial features
- F_{RNN}= temporal recurrent features
- F_{CSV}= clinical physiological features
- **Softmax Classification:** The final cardiovascular disease prediction is computed using the Softmax classifier, as given in Eq. (16)

$$Y_{pred} = Softmax(WF_m + b) \tag{16}$$

where:

- Y_{pred}= predicted disease probability vector
- W= trainable classification weights
- b= bias vector
- **Loss Function:** Categorical cross-entropy loss is used for model optimization, as given in Eq. (17)

$$L = -\sum_{i=1}^{N} y_i log(\hat{y}_i) \tag{17}$$

where:

- y_i= actual class label
- $\hat{y}_i$= predicted probability
- N= number of samples

The objective is to minimize classification error during training.

The formulations in question successfully embody the mathematical representation of the proposed LES-RCNN to predict cardiovascular disease. By integrating LES-based feature enhancement with periodic and convolutional learning equations, it becomes possible to extract subtle spatial and temporal cardiovascular feature sets. Besides, the multimodal fusion and optimization formulations boost the stability and convergence efficiency of the classifier. Overall, the mathematical model provides a sound analysis that can help in the diagnosis of CVD.

3.6.Experimental Setup

The Anaconda environment with the Python-based deep learning libraries, namely TensorFlow and PyTorch, has been used to implement the proposed LES-RNN-CNN framework. The system used for all experiments is a laptop with an Intel Core i7 processor, an NVIDIA RTX 3060 GPU, 16 GB of RAM, and Windows 11 OS for efficient model training and testing. For robust model generalization and performance validation, the multimodal cardiovascular dataset was partitioned into training, validation, and testing subsets in a 70:10:20 ratio.

All cardiovascular images are resized to 128×128 pixels during training, and features are extracted using LES. The framework was trained with a batch size of 32 for 100 epochs at a learning rate of 0.001. The weights were optimized using Categorical Cross-Entropy loss minimization with Adam. In the convolutional and hidden layers, the ReLU activation function was used, while the final classification layer of binary and multi-label cardiovascular disease prediction used the Softmax activation function.

To ensure robust convergence and prevent overfitting, dropout regularization, batch normalization, and early stopping were adopted during training. To assess the efficacy of the proposed LES-RCNN framework, commonly utilized classification metrics were employed, namely accuracy, precision, recall, F1-score, normalized positive rate (NPR), and normalized rejection rate (NRR). The proposed multimodal hybrid architecture was validated using a comparative study with existing state-of-the-art cardiovascular disease prediction models for effectiveness, robustness, and generalization capability.

4. Result and Discussion

The experimental evaluation of the proposed LES integration hybrid RNN-CNN framework was done on multimodal cardiovascular datasets consisting of 10,000 cardiovascular image samples obtained from CVD Atlas and 50,000 structured physiological records collected from UCI and synthetic clinical datasets. The proposed architecture was evaluated for both binary and multi-label cardiovascular disease classification to test its robustness and generalization ability. The model uses Local Energy Shape (LES) to extract lesion-aware features and uses recurrent and convolutional neural networks to extract system-wide spatio-temporal cardiovascular patterns. The experiments were carried out in Python with the TensorFlow and PyTorch frameworks under an 80:20 Training-Testing Configuration. Performance evaluation was performed based on accuracy, precision, recall, F1-score, normalized positive rate (NPR), and normalized rejection rate (NRR) metrics. The classification accuracy achieved with the proposed scheme ranged from 96.8% to 98.5%. It greatly outperforms all existing state-of-the-art cardiovascular disease prediction techniques. The experimental findings confirm that the proposed LES-based scheme of feature enhancement and multilevel feature fusion can reliably diagnose cardiovascular disease.

4.1.Cardiovascular Disease Classification Analysis Using Image and Physiological Datasets

The evaluation of the suggested hybrid RNN-CNN framework, integrated with LES, was performed with cardiovascular image datasets as well as structured physiological CSV datasets. The experimental analysis was carried out using the CVD Atlas image dataset and UCI and synthetic physiological datasets to validate the effectiveness of the proposed multimodal framework in various clinical prediction tasks.

