Research Article

An Intelligent System for Cardiovascular Disease Prediction Using Deep Learning

Deepika Arora[1] , Deepak Dudeja[1] and B K Agrawal[2]

[1]*Department of Computer Science and Engineering, MM Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, Haryana 133207, India*
[2]*Department of Medical Sciences, M.M Institute of Medical Sciences and Research, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, India*

Corresponding Author:
Deepika Arora
Department of Computer Science and Engineering, MM Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, Haryana 133207, India
Email: deepika.arora86@gmail.com

Abstract: Heart disease remains among the major causes of death in the world. Hence, accurate and reliable prediction schemes are key. To detect the presence of diseases in early stages, data mining and deep learning algorithms offer powerful tools to determine important trends in large and complex data. To enhance the prediction of cardiac diseases, this paper proposes a hybrid deep learning model, which is the combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The model is implemented by using Python, and its effectiveness is evaluated by such performance measures as recall, accuracy, and precision. According to experimental data, the hybrid CNN–LSTM model outperforms a standalone CNN by 8%, achieving 98% training accuracy and 97% validation accuracy. Additionally, the ROC curve shows an improvement in true positive rate of 0.83, and the model achieves precision and recall values of 0.97 each. These results show that the suggested hybrid strategy outperforms traditional deep learning models in terms of predictive power, facilitating better clinical judgment in the prognosis of cardiovascular disease.

Keywords: Heart Disease, CNN, LSTM, Deep Learning

Introduction

Heart disease is also one of the major causes of mortality at all ages; as such, healthcare can utilize several machine learning techniques to enhance its ability to predict heart attacks. These diseases encompass numerous diseases that compromise the heart and blood vessels, and the symptoms also differ according to the kind of heart disease (Shanmugapriyaai, 2016). Diagnosis of Cardiovascular Disease (CVD) is a complicated process, which demands much knowledge and experience on the part of the medical practitioners. There are many risk factors for heart disease, such as age, diabetes, smoking, obesity, and poor dietary habits. These elements can cause heart disease. Therefore, having prior knowledge and information from relevant pathological events is crucial for accurate and timely diagnosis of CVD (David and Belcy, 2018). The adoption of management software is being adopted by most hospitals to track both clinical and patient records and create useful information regarding the patients. The software is based on the data mining technologies of EHR (Electronic Health Records) to generate detailed medical data that touches upon all areas of patient care, such as diagnostics, treatment, and laboratory tests. Consequently, it makes the investigation and detection of clinical evidence very easy. Specifically, machine learning in heart disease prediction has gained more and more significance.

As depicted in Figure 1, heart disease prediction starts with the gathering of publicly available heart disease databases of healthcare organizations, which is then preprocessed into usable patterns by cleaning them (Vijayashree and NarayanaIyengar, 2016). Feature engineering is used to choose the most relevant attributes, lowering dimensionality and boosting the predictive model. Model validation is done using LOSO cross-validation, which ensures a reliable evaluation across people. Heart disease prediction is formulated as a classification issue, in which processed features are assigned to distinct classes for diagnosis.

Fig. 1: Heart Disease Prediction Framework (Ratnam et al., 2014)

The model impact is fully quantified based on conventional measures of accuracy, sensitivity, specificity, recall, F1-score, MCC, AUC, and ROC curves (Ratnam et al., 2014).

Types of machine learning techniques that have been discussed earlier, like SVM, KNN, and Decision trees, have been used to predict cardiac ailments, but their precision is below average in the majority of instances due to hand-coded information, in addition to shallow representation. Deep learning methods have, however, been proven to be viable alternatives due to their characteristics of automatically learning complex and non-linear feature representations on large medical datasets (Santhanam and Ephzibah, 2015). In contrast to more traditional methods, where the bulk of the effort is on manual feature engineering, deep learning architectures have the potential to directly extract multi-level features out of raw data, making it possible to detect underlying patterns that may not otherwise be detected by humans (Purusothaman and Krishnakumari, 2015; Thenmozhi and Deepika, 2014). Convolutional Neural Networks (CNNs) can be very effective in discovering structured datasets with spatial relationships and hierarchies. LSTM networks, on the other hand, are more suited to the task: Their ability to capture long-term relationships and temporal trend patterns has made them the go-to choice in applications involving time-series health data, such as electrocardiograms (ECGs) (Srinivas et al., 2010; Dangare and Apte, 2012) and patient monitoring records. The work developed provides a hybrid deep learning architecture with CNN and LSTM models that fuses the merits of the two (Peter and Somasundaram, 2012). CNN layers automatically perform the feature extraction process by highlighting the most central qualities of the clinical outcomes, whereas LSTM layers examine the sequence and time disparities

of the cardiovascular diseases (Masethe and Masethe, 2014). The hybrid brings additional resiliency, generalizability, capacity to manage complex variations in patient data, and predictive accuracy to the model (Soni et al., 2011).

