# Hybrid model of convolutional neural network and long short term memory for heart disease prediction

**Shubham Gupta, Pooja Sharma**
Department of Computer Science, Rayat-Bahra University Punjab, Mohali, India

**Article Info**

***Article history:***

Received May 28, 2023
Revised Oct 5, 2023
Accepted Oct 25, 2023

***Keywords:***

Convolutional neural network
Deep learning
Feature engineering
Heart disease
Long short-term memory

**ABSTRACT**

Data mining is a process that assists in uncovering meaningful data from large, disorganized datasets. This research is being conducted to predict heart disorders by using available data to make predictions for the future. The approach is carried out in several stages, such as pre-processing the data, extracting relevant features, and classifying the data. all of these steps are essential for predicting heart disease. The deep learning models is already proposed by the researches for the heart disease prediction. This work introduces a hybrid deep learning model that combines convolutional neural network (CNN) and long short-term memory (LSTM) to predict heart disease. The proposed model has been implemented in python, and its accuracy, precision, and recall have been evaluated.



***Corresponding Author:***

Shubham Gupta
Department of Computer Science, Rayat-Bahra University Punjab
Mohali, India
Email: guptashubham96222@gmail.com

## 1. INTRODUCTION

Recently, heart disease leads to cause life loss among the people of all ages. Therefore, an essential task of healthcare field is to enhance the way of predicting heart attacks based on different machine learning methods. This disease is consisted of diverse disorders that lay impact on heart and veins of individuals. The symptoms are found different, and varied according to the kind of a coronary disease [1]. To recognize and diagnose the cardiovascular disease (CVD) is a complex task, and the expert doctors require more information and experience to accomplish this task [2] HDPM: an effective heart disease prediction model for a clinical decision support system. Diverse component related to cardiac disorder are age, diabetes, smoking, overweight, and bad eating habits. The heart disease is occurred and maximized due to various factors and metrics. The fundamental factor to diagnose this disease accurately and exactly, is the earlier knowledge and information, captured from related pathological events.

A number of hospitals deploy a management software to monitor the clinical and patient's data. This kind of system attains much attention and helps in producing huge volume of information regarding patient. The process to mine electronic health records (EHRs) [3], is utilized for capturing the medical data related to all aspects of patient care (including analysis, treatment, and laboratory test). Hence, the comprehensive investigations and detection of clinical evidence becomes easy on an unparalleled scale. The field of predicting heart disease makes the implementation of machine learning. There are significant possibilities for enhancing treatment results through the analysis of data, which can help identify patterns and create prognostic systems for personalized assessment, forecast, and therapy [4]. Machine learning holds significant promise in advancing medical solutions across the entire spectrum, spanning from initial discovery and prediction to informed decision-making. Machine learning has the ability to detect disease patterns and notify healthcare providers of any irregularities by examining computerized medical records. Figure 1 depicts a heart disease prediction framework [5].

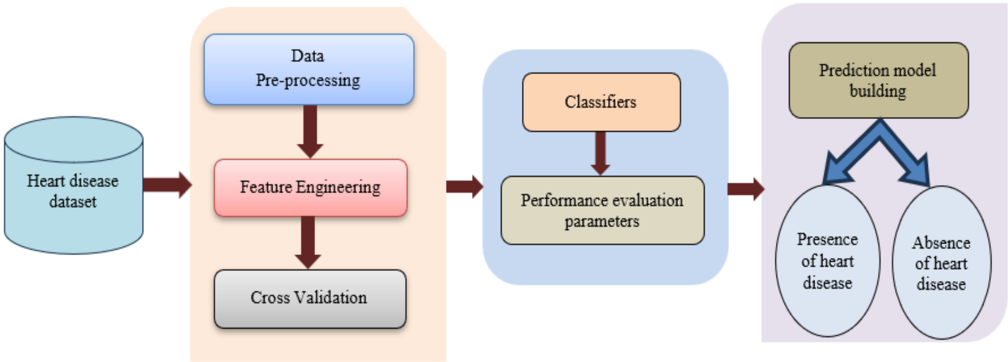


Figure 1. Heart disease prediction framework [4]

The initial stage is to gather public databases for heart disease diagnosis that have been provided by various healthcare organizations. The next stage of data processing is transforming unprocessed data into useful patterns [6]. This refers to a critical stage in machine learning since the amount and quality of meaningful information that can be obtained from the input directly influence the model's ability for learning. Features are selected for further processing after the pre-processing of the data is completed [7]. Generally speaking, feature engineering is essential for building the foundation for classifiers. In order to create successful predictive frameworks, the primary goal of this step is to decrease the number of features that are fed into a classifier [8]. A crucial step in machine learning is validating the prediction model.