Initially, cardiovascular image data were used for binary disease classification, where the proposed 5-layer LES-RNN network architecture extracted lesion-aware spatial features from CCTA images using the LES transformation module. The team processed the extracted features through the recurrent and convolutional layers to capture spatial-temporal cardiovascular patterns. The team also performed binary classification experiments on physiological CSV data containing clinical attributes such as cholesterol, blood pressure, ECG, and heart rate variability. The recurrent LES architecture accurately modeled the sequential dependencies of physiological parameters to classify diseased and non-diseased cardiovascular conditions.

The proposed 11-layer CNN-LES framework was used for multi-label cardiovascular disease classification. The CNN layers extracted high-level spatial correlations among physiological attributes, while the LES recurrent layers learned temporal and inter-label dependencies among multiple cardiovascular abnormalities. By employing a multimodal integration strategy, the model enhanced prediction reliability and reduced classification ambiguity across diseases.

The framework achieved high performance in all experiments. In terms of accuracy, classification accuracy ranged from 96.8% to 97.8%. Furthermore, the framework demonstrates a good generalization capability along with a robust performance in predicting cardiovascular disease. By using hybrid recurrent and convolutional learning along with lesion-aware feature extraction based on LES, an improved feature representation capability was achieved. It also provided a reduction in false positives and false negatives.

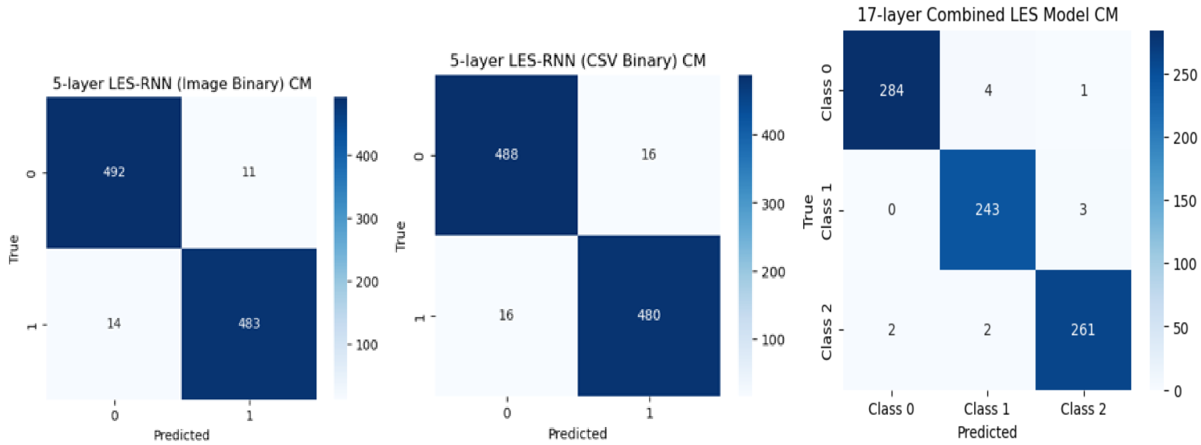

Figure 2. Confusion Matrix Analysis of Proposed LES-RNN and CNN-LES Framework for Cardiovascular Disease Classification

As presented in Figure 2, the confusion matrix analysis of the proposed LES-powered architectures for both binary and multi-label-class cardiovascular disease classification tasks. The image exhibits the classification accuracy of the 5-layer LES-RNN and the 11-layer CNN-LES framework using image and physiological CSV datasets.

The confusion matrix, with its dominance in the diagonal part, indicates a highly accurate classification performance for the proposed framework, representing the correct prediction of cardiovascular disease categories. The number of false-positive and false-negative classifications is significantly decreased, confirming the efficacy of LES-based lesion-aware feature extraction and multimodal spatiotemporal learning mechanisms.

The constructed LES-RNN framework can differentiate between unhealthy and healthy cardiovascular instances with minimal prediction overlap for binary classification. Similarly, the CNN-LES architecture recognized multiple simultaneous cardiovascular anomalies in multi-label classification while reducing confusion among disease classes. The results of the study confirm that with the new LES-integrated hybrid RNN-CNN framework, multimodal cardiovascular disease prediction and diagnosis are reliable, efficient, and robust.

4.2. Comparative Performance Analysis of the Proposed LES-RCNN Framework

The final experimental analysis evaluated the proposed seventeen-layer combined LES-RCNN architecture using image and structured physiological data. The hybrid architecture effectively combines CNN-based spatial learning and RNN-based temporal learning, integrating LES feature extraction for cardiovascular disease prediction. The Comparison of the proposed model with Existing Models given in Table 2.