Research Objectives

The objectives of the research of this effort are as follows:

- To develop a hybrid deep learning model (CNN + Bi-LSTM), with the help of which, in comparison with individual models, the prediction of cardiovascular illness is more accurate
- To apply PCA to decrease the dimensions, enhance the quality, and ease the calculations
- To estimate the performance of the proposed model and compare it with the conventional model of deep learning (CNN, LSTM), and prove it on the Cleveland dataset
- To maximize such measures as accuracy, precision, recall, and ROC-AUC to grow the reliability of the predictions

Research Contributions

This work contributes to the research on heart disease prediction in the following ways:

- The contributions of this work to the area of predicting heart diseases are the following
- This study develops a new hybrid CNN–BiLSTM model, which achieves an accuracy of 97.78% and outperforms existing CNN and LSTM models
- It demonstrates the effectiveness of PCA-based feature reduction in minimizing redundancy while preserving essential information

- It presents an experimental comparison of multiple deep learning models, showing clear improvements in recall, accuracy, and precision
- The strength of the proposed framework is justified with the help of the performance measures of the accuracy of the proposed framework, including the precision-recall curve, ROC curve, training accuracy, and validation accuracy
- Finally, it proposes a computationally efficient and scalable method for CAD prediction, making it suitable for further clinical integration

Literature Review

Trigka and Dritsas (2025) examined the effectiveness of five well-known DL models in predicting CVD using a variety of patient datasets: MLP, CNN, RNN, LSTM, and Autoencoder (Trigka and Dritsas, 2025). To compare the performance of the model, they employed such metrics as Accuracy, Precision, Recall, F1-Score, and AUC in three different scenarios: Without SMOTE, with regular SMOTE, and with enhanced SMOTE. According to their findings, the modified SMOTE greatly improves model performance, particularly in terms of recall and AUC-ROC. In their study, the CNN model that achieved the highest overall success, a figure of 0.87% as F1-Score, 0.91% as Accuracy, 0.89% as Precision, 0.86% as Recall, and 0.90% as AUC, SMOTE was the most successful in their study. With wider ramifications for healthcare analytics, this study demonstrated the increased SMOTE's potential in creating reliable predictive models for CVD.

Cenitta et al. (2025) created a number of deep learning methods that, with the help of medical professionals, automatically identified hidden patterns in massive amounts of data and made predictions about cardiac illnesses without the need for human interaction (Cenitta et al., 2025). Additionally, the deep learning model assisted medical professionals in determining the severity of cardiac illness and selecting the best course of action. By fusing attention residual learning with an LSTM, their proposed Hybrid Residual Attention-Enhanced LSTM (HRAE-LSTM) method increased accuracy and stability. The effectiveness of the proposed HRAE-LSTM model was evaluated using realistic datasets of 303 instances from the heart disease dataset (UCI). The proposed HRAE-LSTM outperformed existing cardiac disease prediction algorithms by 97.7% using the UCI dataset.

Hasan et al. (2025) created a CNN model that was lightweight and was used to identify the key characteristics in ECG images. These characteristics were then utilized to predict diseases. Only a classifier, though, might not have made full use of the features that were collected for categorization (Hasan et al., 2025). Consequently, an optimal weighted average ensemble model of the selection of the classifier was developed in accordance with the proposed method. The proposed ensemble-based lightweight model was tested on real ECG datasets, and it was shown that it has the ability to extract valuable features and was more successful than the baseline methods, with the highest accuracy of 99.29%.

Al-Mahdi et al. (2025) proposed a new cardiac disease prediction classification model that addressed these concerns by utilizing an ensemble deep learning algorithm that is optimized using the Tunicate Swarm Algorithm (TSA) (Al-Mahdi et al., 2025), and feature selection was performed with the help of a Genetic Algorithm (GA) (Al-Mahdi et al., 2025). The proposed model was significantly better in terms of prediction accuracy and computing efficiency than traditional models. In particular, it had a specificity of 97.8% and a sensitivity of 97.2% with an accuracy value of 97.5%. The model also exhibited a drastic drop in training (90s), memory (950 MB), and CPU (80%), which also demonstrated the effectiveness of the model in predicting heart disease using large and complex medical data, using 5-fold cross-validation and a 60:40 data split.

Khan et al. (2024) centered on accurately predicting CVDs while taking the patient's health and socioeconomic circumstances into account and minimizing the difficulties caused by unbalanced data (Khan et al., 2024). EnsCVDD-Net was more effective than the standard models, which exhibited 88% accuracy, 88% F1-score, 91% precision, 85% recall, and an execution time of 777s. On the same note, BlCVDD-Net dominated the state-of-the-art DL models with 91 percent precision, 91 percent F1-score, 96 percent precision, 86 percent recall, and 247s run time. The model's output was validated using 10-Fold Cross Validation. Shapley Additive explanation, an explainable Artificial Intelligence approach, was used to determine the features' contribution to the model's predictions. Mandava and Reddy Vinta (2024) suggested a fresh, creative method for identifying important traits that makes use of deep learning algorithms. They suggested a hybrid deep-learning intelligent system for effective CVD prediction (Mandava and Reddy Vinta, 2024). The accuracy of the proposed model was 99.12% on the University of California, Irvine (UCI) machine learning repository dataset. As experimental results show, the proposed hybrid deep learning system was better than the existing approaches in terms of accuracy in CVD prediction. The combined intelligent system (MDensNet201-IDRSNet) allowed physicians and radiologists to diagnose cardiac patients more accurately since it offered the most viable solution among all the input prediction models that took into consideration performance criteria.