The leave-one-subject-out (LOSO) cross validation technique is well recognized. This validation technique divides the remaining individuals into groups for model training and uses one instance as test data. In order for a person to be classified as normal, heart disease must be predicted of the target person [9]. In machine learning, heart disease prediction can be viewed as a classification or clustering problem. The obtained features are grouped in the fourth stage of prediction. A data mining task requires classification, which is of utmost importance. Separate classes and features must be created from the abstracted data. The type of partitioning depends on the size and content of the index [10]. In order to improve prediction accuracy for hidden specific objects in data class, classification aims to identify and classify the class. The prediction for each class model is not achieved until the data class objects are separated for each class framework.

Several performance indicators are utilized to assess the performance of classifiers during the evaluation stage [11]. The confusion matrix is employed to compute these indicators. Various performance indicators such as accuracy, sensitivity, specificity, recall, F1-score, Matthew correlation-coefficient (MCC), AUC-score, and receiver operating characteristics (ROC) curve are included [12]. For the purpose of predicting heart disease, the predictive framework is developed using traditional, statistical, and machine learning algorithms. The support vector machine (SVM) is a linear classifier that has the widest gap in the feature space and is used for binary classification [13]. It uses interval maximization as a learning strategy, which may serve as a regularized loss function minimization issue as well as a convex quadratic problem to solve. The separation hyperplane must be solved with the widest geometric interval possible in order for SVM to successfully split the training set of data [14].

The hyperplane that has the greatest geometric distance between its points of separation is identified, even if there are an infinite number of them for the linearly separable data set. A reliable technique for identifying the data is K-nearest neighbors (K-NN) [15]. This algorithm makes the assumption that similar objects are nearby and may be found by looking at their K-NN. The training examples are the space vectors with multidimensional elements and their class labels. The K-NN algorithm is trained using testing sample feature vectors and class labels [16]. In the categorization process, k is used to define a user-defined constant. In order to identify an unlabelled vector, a label is applied to the well-known sample out of k samples. The k is chosen by the dataset. The impact of noise may generally be reduced by classifying the data and introducing less defined class borders with higher k values. Due to majority or average voting, the maximizing k reveals the prediction accuracy [17]. This algorithm is simple to construct and is considered to be least complex.

Currently, the decision tree algorithm is widely used and relies on the concept of information entropy. The idea is to create and prune trees until you find the best feature. To construct the decision tree incrementally, the information gain for each node is utilized in the selection process [18]. At the outset, the information gain for all potential features is calculated, commencing from the root node. The feature that exhibits the maximum information gain is subsequently preferred as the node feature. In the event that none of the features meet the selection criteria and the information gain for all features is exceedingly low, the final decision tree model may be formed [19].

The categorization of the unknown test data by the created decision tree is frequently not as accurate as the classification of the training data, which causes an over fitting phenomena. As a result, it is important to process paper cuts and simplify the decision tree that was formed. In order to simplify the classification tree model and minimize the overall loss function of the decision tree, certain subtrees are pruned or removed from the constructed tree, and the root node is considered as a fresh leaf node [20]. This method helps to improve the comprehensibility of the decision tree and optimize its efficiency.

## 2. RELATED WORK

Mienye *et al.* [21] recommended a two-fold method for predicting the coronary disease in efficient way. First of all, an unsupervised neural network algorithm called an enhanced sparse autoencoder (SAE) algorithm was adopted for learning the effective illustration of the data used to train the system. The health state of the learned information was predicted in the following phase using an artificial neural network (ANN). The major emphasize was on optimizing the initial model for training an effectual system. According to the simulation result, the recommended method was applicable for improving the efficacy of second algorithm and its robustness was proved over the traditional methods.

Yang *et al.* [22] suggested a predictive framework known as HY_OptGBM for predicting the coronary disease for which the optimized light gradient boosting machine (LightGBM) algorithm was implemented. This algorithm was optimized after adjusting its hyper-parameters. Moreover, these hyper-parameters assisted in enhancing the loss function (LF) and training the suggested framework. Framingham dataset was executed for simulating the suggested framework with regard to accuracy and area under curve. The suggested framework attained area under curve around 0.978 as compared to the traditional methods. This framework was useful for predicting the coronary disease at initial stage and alleviating the costs related to the medical cure of patients of heart disorders.