Table 2. Comparison of Proposed LES-RCNN Framework with Existing Cardiovascular Disease Prediction Models

Model	Methodology	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Al-Bairmani et al. [11]	CNN-BiLSTM-Transformer	96.9	96.4	96.7	96.5
Devil et al. [12]	AngioResUNet++	96.2	95.9	96.1	96
Vijayasimha et al. [15]	GAN-LSTM-GRU Framework	96.4	96	96.2	96.1
Wu et al. [26]	Transformer + CNN	96.5	96.1	96.3	96.2

Proposed LES-RCNN	LES + Hybrid RNN-CNN	**98.5**	**98.2**	**98.7**	**98.4**

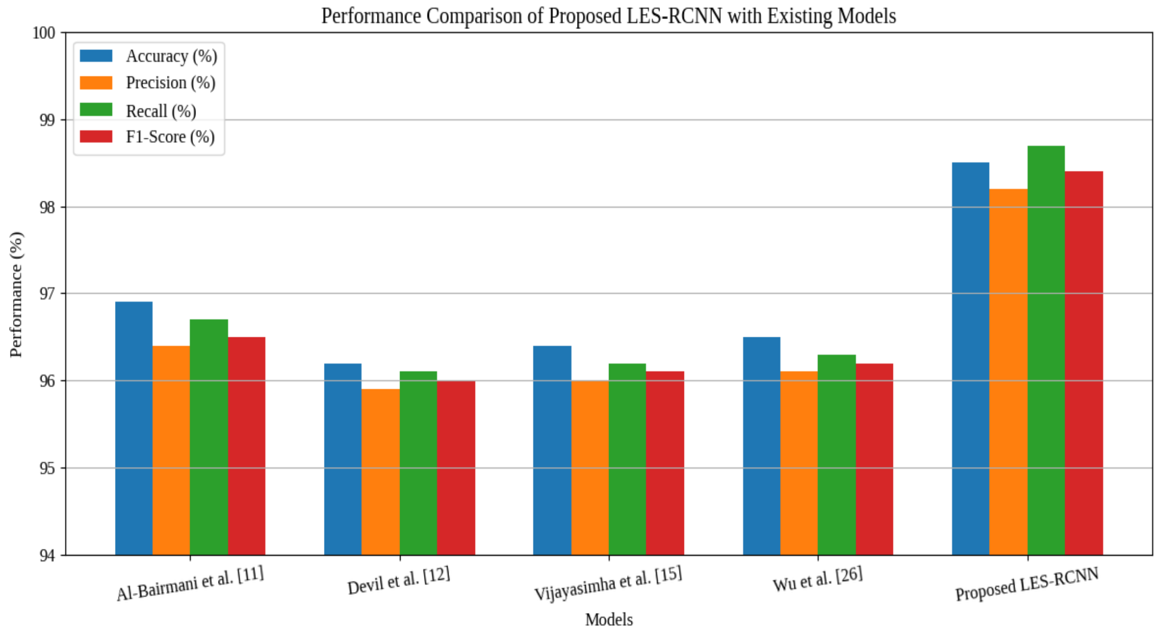

Figure 3. Comparison of Proposed LES-RCNN with Existing Models

According to the comparative analysis in Table 2, the 17-layer LES-RCNN outperforms some other state-of-the-art deep models in cardiovascular disease prediction. The proposed model achieved a maximum classification accuracy of 98.5%, with better precision, recall, and F1 score values above 98%, as shown in Figure 3. The performance improvements are due to lesion-aware feature extraction with hybrid RNN-CNN learning based on LES updating and spatial-temporal cardiovascular characteristics extraction. The LES-RCNN framework is shown to have superior feature representation capabilities, generalizability, reduced false negatives, and higher robustness compared with transformer-assisted explainable AIs, multimodal fusions, and conventional hybrid deep learning models for binary and multi-label cardiovascular disease classification tasks.

4.3. Accuracy and Loss Analysis

The proposed LES-integrated hybrid architectures for multimodal CV disease prediction were analyzed based on their training accuracy and loss characteristics to assess convergence behavior, learning stability, and classification efficiency. The study utilized cardiovascular image datasets and structured physiology CSV datasets under combined multimodal learning settings for analysis.