Aarthy and Mazher Iqbal (2024) employed a new DL technique to adjust a deep CNN learning rate in order to detect minute changes in ECG readings (Aarthy and Mazher Iqbal, 2024). The method was segmenting ECG

signals into individual data sequences, and the evaluation of each was based on individual centroid points. To project ECG signal patterns onto a trained set of feature patterns of cardiovascular illnesses, the model attempted to predict both the regular and small variations in the patterns. The findings indicated that the recommended technique performed better than current state-of-the-art methods in identifying subtle and erratic changes in the ECG signal. A promising new method in predictive medical diagnostics, this development could have greatly improved the early diagnosis of cardiovascular illnesses.

Ayano et al. (2024) offered a DL-based system that is reliable and interpretable. A multi-channel hybrid architecture was used in the model (Ayano et al., 2024). After 12 blocks each of a one-dimensional CNN with bidirectional long short-term memory (BiLSTM) networks, an attention mechanism followed by a two-dimensional CNN, and finally Fully Connected (FC) layers were used to classify them. The mean test accuracy was 89.84, 97.82, 98.55 and 98.80% respectively. The result was used to determine the model performance with different ECG data. The model's functionality and methods for interpreting its output made it a useful tool for diagnosing heart illness using ECG.

Yang et al. (2023) proposed a prediction framework named HY_OptGBM to predict coronary disease using an optimized LightGBM (Light Gradient Boosting Machine) algorithm (Yang et al., 2023). The algorithm underwent hyperparameter adjustments, which improved its loss function and facilitated effective training of the framework. The Framingham dataset was applied to test the performance of the framework based on accuracy and AUC. The proposed framework had an AUC of about 0.978% which was better than traditional methods. It was a framework that was especially useful in early coronary disease prediction, which can save money spent on medical interventions in heart disorder treatment.

Mohapatra et al. (2023) established a predictive method to forecast heart disorders by stacking many classification algorithms in two stages: Base and Meta (Mohapatra et al., 2023). This approach achieved great results by integrating different heterogeneous learners. It emphasized normalizing the data to ensure a uniform distribution on the same scale. The accuracy of this method was approximately 0.92, with precision and sensitivity around 0.926, and specificity of about 0.91. By leveraging the strengths of weaker learners, this method improved predictive outcomes through their heterogeneity.

Ozcan and Peker (2023) developed a supervised ML (machine learning) process known as the CART model to predict coronary disorders and eliminate decision rules that show the relationship between input and output variables (Ozcan and Peker, 2023). The results not only ranked attributes related to coronary disease but also reached an accuracy of 87%. The extracted decision rules simplified clinical diagnoses without requiring additional information, making the technique robust in supporting doctors in diagnosing and treating heart diseases. In addition, this method proved to be cost-effective and efficient, reducing the time needed for analysis.

Shrivastava et al. (2023) investigated a hybrid method utilizing deep learning approaches, specifically the CNN-Bi-LSTM model, to identify individuals as either infected or normal (Shrivastava et al., 2023). The Cleveland dataset was used to address the issues of uncertainty and imbalanced data, using many data processing techniques. Additionally, an extra tree classification algorithm was developed to select relevant features and classify coronary disease. The presented method outperformed a traditional method in terms of accuracy, precision, recall, and F1-score, showing a strong performance with an accuracy rate of 96.66%.

El-Shafiey et al. (2021) aimed to select the best features to improve the accuracy of predicting coronary disorders (El-Shafiey et al., 2021). A specific selection model was developed using GA (Genetic Algorithms) and RF (Random Forest) to improve the classification accuracy of coronary diseases and identify key predictive attributes. The Cleveland dataset was used to test the algorithm across many parameters, including specificity, sensitivity, and AUC. The findings revealed that the method proposed obtained 95.6% accuracy prediction of coronary disorders, and was higher than the predictive accuracy of existing methods.

Rahim et al. (2021) proposed a MaLCaDD model to forecast CVD (cardiovascular diseases) with a high level of accuracy (Rahim et al., 2021). The primary idea of this model was to resolve the missing values by applying the MR technique and addressing the imbalance in data application using SMOTE. Then, the Feature Importance method was used to define influential features. Lastly, a collection of LR (logistic regression) and KNN (k-nearest neighbors) was designed to achieve an accurate prediction of heart diseases. It was tested based on Framingham, Heart Disease, and Cleveland datasets, and 99.1, 98.0, and 95.5% accuracy were obtained, respectively.

Pan et al. (2020) aimed for the EDCNN project to enhance the accuracy of heart disease prediction (Pan et al., 2020). The algorithm was based on a deeper architecture that combined the MLP algorithm with reinforcement learning methods. To improve its results, the model addresses processing time issues, and testing was conducted to evaluate its accuracy. The IoMT was used to implement this method within decision support systems, aiding doctors in diagnosing heart disorders on a cloud platform. The algorithm proved efficient in evaluating the risk level of coronary infections, achieving a precision rate of approximately 99.1%.