Ashri *et al.* [23] focused on selecting the optimal attributes for maximizing the accuracy of predicting the coronary disorder. An fuzzy system (FS) model was introduced on the basis of genetic algorithm (GA) and random forest (RF), for enhancing the accuracy of classifying the coronary diseases, and determining the optimal attributes for predicting the heart-disease. Cleveland dataset executed to simulate the introduced algorithm with regard to different parameters such as specificity, sensitivity, and AUC. The experimental outcomes revealed the efficiency of the introduced algorithm concerning accuracy of 95.6% for predicting the coronary disorders. Additionally, the supremacy of this algorithm was proved against the existing techniques.

Pan *et al.* [24] projected enhanced deep learning convolutional neural network (EDCNN) model in order to assist and maximize the efficacy for predicting the heart disease. This algorithm considered a deeper model for covering multi-layer perceptron (MLP) algorithm with reinforcement learning (RL) methods. The effectiveness of the projected algorithm was affected by mitigating the attributes concerning processing time, and the testing was performed for analyzing the accuracy. Internet of medical things platform (IoMT) was utilized to implement the projected algorithm for decision support systems (DSSs) so that the doctors were assisted in effectively diagnosing heart disorder in cloud platforms. This algorithm was efficient for determining the risk level of coronary infection. The outcomes indicated that the projected algorithm offered a precision around 99.1%.

Ozcan and Peker [25] established a supervised machine learning technique called classification and regression tree (CART) model with the objective of predicting the coronary disorder, and extracting the decision rules (DRs) so that the associations amid input and output variables were specified [25]. Moreover, the results helped in ranking the attributes related to coronary disease. The accuracy of the established technique was measured 87%. The extracted DRs were capable of simplifying the use of clinical diagnosis without any requirement of additional information. The established technique was robust for supporting the doctors to diagnose and cure the heart diseases. Moreover, this technique was cost-effective and consumed least time.

Rahim *et al.* [26] developed a machine learning based CVD diagnosis (MaLCaDD) model to predict CVD at higher precision. The fundamental gaol of this model was of tackling the missing values with the help of mean replacement (MR) method, and SMOTE was implemented for tackling the data imbalance. After that, the feature importance method was adopted to select the attributes. In the end, an ensemble of LR and K-NN algorithm was suggested for predicting the heart disease at superior accuracy. Framingham, heart disease and cleveland datasets executed to compute the designed algorithm. Based on experiments, the designed algorithm offered an accuracy of 99.1% on first dataset, 98.0% on second and 95.5% on last dataset.

Mohapatra *et al.* [27] designed a predictive method for predicting the heart disorder depending upon the stacking of numerous classification algorithms in 2 levels: base level and meta level [27]. The promising result was generated by integrating diverse heterogeneous learners. This method focused on normalizing the data for ensuring the distribution of data as even and available on same scale. The designed method attained an accuracy around 0.92, precision of 0.926, sensitivity around 0.926, and specificity around 0.91. The designed method provided the benefits to integrate the weak learners and deploy their heterogeneity for enhancing the predictive results.

Shrivastava *et al.* [28] investigated a hybrid framework on the basis of deep learning (DL) approaches for predicting an individual as infected or normal via CNN-Bi-LSTM. Cleveland dataset was considered to tackle the issue related to redundant data and imbalanced data. For this, the techniques used to process the data were employed. An extra tree classification algorithm was developed to select the attributes and classify the coronary disease. The investigated framework performed well with regard to accuracy, precision, recall and F1-score. The findings exhibited that the investigated framework offered 96.66% accuracy as compared to the traditional methods.

### 2.1. Research gap

Heart disease is a significant global health issue, leading to high morbidity, mortality, and healthcare costs. Despite extensive research on heart disease, there are still several research gaps that need to be addressed. This section will discuss various research gaps that have been concluded viewing the existing literature mentioned above in section 2. Follwing are the existing research gaps of heart disease prediction work:

- The research gap in these studies collectively lies in the absence of comprehensive comparisons with existing methods and a lack of investigation into the model's robustness across diverse patient populations, hindering a complete understanding of their real-world effectiveness and potential limitations [21], [24].
- The combined research gap in these studies encompasses the need for further exploration into the generalization and applicability of the proposed predictive models across diverse datasets and patient populations. Additionally, a comparative analysis with other advanced predictive frameworks is lacking, limiting a comprehensive understanding of their relative performance and potential advantages in real-world clinical scenarios [22], [26].
- There is the absence of a comprehensive assessment of the proposed methods' performance across a wider range of diverse datasets and their adaptability to varying clinical contexts. Furthermore, the studies lack an exploration of potential limitations or challenges that could arise when applying these models to real-world scenarios, hindering a complete understanding of their practical effectiveness and broader utility in clinical decision-making [23], [27].
- The research gaps include the necessity for broader validation of the proposed models on diverse datasets and populations, along with comparative assessments against advanced techniques. Addressing potential limitations and biases is crucial for ensuring robustness and reliability in real-world healthcare contexts [25], [28].