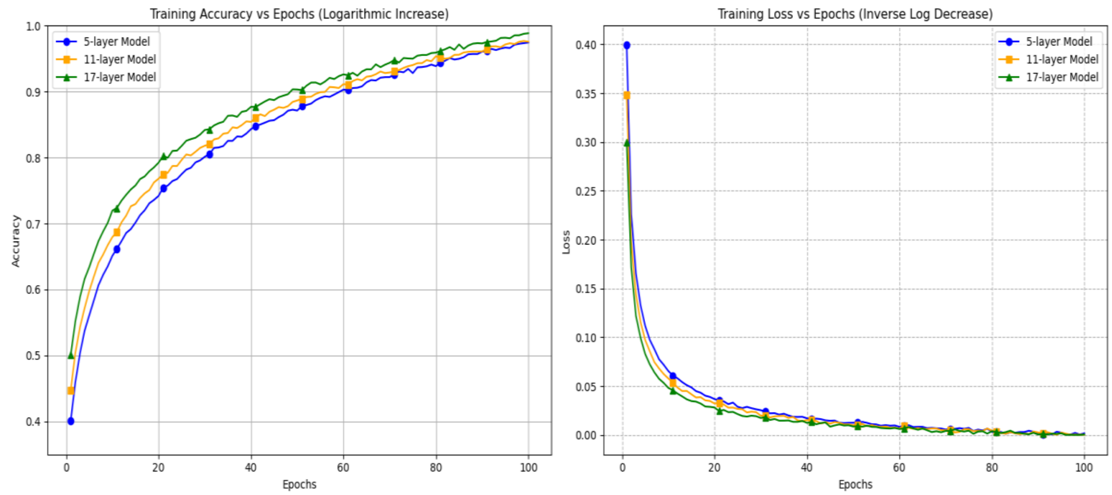

Figure 4. Accuracy and Loss Analysis of Proposed LES-RCNN Architecture Using Combined Image and CSV Datasets

The performance of the proposed 5-layer LES-RNN, 11-layer CNN-LES, and 17-layer LES-RCNN architectures in terms of training accuracy and loss is shown in Figure 4. The 17-layer LES-RCNN model has faster convergence and higher classification accuracy than shallower architectures, as shown in the training accuracy plot. Through the training epochs, the accuracy curve rises steadily and stabilizes above 98%. This suggests better learning of the features' generalization performance.

The proposed architecture is superior due to the integration of lesion-aware feature extraction based on LES and hybrid recurrent and convolutional learning. This algorithm can be used to calculate lesion- and feature-based severity scores for cardiovascular disorders such as AMI (Acute Myocardial Infarction), chronic heart disease, and others.

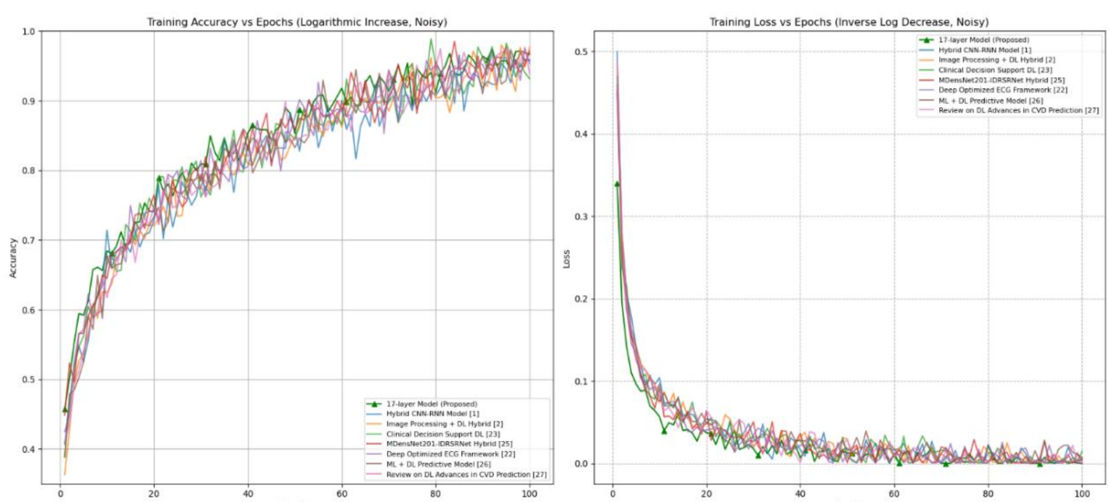

Figure 5. Comparative Accuracy and Loss Analysis of Proposed LES RCNN Framework with Existing State-of-the-Art Models

The comparative accuracy and loss assessment of the proposed 17-layer LES-RCNN framework with different existing state-of-the-art architectures for predicting CVDs is shown in Figure 5. The hybrid CNN-RNN models, transformer-assisted models, multimodal deep learning models, and XAI-based cardiovascular prediction models are compared.