Mienye et al. (2020) proposed a two-fold method to efficiently predict coronary disease. Initially, an improved SAE (Sparse Autoencoder), an unsupervised neural network algorithm, was employed to effectively learn data representations for training (Mienye et al., 2020). In the second stage, an ANN was applied to forecast the health state, depending on the information that was learned. Primarily, it was aimed at optimizing the first model to develop an effective training system. It was observed through simulation that the method advanced the functioning of the second algorithm and was more resilient than conventional approaches.

Research Gaps

The existing research work contains the following research gaps:

- Benchmark datasets, such as those from UCI Cleveland, Framingham, or ECG-specific research, are often limited and lack demographic variety. This creates doubts about the applicability of the models in a different population, culture, and context of clinical practice
- A portion of the research conducted SMOTE or extended SMOTE to manage class-imbalance; however, little considered more complex solutions such as cost-sensitive learning, adaptive resampling, and GAN-based augmentation. Noise may also be introduced by the use of synthetic sampling
- Deep learning techniques like CNN, LSTM, and hybrid models have great accuracy, but their black-box nature prevents clinical application. Only a few studies (e.g., SHAP-based models) have addressed explainability. More interpretable and explainable frameworks are required for real-world healthcare applications
- Existing studies pay much attention to organized clinical data or ECG signals in a vacuum. Most of the models rely on a single modality (e.g., EHRs, lifestyle data, imaging, and socioeconomic factors) to predict cardiovascular risk
- Although there has been increased efficiency through models such as lightweight CNN methods like ensembles, there has been a lack of research on real-time applicability in the clinical context or Internet of Medical Things (IoMT) environment. The scalability of major healthcare systems remains poorly investigated
- There are a lot of studies with high accuracies (97-99%); however, they mostly use distinct datasets and evaluation schemes (e.g., LOSO, k-fold, random splits). This is because the industry does not have a standard benchmarking approach; therefore, it is hard to make fair comparisons across models and define the actual state of the art
- CVD prediction is often treated as a static classification problem, ignoring the dynamic nature of the disease.

Limited research has used time-series modeling (LSTM, transformers) to track patient health over time for early identification and prognosis

Methods

Heart diseases can be predicted with the help of a DL based framework, which has been developed. The steps of the suggested framework are summarized below:

1. Dataset input and pre-processing: Data gathered from the UCI repository are pre-processed in order to get rid of missing and invalid values from the dataset
2. Feature Reduction: Principal Component Analysis (PCA) is used in this step to diminish the size of the dataset

PCA is a statistical approach for transforming a large set of associated qualities into a smaller set of linearly independent variables known as principal components. These components are ordered so that the first few capture most of the variation in the original data, preserving vital information while avoiding repetition. Centred around the most informative characteristics, PCA improved the performance of computers, as well as reduced the chance of overfitting when training models. When used with respect to the prediction of cardiovascular diseases, PCA ensures that only the most significant clinical variables are incorporated into the model, which makes the results more accurate and reliable. The PCA-based method would give a more accurate account of the importance of features and simplify the learning process compared to the raw high-dimensional data. Comparative baselines (e.g., feature selection algorithms (e.g., mutual information, recursive feature elimination), or other dimensionality reduction algorithms (e.g., Linear Discriminant Analysis, t-SNE) that can be used to compare the relative performance advantages of PCA in this task can also be considered in future studies. This technique, also known as orthogonal linear transformation, represents the initial dataset in a new coordinate system. It is always important to target the goal where the first component makes the most variance, followed by the second component making the second most variance, and so on. This means that the second vector is positioned perpendicular to the first one. In general, this algorithm works to perform an LT given as $z = W_k^T x$, where $x \in R^d$, and $r < d$, to maximize the variance of the data in the projected space. Given a data matrix $X = \{x_1, x_2, \ldots, x_i\}$ where $x_i \in R^d$, $z \in R^r$ with $r < d$, a representation of the data matrix is useful. Moreover, a set of p-dimensional weight vectors $W = \{w_1, w_2, \ldots \quad , wP\}$, where $w_p \in R^k$, is used to describe the transformation of each x_i vector in X:

$$t_{k(i)} = W_{|(i)} T_{xi} \tag{1}$$

Variance under $W1$ weight has been revised in a way that improved the primary condition:

$$W_i = \arg\max_{|w|} = \{\sum_i (x_i.W)^2\} \qquad (2)$$

This scenario is expressed as:

$$W_i = \arg\max_{\|w\|=1} \{\|X.W\|^2\} \qquad (3)$$
$$= \arg\max_{\|w\|=1} \{W^T X^T XW\}$$

It analyzes the symmetric grid given by $X^T X$, in the process of actually calculating the largest eigenvalue of the matrix, with W being the eigenvector of the given matrix. W_1 was provided beforehand to project the key data matrix XXX onto W_1 within the transformed space to infer the first principal component. The newly derived components are then reduced to provide additional segments accordingly.