## 3. RESEARCH METHOD

As DL has evolved over time, it has become increasingly capable of handling more complex tasks and improving accuracy in image classification [29]. The proposed study designed a framework based on DL to predict the heart disorder. The designed model framework will predict heart disease with high accuracy as compared to existing DL models. The phases of the proposed model are explained below:

### 3.1. Dataset input and pre-processing

UCI repository is considered to gather the data in a dataset and the pre-processing phase is executed to eliminate the missing, unsuitable values from the dataset. There are 76 characteristics in this database. The database relates to the patient's existence of cardiac disease. It has an integer value ranging from 0 (no presence) to 4. Experiments with the Cleveland database have mostly focused on trying to discern between presence (values 1, 2, 3, and 4) and absence (value 0).

### 3.2. Feature reduction

This stage makes the implementation of principal component analysis (PCA) model for mitigating the attributes. This statistical model tries to provide the uncorrelated variables by transforming the set of interconnected elements into a collection of linearly discrete subsets according to a transformation. Orthogonal linear transformation is a synonym for this algorithm, which is used to project the main dataset to a different projecting system. Moreover, the end objective is also considered which employs the largest variance containing a projection of the 1st element. Furthermore, a projection of the 2nd coordinate is comprised in the 2nd largest variance. It implies that it is placed vertically against the 1st element. Generally, this algorithm helps in performing LT represented with $z = W_k^T x$ and $x \in R^d$, and $r < d$, to increase the variance of the data in the projected space. $X = \{x_1, x_2, \dots\dots.., x_i\}$, $x_i \in R^d$, $z \in R^r$and $r < d$ is useful for illustrating the data matrix, and a set of p-dimensional vectors of weights $W = \{w_1, w_2, \dots\dots.., w_p\}$, $w_p \in R^k$ is considered to define the entire process for matching every $x_i$ vector of X to a;

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

the variance is increased after observing the condition primary with weight W1 as:

$$W_i = \arg max_{|w|} = \{\textstyle\sum_i (xi.W)^2\} \tag{2}$$

this condition is enlarged as (3).

$$W_i = argmax_{||w||=1} \left\{ \left||x.w|\right|^2 \right\} = argmax_{||w||=1} \{W^T\, X^T\, XW\} \tag{3}$$

An analysis is performed on the symmetric grid defined with the $X^T X$ successfully, subsequent to attain the largest eigenvalue of the matrix and W denotes the related eigenvector. The $W_1$ is provided prior to project the primary data matrix X onto the $W_1$ within the space of transformation to infer the initial principal element. The newly obtained elements are mitigated to offer further segments along these lines.

### 3.3. Classification

This stage implements a hybrid DL framework which is combination of LSTM and CNN. A CNN is a vast network that replicates and comprehends stimuli in a manner similar to how the brain's visual cortex functions. It is also a deep neural network with numerous hidden layers. The neural network is often used for multiclass categorization in CNN’s output layer. Instead of doing it by hand, CNN utilises a feature extractor in the training phase. The feature extractor used by CNN is made up of unique neural network types, the weights of which are determined during training. When CNN’s neural network feature extraction is more in-depth (has more layers), it improves picture identification, but at the expense of the learning process complexity that had previously rendered CNN ineffective and unappreciated. CNN helps in extracting the attributes of the input images, though a different neural network categorises features. The input image is employed as an initial point in this algorithm. The retrieved feature signals are deployed to classify the disease. For categorization, this process used a Bi-directional model. The forward LSTM and Backward LSTM of Bi-LSTM are coupled by Softmax regression layer. Due to the normal LSTM network’s lack of future context information and inability to learn all the sequences, the prediction effect is lost when dealing with time series. The process of predicting heart disease of the proposed model is demonstrated Figure 2. First of all, the images of heart diseases are utilized as input in the proposed model and are pre-processed for removing the missing values. After that, the PCA algorithm is implemented for mitigating the features. The data is split into two sets: training and testing. In initial set, the CNN model is implemented to extract the features and Bi-directional model is deployed later on. Thereafter, the training system is created. In latter set, the training model is developed directly. Afterward, the testing data is predicted. At last, diverse metrics, such as accuracy, precision and recall are employed for computing the performance. Hence, the results are obtained.