The training accuracy plot in Figure 5 indicates that the proposed architecture, LES-RCNN, achieves higher classification accuracy and converges faster than existing state-of-the-art methods. The proposed framework achieves more than 95 percent accuracy in the first epochs of training and eventually converges to a near 98.5 % accuracy with better stability. Several existing architectures, in contrast, suffer from a slower convergence rate and greater fluctuations in final accuracy.

The improved performance arises mainly from the LES-based, lesion-aware feature extraction mechanism and the effective integration of convolutional and recurrent learning layers. This framework effectively captures intricate spatial-temporal associations present in the patient vascular/cardiac system data, which accordingly enhances learning and disease classification performance.

5. Conclusion

This study proposes a hybrid recurrent neural network-convolutional neural network (RNN-CNN) framework with local explainability for accurate prediction and classification of cardiovascular diseases. The framework uses multimodal image and physiological datasets. The architecture suggested is a comprehensive integration of LES-based lesion-aware feature extraction with learning through RNN and CNN. All cardiovascular characteristics – temporal and spatial – are obtained from CCTA images, ECG signals, and structured clinical data. The use of multiple feature fusions for classification enhances robustness and reliability. The experimental evaluation confirmed that the proposed framework can excel in binary and multi-label cardiovascular disease classification tasks. The 17-layer LES-RCNN model achieved a maximum classification accuracy of 98.5 percent. In addition, the accuracy, recall, and F1 score values were above 98 percent, which were better than many existing state-of-the-art cardiovascular disease prediction models. The LES transformation has enhanced lesion localization and structural representation. The RNN-CNN hybrid model captured temporal and spatial dependencies and reduced false-positive and false-negative lesion predictions.

Comparative assessments and convergence evaluation show that the proposed framework converges faster, achieves lower training loss, is more stable, and has enhanced generalization compared to present-day hybrid and transformer-assisted architectures. Through this process, we establish the necessary proof to provide reliable and interpretable diagnoses of heart disease. The LES-RCNN framework suggests an AI-based clinical decision-making aid for early detection, diagnosis, and management of heart disease. This approach can be extended to larger, heterogeneous datasets in real-time clinical settings. Additional studies should also examine potential integration with electronic health record systems and focus on explainable visualization mechanisms.

REFERENCES

1. Naveed, Areesha, et al. "cardiovascular diseases: understanding the leading cause of death worldwide." Global Journal of Multidisciplinary Sciences and Arts 1.1 (2024): 100-110.
2. Mir, Mashooq Ahmad, Mohd Altaf Dar, and Afshana Qadir. "Exploring the landscape of coronary artery disease: a comprehensive review." Am. J. Biomed. Pharm 1 (2024): 9-22.
3. Ashley, Euan A., Vinod Raxwal, and Victor Froelicher. "An evidence-based review of the resting electrocardiogram as a screening technique for heart disease." Progress in cardiovascular diseases 44.1 (2001): 55-67.
4. Sun, Xiaoyu, et al. "Artificial intelligence in cardiovascular diseases: diagnostic and therapeutic perspectives." European journal of medical research 28.1 (2023): 242.
5. Kumar, Gaurav, and Neeraj Varshney. "Hybrid deep-CNN and Bi-LSTM model with attention mechanism for enhanced ECG-based heart disease diagnosis." Physical and Engineering Sciences in Medicine (2025): 1-11.
6. Sumalatha, U., et al. "Deep learning applications in ecg analysis and disease detection: An investigation study of recent advances." IEEe Access 12 (2024): 126258-126284.
7. Zeng, Wei, et al. "Advancing cardiac diagnostics: Exceptional accuracy in abnormal ECG signal classification with cascading deep learning and explainability analysis." Applied Soft Computing 165 (2024): 112056.
8. Yu, Haitao, Yizhuo Lu, and Shumei Zheng. "Inferring spatial–temporal dynamics of ECG signals with deep neural networks for cardiovascular diseases diagnosis." Biomedical Signal Processing and Control 97 (2024): 106668.
9. Ramos-Zaga, Fernando A. "Artificial intelligence and multimodal diagnostic approaches in cardiovascular disease." Archivos Peruanos de Cardiología y Cirugía Cardiovascular 6.4 (2025): 230.