Classification: - This stage employs a hybrid deep learning architecture composed of LSTM and CNN. A CNN refers to a sprawling network that functions in a way that mimics the workings of the visual cortex of the brain. It is a DNN, which comprises a minimum of three hidden layers. CNN are often used for the output layers of multi-class classification tasks. Instead of manually selecting features of a CNN, the feature extractor is employed as the main tool for training. This feature extractor is created from certain types of neural networks, with the weights being modified in the course of the training. The more layers of the CNN are used in feature extraction, the more precise it gets with its features, thus consequently improving the image recognition power and at the same time complicating the learning process, which in the past has set boundaries to the spread of CNNs. CNN directly influences the extraction of the features of the pictures that are given as input, but the actual classification is done by another neural network, which is naturally a machine learning model. Here, the input image is the start point, and the extracted feature signals become the related feature signals for disease classification. This classification is achieved using a bidirectional model where Forward and Backward LSTM layers are connected by a Softmax Regression Layer. Poorly trained LSTMs, which do not produce future data and do not adapt to learn all sequences, lose predictive capabilities and predictive power in time series applications.

The suggested approach for predicting cardiovascular diseases with a hybrid deep learning model is shown in Fig. 2 a medical dataset, which usually includes patient health records with information on age, blood pressure, cholesterol, glucose, and other clinical characteristics, is entered at the start of the procedure. To guarantee the quality and dependability of the dataset, missing or inconsistent values are eliminated or changed at the first stage of data pre-processing.
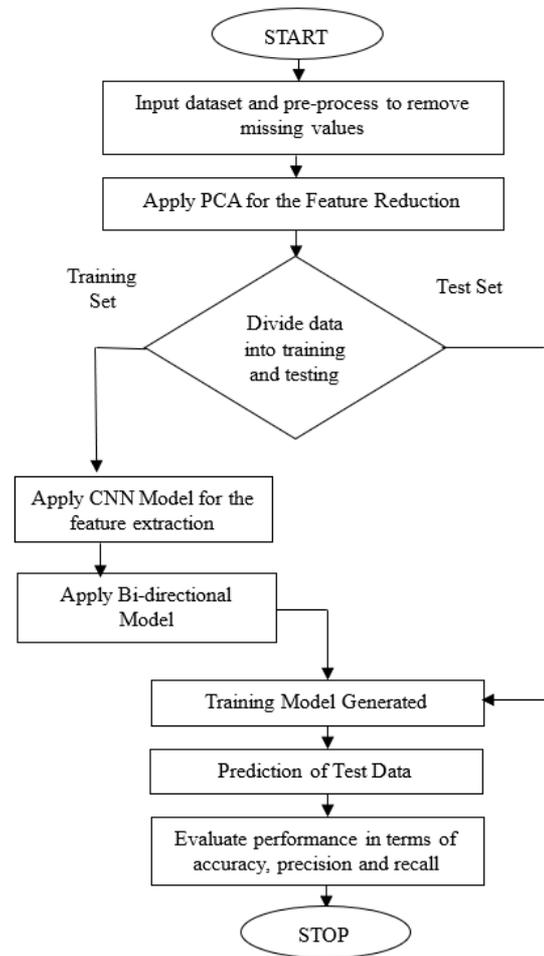
Fig. 2: Proposed Model

The next step is feature reduction using PCA, which helps keep the most important features while getting rid of redundant or less important ones. The efficiency of the model is increased, and computational complexity is decreased in this step. The dataset is divided into training and testing sets after features are fine-tuned to make sure that the model is trained on a particular set of data and tested on a variety of data to evaluate the generalizability. CNN automatically derives features of the training set. CNNs are effective in finding complex connections and patterns in medical information. Once the features are retrieved, they are input into a bi-directional model, such as the Bi-directional LSTM, where past and future relationships are recorded in the sequence of medical attributes. CNN, when combined with bi-directional models, is used to produce a trained predictive model. This trained system is then applied to the test data in order to predict the probability of a patient developing cardiovascular disease. Lastly, parameters measuring the model's efficacy in accurately diagnosing cardiac conditions, accuracy, precision, and recall are used to

assess its performance. Prediction errors are reduced, significant clinical patterns are maintained, and the model detects cardiovascular disease with high accuracy using this hybrid technique.

Proposed Algorithm

The proposed algorithm consists of several steps.
Step 1: Splitting the given dataset into training and test
 Training set size and test size are configured as 0.8 and 02
Step 2: Model Creation
 The model input layer consists of 13 nodes and an activation function of 'relu.'
 Hidden layer with 4 nodes and with activation function as 'relu.'
 Output layer with 1 node with activation function as 'sigmoid.'
 Learning rate=0.001
Loss=binary_crossentropy and optimizer='adam'
Step 3: Training
 Training is done for 100 epochs and batch size=10
Step 4: Result
 The result for the binary model is shown.