The presented algorithm involves the creation and training of a neural network model with specific architecture and parameters. The dataset is split into a training set (80%) and a test set (20%). The neural network has an input layer with 13 nodes and a rectified linear unit (ReLU) activation function, a hidden layer with 4 nodes and ReLU activation, and an output layer with 1 node and a sigmoid activation function. The learning rate is set to 0.001, utilizing binary crossentropy as the loss function and the Adam optimizer. Training is performed over 100 epochs with a batch size of 10. The results obtained from this binary model, such as accuracy and loss metrics, would provide insights into the model's performance on the given dataset.

Algorithm 1. Heart disease prediction

```
Begin
        Splitting the given dataset into training and test

        Training set size and test size is configuring as 0.8 and 02
        Model Creation
        In the model input layer consists of 13 nodes and activation function
        as’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'
        Training is done for 100 epochs and batch size=10
        The result for binary model is obtained is below
    end
```

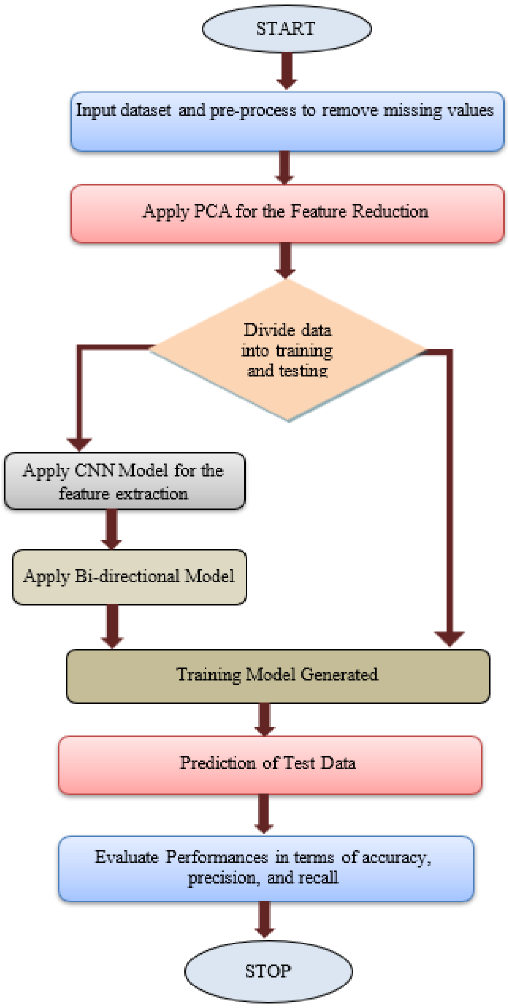


Figure 2. Proposed model

## 4. RESULTS AND DISCUSSION

A dataset named Cleveland, consisting of 14 features, has been employed for this study to forecast CVD. Various algorithms are employed and compared to predict cardiac disorders. Additionally, a deep learning model is suggested for CAD prediction and compared to existing deep neural networks such as CNN and LSTM. Figure 3 depicting the result of proposed work in which Figure 3(a) indicates the evaluation of the training accuracy of the hybrid model. The testing and validation results are analysed. It is indicated that the suggested model offered the training accuracy up to 98%. The training accuracy is shown with the validation accuracy of the developed system, Figure 3(b) illustrates the computation of the loss trend of the hybrid deep learning model. The testing and validating results are analysed. It is depicted that the suggested model has achieved an approximate 3% loss in predicting heart disease in testing and validation cases, Figure 3(c) represents precision and recall curve of the proposed model for heart disease prediction. The x axis represents recall value and y axis represents precision value. The recall value is found 0.97 and the precision value is found 0.97 while predicting heart disease and Figure 3(d), the effectiveness of hybrid deep learning model evaluated against CNN and LSTM. The hybrid deep learning model achieves accuracy of 97% which 8% high as compared to CNN model. The precision and recall values of the hybrid deep learning model is also high as compared to CNN, LSTM models.