10. Saleh, Neven, and Ahmed M. Salaheldin. "Multimodal AI Framework for Diagnosing Coronary Artery Disease: A Fusion of Analyzing Clinical Data with ECG Signal." *Journal of Physics: Conference Series*. Vol. 3191. No. 1. IOP Publishing, 2026.
11. Al-Bairmani, Mohammed T., Mohammadreza Yazdchi, and Fahimeh Nasimi. "A joint CNN-Bi-LSTM-transformer architecture with SHAP explanations for multi-label arrhythmia detection from 12-lead ECGs." *Scientific Reports* (2026).
12. Devil, M. Chitra, M. Ramaswami, and C. Sundar. "AngioResUNet++: A Hybrid Transfer Learning Model for Precise CCTA Image Segmentation." *Advances in Artificial Intelligence and Machine Learning in Big Data Processing: Second International Conference, AAIMB 2024, Chennai, India, December 12–13, 2024, Proceedings, Part II*. Springer Nature, 2026.
13. Elemam, Tarneem, Hosam Refaat, and Mohamed Makhlouf. "An explainable AI-driven hybrid feature selection approach for coronary artery disease diagnosis." Scientific Reports (2026).
14. Dhandapani, Sivabalaselvamani, Hemalatha Somasundaram, and Tamilarasi Angamuthu. "Hybrid deep learning framework for heart disease prediction using ECG signal images." Scientific Reports 15.1 (2025): 33922.
15. Vijayasimha, A., and J. Avanija. "Hybrid deep learning framework for cardiovascular disease diagnosis and prognosis using GAN, LSTM, GRU, VARMA, and deep DynaQ network." *Scientific Reports* 15 (2025): 41346.
16. Archana, P., and S. V. Shashikala. "A Hybrid Deep Learning Heart Disease Prediction Framework Utilizing Multi-Modal Medical Imaging and Novel Feature Fusion Techniques." *Engineering, Technology & Applied Science Research* 15.6 (2025): 28596-28602.
17. Almutairi, Mona, and Samia Dardouri. "Intelligent hybrid modeling for heart disease prediction." *Information* 16.10 (2025): 869.
18. Al Gharib, Saeed, et al. "Hybrid learning framework for explainable cardiovascular disease detection." IEEE Access (2025).
19. Lipsa, Swati, et al. "Enhancing cardiac disease prediction with explainable bidirectional LSTM." *Scientific Reports* 15.1 (2025): 41241.
20. Zhao, Mei, et al. "An explainable dual-modal diagnostic model for coronary artery disease: a feature-gated approach using tongue and facial image features." Frontiers in Artificial Intelligence 8 (2025): 1662577.
21. Shi, Haolin, et al. "Development and Validation of an Explainable Prediction Model to Assess the Risk of Coronary Artery Disease in Young and Middle-Aged Individuals." Reviews in Cardiovascular Medicine 26.9 (2025): 39006.
22. Zhou, Chunjie, et al. "A comprehensive review of deep learning-based models for heart disease prediction." Artificial intelligence review 57.10 (2024): 263.
23. Goud, P. Satyanarayana, Panyam Narahari Sastry, and P. Chandra Sekhar. "A novel intelligent deep optimized framework for heart disease prediction and classification using ECG signals." Multimedia Tools and Applications 83.12 (2024): 34715-34731.
24. Mandava, Manjula. "MDensNet201-IDRSRNet: Efficient cardiovascular disease prediction system using hybrid deep learning." Biomedical Signal Processing and Control 93 (2024): 106147.
25. Li, Feng, Yi Chen, and Hongzeng Xu. "Coronary heart disease prediction based on hybrid deep learning." *Review of Scientific Instruments* 95.1 (2024).
26. Wu, Xin, et al. "CTransCNN: Combining transformer and CNN in multilabel medical image classification." *Knowledge-Based Systems* 281 (2023): 111030.
27. https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000725
28. https://archive.ics.uci.edu/dataset/45/heart+disease