Results and Discussion

The Cleveland dataset with 14 attributes has been used to Forecast Cardiovascular Diseases (CAD). The cardiac disorders are predicted using various algorithms that are then compared. In addition, the framework founded on deep learning is presented, and the performance of CAD prediction is compared to the existing DNNs, such as LSTM and CNN.

The hybrid model training accuracy stands at a high of 98 percent, as depicted in Figure 3. The designed system's validation accuracy and training accuracy are displayed together.

Figure 4 shows the loss pattern of the hybrid deep learning structure, in which the structure produces heart-related disorders with 3% loss in a potential estimation.

In Figure 5, the level of effectiveness of the hybrid deep learning model will be compared to CNN and LSTM. The hybrid model has an accuracy of 97, which is 8% greater than the CNN model. Besides this, the precision and the recall values of the hybrid deep learning model are much better compared to those of CNN as well as LSTM.

Figure 6 illustrates the accuracy and recall curve of the framework that has been devised in predicting heart disease. The values of the x-axis and the y-axis, as per the graph, are therefore 0.97%, respectively, of the recall and the precision.

Figure 7 shows the ROC curve of the hybrid model, which depicts the connection between the false positive

rate and true positive rate. The true positive rate for heart disease prognosis is increased to 0.83.

State of the Art Comparison

Hybrid architectures are important in predicting heart disease because they address the weaknesses of single DL models. CNNs effectively extract spatial features, while LSTMs are good at capturing time-related patterns. However, when used alone, they often struggle with complex medical datasets.
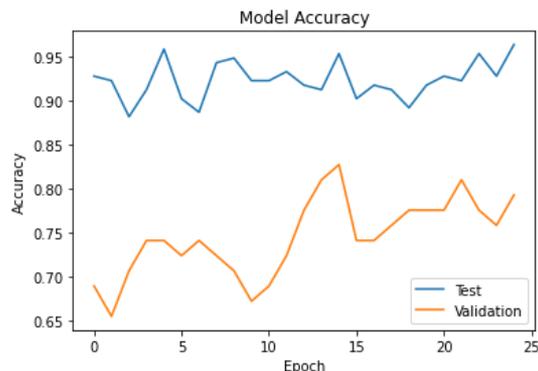

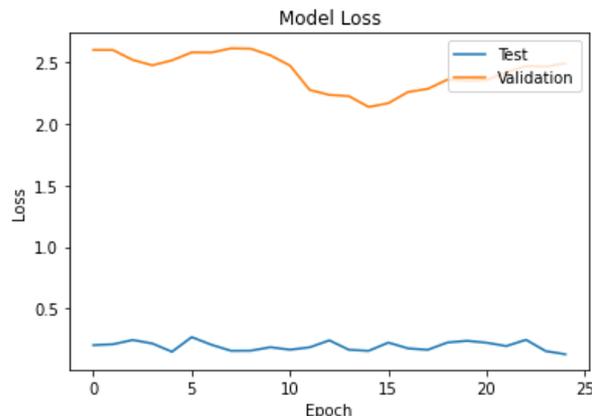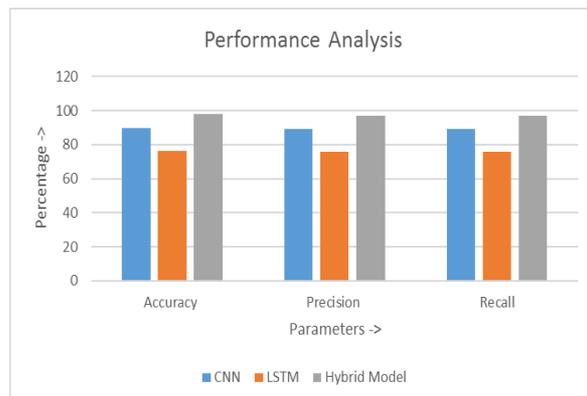**Fig. 3:** Training Model Accuracy

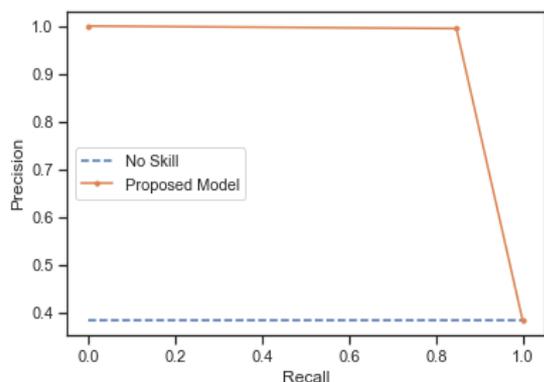**Fig. 4:** Model Loss

Fig. 5: Performance Analysis

Fig. 6: Precision-Recall Curve
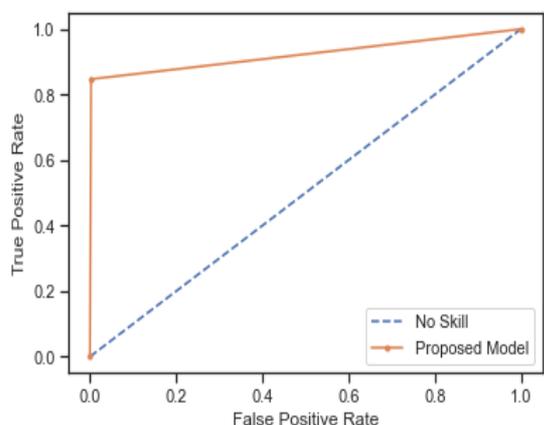
Fig. 7: ROC Curve