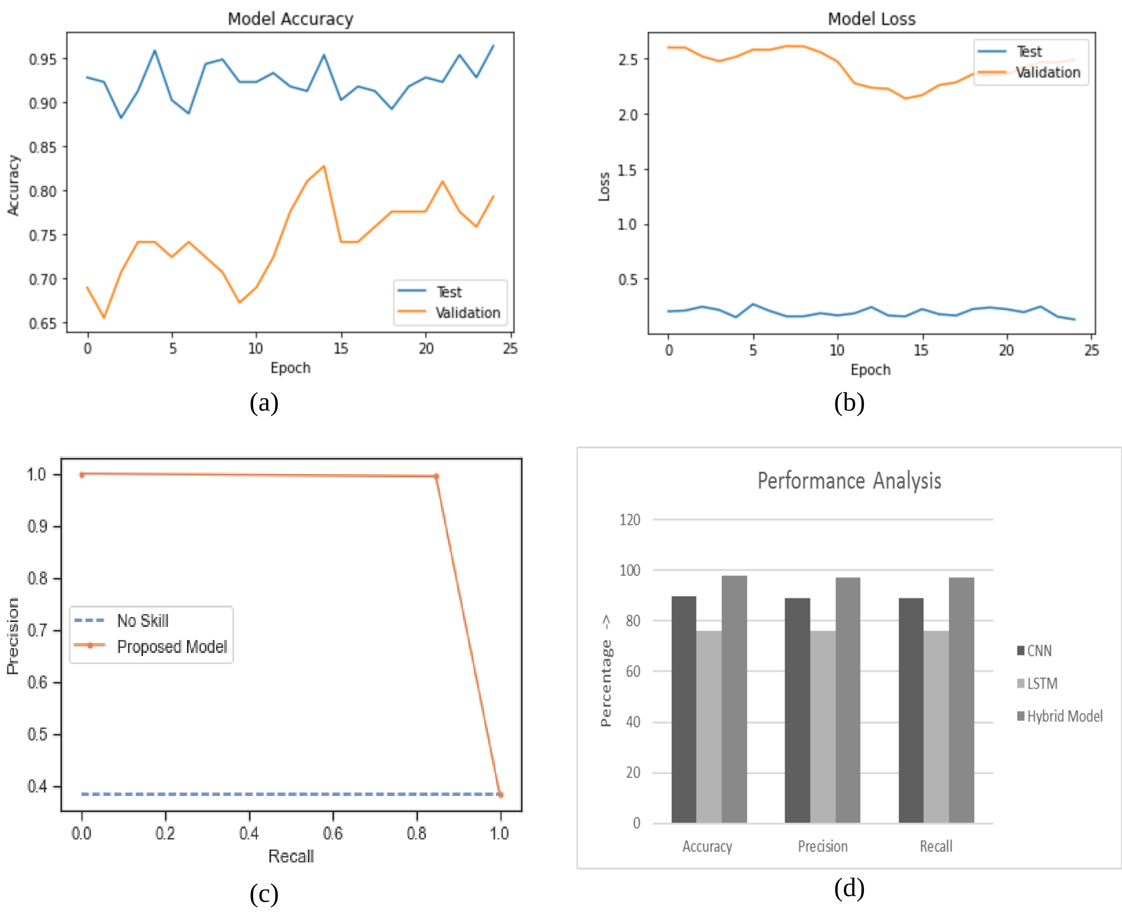


Figure 3. Depicting the results of our proposed research work in which; (a) shows the training model accuracy, (b) shows the result of model loss, (c) provides the ROC curve, and (d) analysing the performance on the basis of accuracy, precision, recall

Table 1 depicts the classification-based analysis of suggested model. The data is classified into 2 classes: 1 and 0. Precision, recall, F1-score and support, metrics are employed. The suggested model yielded precision of 0.97 and 0.96, recall of 0.96 and 0.97, F1-score of 0.95 and 0.96, and support of 35 and 26 for both classes respectively.

Table 2 depicts the comparison of the CNN, LSTM and suggested hybrid model for analysing the performance. For this, different metrics, such as accuracy, precision and recall are considered. It is analysed that the suggested model yielded an accuracy of 97.78%, recall of 97% and precision of 97%, which are found higher in contrast to other models.

Figure 4 displays the hybrid model's ROC curve while predicting the heart disease. This curve is drawn between false positive rate and true positive rate. The value of true positive rate is raised to 0.83 for the heart disease prediction.

Table 1. Classification based analysis of proposed model

| Class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| 1 | 0.97 | 0.96 | 0.95 | 35 |
| 0 | 0.96 | 0.97 | 0.96 | 26 |

Table 2. Performance analysis

| Model | Accuracy | Precision | Recall |
|---|---|---|---|
| CNN | 89.91% | 89% | 89% |
| LSTM | 76.16% | 76% | 76% |
| Hybrid model | 97.78% | 97% | 97% |

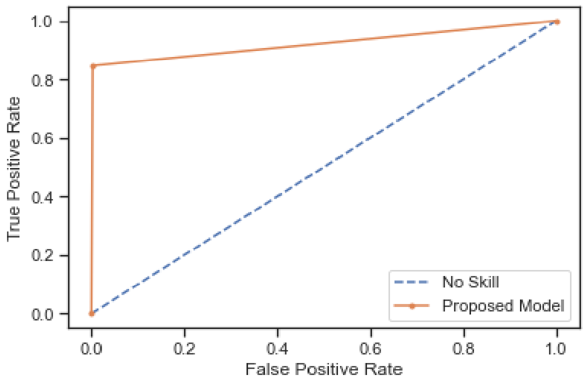


Figure 4. ROC curve

## 5. CONCLUSION

Heart diseases encompass various disorders that affect the heart, including the coronary arteries and the circulatory system. The CDC has reported that heart disease is a primary cause of mortality in developed nations. Coronary disease, in particular, has been responsible for a significant number of deaths in the US. It can lead to various diseases that affect various segments of the heart and its functioning. This research represents that the complex issue is related to predicting the coronary disorder because of the occurrence of great volume of attributes. The CNN and LSTM models is implemented previously for the prognosis of heart diseases. The hybrid deep learning model is suggested which is the combination of CNN and LSTM. This model achieves accuracy of 97% which is 8% higher than other deep learning models. In future transfer learning models can be applied with hybrid model for the heart disease prediction.

## REFERENCES

[1] Y. Huang *et al.*, "Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment," *Computers in Biology and Medicine*, vol. 151, p. 106297, Dec. 2022, doi: 10.1016/j.compbiomed.2022.106297.

[2] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: an effective heart disease prediction model for a clinical decision support system," *IEEE Access*, vol. 8, pp. 133034–133050, 2020, doi: 10.1109/ACCESS.2020.3010511.

[3] S. A. Ali *et al.*, "An optimally configured and improved deep belief network (OCI-DBN) approach for heart disease prediction based on ruzzo-tompa and stacked genetic algorithm," *IEEE Access*, vol. 8, pp. 65947–65958, 2020, doi: 10.1109/ACCESS.2020.2985646.

[4] D. Cenitta, R. V. Arjunan, and K. V. Prema, "Ischemic heart disease prediction using optimized squirrel search feature selection algorithm," *IEEE Access*, vol. 10, pp. 122995–123006, 2022, doi: 10.1109/ACCESS.2022.3223429.

[5] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthcare Analytics*, vol. 2, p. 100060, Nov. 2022, doi: 10.1016/j.health.2022.100060.

[6] T. R. Dawber, W. B. Kannel, and P. M. Mcnamara, "The prediction of coronary heart disease.," *Transactions of the Association of Life Insurance Medical Directors of America*, vol. 47, pp. 70–105, 1964.

[7] D. Bertsimas, L. Mingardi, and B. Stellato, "Machine learning for real-time heart disease prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3627–3637, Sep. 2021, doi: 10.1109/JBHI.2021.3066347.

[8] W. Li, M. Zuo, H. Zhao, Q. Xu, and D. Chen, "Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks," *Methods*, vol. 198, pp. 96–106, Feb. 2022, doi: 10.1016/j.ymeth.2021.12.009.

[9] R. P. Cherian, N. Thomas, and S. Venkitachalam, "Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm," *Journal of Biomedical Informatics*, vol. 110, p. 103543, Oct. 2020, doi: 10.1016/j.jbi.2020.103543.

[10] B. Wang *et al.*, "A multi-task neural network architecture for renal dysfunction prediction in heart failure patients with electronic health records," *IEEE Access*, vol. 7, pp. 178392–178400, 2019, doi: 10.1109/ACCESS.2019.2956859.

[11] W. Chang, Y. Liu, X. Wu, Y. Xiao, S. Zhou, and W. Cao, "A new hybrid XGBSVM model: application for hypertensive heart disease," *IEEE Access*, vol. 7, pp. 175248–175258, 2019, doi: 10.1109/ACCESS.2019.2957367.

[12] S. J. Pasha and E. S. Mohamed, "Novel feature reduction (NFR) model with machine learning and data mining algorithms for effective disease risk prediction," *IEEE Access*, vol. 8, pp. 184087–184108, 2020, doi: 10.1109/ACCESS.2020.3028714.

[13] S. S. Sarmah, "An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network," *IEEE Access*, vol. 8, pp. 135784–135797, 2020, doi: 10.1109/ACCESS.2020.3007561.