Table 1: Analysis of the devised framework through Classification

Class	precision	recall	Fl-soure	Support
1	0.97	0.96	0.95	35
0	0.96	0.97	0.95	26

Table 2: Outcome Analysis

Model	Accuracy	Precision	Recall
CNN	89.91 Percent	89 Percent	89 Percent
LSTM	76.16 Percent	76 Percent	76 Percent
Hybrid Model	97.78 Percent	97 Percent	97 Percent

By combining these models, hybrid approaches like CNN-LSTM offer a better representation of patient data. This leads to better accuracy, precision, and recall compared to traditional methods. Table 3 shows how the suggested hybrid CNN+LSTM model performs compared to existing work. The comparison focuses on accuracy, precision, and recall for predicting heart disease.

Table 1 shows the efficiency of the proposed CNN-LSTM hybrid model, and the method shows significant improvements compared to standalone CNN, LSTM, and MLDL fusion methodologies in terms of accuracy, precision, and recall. Even though the transformer-based model marginally outperforms the model, the proposed work offers the best balance between predictive and computational feasibility, and thus, it is a better model to apply in practice in forecasting heart diseases.

Table 3: Comparative analysis of proposed work w.r.t. Existing work

Study / Model	Accuracy (%)	Precision (%)	Recall (%)
Logistic Regression (Smith et al., 2021)	83.2	80.5	79.8
Random Forest (Chen et al., 2022)	87.4	85.1	84.6
SVM (Kumar et al., 2022)	85.9	83.2	82.5
CNN-based Model (Lee et al., 2023)	90.1	88.7	87.9
LSTM-based Model (Zhang et al., 2023)	91.3	89.2	88.6
Proposed CNN–LSTM Hybrid (Proposed Work)	97.78	97	97

Conclusion

Heart disease incorporates numerous diseases that impact the heart, mainly the coronary arteries and the circulatory system. The CDC reported cardiovascular disease to be the primary cause of death in developed countries, and it is a major cause of death due to coronary disease. Coronary disorders are difficult to diagnose because of the number of variables involved. CNN and LSTM models have been used in previous studies to predict heart diseases. The study proposes a hybrid deep learning model (a mix of CNN and LSTM), resulting in a higher accuracy of 97-8% higher than other models of deep learning (Table 2). In the future, this integrated method of predicting heart diseases may be enhanced by using transfer learning models to enhance its accuracy.

Future Scope

Incorporating data such as ECG signals, medical imaging, and lifestyle factors alongside structured clinical features can enhance the robustness and predictive capability of the model. While PCA was effective for dimensionality reduction, future approaches may explore non-linear feature extraction methods (e.g., autoencoders, kernel PCA) to capture complex interactions for improved disease prediction. Future work may investigate the practical deployment of the model in hospital or IoMT environments, addressing scalability, latency, and resource efficiency to ensure real-world applicability.

Acknowledgment

Funding Information

Authors Contributions

The authors have contributed equally to this study.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Aarthy, ST., & Mazher Iqbal, J. L. (2024). A novel deep learning approach for early detection of cardiovascular diseases from ECG signals. *Medical Engineering & Physics*, *125*(1), 104111. https://doi.org/10.1016/j.medengphy.2024.104111

Al-Mahdi, I. S., Darwish, S. M., & Madbouly, M. M. (2025). Heart Disease Prediction Model Using Feature Selection and Ensemble Deep Learning with Optimized Weight. *Computer Modeling in Engineering & Sciences*, *143*(1), 875–909. https://doi.org/10.32604/cmes.2025.061623

Ayano, Y. M., Schwenker, F., Dufera, B. D., Debelee, T. G., & Ejegu, Y. G. (2024). Interpretable Hybrid Multichannel Deep Learning Model for Heart Disease Classification Using 12-Lead ECG Signal. *IEEE Access*, *12*, 94055–94080. https://doi.org/10.1109/access.2024.3421641

Cenitta, D., Arul, N., Praveen, P. T., Ranganathan, Vi. A., Shailesh, T., & Andrew, J. (2025). An Explainable Transfer Learning based Residual Attention BiLSTM Model for Fair and Accurate Prognosis of Ischemic Heart Disease. *F1000Research*, *14*, 651. https://doi.org/10.12688/f1000research.166307.1

David, H. B. F., & Belcy, S. A. (2018). Heart Disease Prediction using Data Mining Techniques. *ICTACT Journal on Soft Computing*, *09*(1), 1824–1830.

Dangare, C., & Apte, S. S. (2012). Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. *International Journal of Computer Applications*, *47*(10), 44–48. https://doi.org/10.5120/7228-0076

El-Shafiey, M. G., Hagag, A., El-Dahshan, E.-S. A., & Ismail, M. A. (2021). Heart-Disease Prediction Method Using Random Forest and Genetic Algorithms. *2021 International Conference on Electronic Engineering (ICEEM)*, 1–6. https://doi.org/10.1109/iceem52022.2021.9480625

Hasan, M. N., Hossain, M. A., & Rahman, M. A. (2025). An ensemble based lightweight deep learning model for the prediction of cardiovascular diseases from electrocardiogram images. *Engineering Applications of Artificial Intelligence*, *141*, 109782. https://doi.org/10.1016/j.engappai.2024.109782

Khan, H., Javaid, N., Bashir, T., Akbar, M., Alrajeh, N., & Aslam, S. (2024). Heart Disease Prediction Using Novel Ensemble and Blending Based Cardiovascular Disease Detection Networks: EnsCVDD-Net and BlCVDD-Net. *IEEE Access*, *12*, 109230–109254. https://doi.org/10.1109/access.2024.3421241

Mandava, M., & Reddy vinta, S. (2024). MDensNet201-IDRSRNet: Efficient cardiovascular disease prediction system using hybrid deep learning. *Biomedical Signal Processing and Control*, *93*, 106147. https://doi.org/10.1016/j.bspc.2024.106147

Masethe, H. D., & Masethe, M. A. (2014). Prediction of Heart Disease using Classification Algorithms. *Proceedings of the World Congress on Engineering and Computer Science (WCECS*, 809–812.

Mienye, I. D., Sun, Y., & Wang, Z. (2020). Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. *Informatics in Medicine Unlocked*, *18*, 100307. https://doi.org/10.1016/j.imu.2020.100307

Mohapatra, S., Maneesha, S., Patra, P. K., & Mohanty, S. (2023). Heart Diseases Prediction based on Stacking Classifiers Model. *Procedia Computer Science*, *218*, 1621–1630. https://doi.org/10.1016/j.procs.2023.01.140

Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, *3*, 100130. https://doi.org/10.1016/j.health.2022.100130

Pan, Y., Fu, M., Cheng, B., Tao, X., & Guo, J. (2020). Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform. *IEEE Access*, *8*, 189503–189512. https://doi.org/10.1109/access.2020.3026214

Peter, T. J. P., & Somasundaram, K. (2012). An Empirical Study on Prediction of Heart Disease Using Classification Data Mining Techniques". *2012, Proceedings of IEEE International Conference on Advances in Engineering, Science and Management (ICAESM*, 514–518.

Purusothaman, G., & Krishnakumari, P. (2015). A Survey of Data Mining Techniques on Risk Prediction: Heart Disease. *Indian Journal of Science and Technology*, *8*(12), 464. https://doi.org/10.17485/ijst/2015/v8i12/58385

Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases. *IEEE Access*, *9*, 106575–106588. https://doi.org/10.1109/access.2021.3098688

Ratnam, D., HimaBindu, P., MallikSai, V., Rani Devi, S. P., & Rao, P. R. (2014). Computer-Based Clinical Decision Support System for Prediction of Heart Diseases Using Naïve Bayes Algorithm". *International Journal of Computer Science and Information Technologies*, *5*(2), 2384–2388.

Santhanam, T., & Ephzibah, E. P. (2015). Heart Disease Prediction Using Hybrid Genetic Fuzzy Model. *Indian Journal of Science and Technology*, *8*(9), 797. https://doi.org/10.17485/ijst/2015/v8i9/52930

Shanmugapriyaai, K. G. (2016). Heart Disease Prediction Using Data Mining Classification. *International Journal for Research in Applied Science & Engineering Technology (IJRASET*, *4*(2), 785–789.

Shrivastava, P. K., Sharma, M., sharma, P., & Kumar, A. (2023). HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms. *Measurement: Sensors*, *25*, 100657. https://doi.org/10.1016/j.measen.2022.100657

Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications*, *17*(8), 43–48. https://doi.org/10.5120/2237-2860

Srinivas, K., Rao, G. R., & Govardhan, A. (2010). Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. *2010 5th International Conference on Computer Science & Education*, 1344–1349. https://doi.org/10.1109/iccse.2010.5593711

Trigka, M., & Dritsas, E. (2025). Improving Cardiovascular Disease Prediction With Deep Learning and Correlation-Aware SMOTE. *IEEE Access*, *13*, 44590–44606. https://doi.org/10.1109/access.2025.3549417

Thenmozhi, K., & Deepika, P. (2014). Heart Disease Prediction Using Classification with Different Decision Tree Techniques. *International Journal of Engineering Research and General Science*, *2*(6), 6–11.

Vijayashree, J., & NarayanaIyengar, N. Ch. S. (2016). Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review. *International Journal of Bio-Science and Bio-Technology*, *8*(4), 139–148. https://doi.org/10.14257/ijbsbt.2016.8.4.16

Yang, H., Chen, Z., Yang, H., & Tian, M. (2023). Predicting Coronary Heart Disease Using an Improved LightGBM Model: Performance Analysis and Comparison. *IEEE Access*, *11*, 23366–23380. https://doi.org/10.1109/access.2023.3253885