[14] S. Rao *et al.*, “An explainable transformer-based deep learning model for the prediction of incident heart failure,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 3362–3372, Jul. 2022, doi: 10.1109/JBHI.2022.3148820.
[15] P. Ghosh *et al.*, “Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques,” *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
[16] A. M. Khedr, Z. Al Aghbari, A. Al Ali, and M. Eljamil, “An efficient association rule mining from distributed medical databases for predicting heart diseases,” *IEEE Access*, vol. 9, pp. 15320–15333, 2021, doi: 10.1109/ACCESS.2021.3052799.
[17] M. A. Khan, “An IoT framework for heart disease prediction based on MDCNN classifier,” *IEEE Access*, vol. 8, pp. 34717–34727, 2020, doi: 10.1109/ACCESS.2020.2974687.
[18] T. Amarbayasgalan, V. H. Pham, N. Theera-Umpon, Y. Piao, and K. H. Ryu, “An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets,” *IEEE Access*, vol. 9, pp. 135210–135223, 2021, doi: 10.1109/ACCESS.2021.3116974.
[19] J. Chen, A. Valehi, and A. Razi, “Smart heart monitoring: early prediction of heart problems through predictive analysis of ECG signals,” *IEEE Access*, vol. 7, pp. 120831–120839, 2019, doi: 10.1109/ACCESS.2019.2937875.
[20] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, “Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison,” *Computers in Biology and Medicine*, vol. 136, p. 104672, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104672.
[21] I. D. Mienye, Y. Sun, and Z. Wang, “Improved sparse autoencoder based artificial neural network approach for prediction of heart disease,” *Informatics in Medicine Unlocked*, vol. 18, p. 100307, 2020, doi: 10.1016/j.imu.2020.100307.
[22] H. Yang, Z. Chen, H. Yang, and M. Tian, “Predicting coronary heart disease using an improved lightGBM model: performance analysis and comparison,” *IEEE Access*, vol. 11, pp. 23366–23380, 2023, doi: 10.1109/ACCESS.2023.3253885.
[23] S. E. A. Ashri, M. M. El-Gayar, and E. M. El-Daydamony, “HDPF: heart disease prediction framework based on hybrid classifiers and genetic algorithm,” *IEEE Access*, vol. 9, pp. 146797–146809, 2021, doi: 10.1109/ACCESS.2021.3122789.
[24] Y. Pan, M. Fu, B. Cheng, X. Tao, and J. Guo, “Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform,” *IEEE Access*, vol. 8, pp. 189503–189512, 2020, doi: 10.1109/ACCESS.2020.3026214.
[25] M. Ozcan and S. Peker, “A classification and regression tree algorithm for heart disease modeling and prediction,” *Healthcare Analytics*, vol. 3, p. 100130, Nov. 2023, doi: 10.1016/j.health.2022.100130.
[26] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim, and A. W. Muzaffar, “An integrated machine learning framework for effective prediction of cardiovascular diseases,” *IEEE Access*, vol. 9, pp. 106575–106588, 2021, doi: 10.1109/ACCESS.2021.3098688.
[27] S. Mohapatra, S. Maneesha, P. K. Patra, and S. Mohanty, “Heart diseases prediction based on stacking classifiers model,” *Procedia Computer Science*, vol. 218, pp. 1621–1630, 2022, doi: 10.1016/j.procs.2023.01.140.
[28] P. K. Shrivastava, M. Sharma, P. sharma, and A. Kumar, “HCBiLSTM: a hybrid model for predicting heart disease using CNN and BiLSTM algorithms,” *Measurement: Sensors*, vol. 25, p. 100657, Feb. 2023, doi: 10.1016/j.measen.2022.100657.
[29] S. Gupta and P. Sharma, “Machine learning approach for heart disease prediction: a survey,” *AIP Conference Proceedings*, vol. 2555, 2022, doi: 10.1063/5.0108884.

## BIOGRAPHIES OF AUTHORS

**Mr. Shubham Gupta** received a Master’s degree in Computer Engineering from University of Jammu. He is currently pursuing Ph.D. Computer Science Department, Rayat Bahra University, Mohali. His research area includes image processing, machine learning and deep learning. He can be contacted at email: guptashubham96222@gmail.com.

**Dr. Pooja Sharma** is an Associate Professor in Department of Computer Science, Rayat Bahra University since 2007. Her area of specialization is computer networks for high performance computing. As an expert in computer networks, she is guiding the next generation of engineers and network administrators as well as contributes to moving the research front in computer networks. She has many good publications and patents titled AI-powered virtual interior design system and method using generative models. She can be contacted at email: pooja.rani@rayatbahrauniversity.edu.in.