VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2024, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 12, Number 3, July-September 2024 pp:68-81

IoT-Driven Approaches for Early Detection and Monitoring of Heart Disease: Current Trends and Future Directions

Areebah [1], **Muhammad Rabbi Butt** [2*], **Faheem Aslam** [3]

[1,2,3]Department of control and computer engineering, NCEPU, Beijing China

Keywords: *IoT based HD classification, Hybrid Models, ECG signals dataset.*

Journal Info:
Submitted:
August 08, 2024
Accepted:
August 21, 2024
Published:
August 26, 2024

Abstract

Cardiovascular disease (CVD) is a serious disease having a widespread effect on individuals across the world. Early and accurate detection of cardiac disease is crucial in healthcare, especially in the domain of cardiology. Currently, a non-invasive ultrasound imaging method is used that evaluates the structure, performance, and blood, allowing for the precise identification of a number of cardiac ailments, such as valve problems, heart failure, and congenital anomalies. These traditional techniques have some limitations, including high cost, the need for medical expertise and equipment, and the fact that they often create incorrect results due to human involvement. Furthermore, the traditional method takes more time to predict heart disease. Electrocardiogram (ECG) signals play a critical role in reducing death rates caused by CVDs, and they provide details regarding the heart patient's health to a medical expert by employing an automated heart failure detection system. Recent developments in deep learning-based health care systems, such as ECG signal analysis, include CNN, LSTM, and other neural networks. In this research, we provide a hybrid deep learning based approach for the timely and accurate diagnosis of cardiovascular disease. The proposed system uses a hybrid of convolutional neural networks (CNN) and long short-term memory (LSTM) and utilizes the MIT-BIH ECG signal dataset for heart disease diagnosis. This study uses two different approaches with MIT-BIH arrhythmia imbalanced and balanced datasets. The first approach uses CNN and CNN-LSTM with an imbalanced dataset, and the second approach uses CNN and CNN-LSTM with a balanced dataset. The performance of both approaches was analyzed. The experimental outcomes show that the overall performance of both CNN, CNN-LSTM was excellent on a balanced dataset compared to imbalanced dataset. The proposed system achieved a better result than the previous suggested methods. Additionally, it is easy to adopt the suggested technique in the field of healthcare in order to identify heart disease.

***Correspondence author email address:** Rabbibutt3@gmail.com
DOI: 10.21015/vtse.v12i3.1860

1 Introduction

Heart disease is one of the most dangerous and persistently fatal diseases globally. In this disease, the heart usually fails to deliver sufficient oxygenated blood to the remaining parts of the human body in order to operate properly [1]. It can be a result of a number of factors and usually grows continually with time. Heart disease occurs when the coronary arteries block suddenly and damage some heart muscles [2]. These arteries serve as channels for supplying blood to the circulatory system. A recent report published shows that Russia has the greatest number of heart patients [3]. The most typical symptoms of heart disease include pain in the chest, shortness of breath, and tiredness. Unhealthy foods with high cholesterol, smoking cigarettes, high levels of blood pressure, poor nutrition, and a lack of body exercise are all factors raising the risk of heart disease [4]. Heart disease is classified into several categories; the most common of them is coronary artery disease (CAD), which may cause heart attack, chest pain, and stroke. Other categories of heart illness include heartbeat problems, cardiovascular disease, and congestive heart failure [5].

Traditional investigation approaches were applied for the diagnosis of heart illness, but they were too complicated to use [6]. Due to insufficient medical knowledge, and a lack of instruments, the diagnosis of heart disease is very difficult in poor countries [7]. However, early and accurate diagnosis of heart disease is very essential to preventing further heart loss [8]. Heart disease has become a fatal disorder that's increasing rapidly in nations across the globe. According to a recent report by the World Health Organization (WHO), approximately 17.9 million individuals lost their lives due to CVD in 2016. This rate reflects 30% of the worldwide deaths. Based on a survey, approximately 0.2 million individuals die in Pakistan each year because of heart disease. The European Society of Cardiology (ESC) released a report that revealed 26.5 million individuals were identified who suffered from heart disease, with an additional 3.8 million reported annually. Approximately 50–55% of heart disease patients die in an initial span of 1-3 years [9]. Invasive diagnostic techniques rely on an individual's medical records, medical test findings, and an analysis of indications by medical experts [10]. Among traditional methods, angiography has become a popular method for the detection of heart disease. However, angiography is still facing some limitations, including high cost, several risks, and lack of medical expertise [11]. Traditional methods usually result in incomplete treatment and require more time due to human involvement [12].

To overcome these challenges, various machine learning and deep learning algorithms have been utilized to develop an expert medical system for early and accurate heart disease detection [13]. Because of this, the annual mortality rate for people with heart disease has decreased [14]. Traditional machine learning and deep learning methods suggested for diagnosis of heart disease has several limitation including, the models might produce biased predictions if there are data imbalances, such as a class of cardiovascular diseases that is noticeably more common than other. Furthermore, these techniques may not be robust for use in real-world medical scenarios due to difficulties with generalization when dealing with a variety of patient demographics as well as data quality differences.

Hybrid models integrates the features of traditional ML and DL models, and having several advantages over classical ML approaches. Additionally, hybrid models can handle large and complex data more efficiently and accurately. Furthermore, these models are adaptive and versatile, which may improve their capacity of robustness and generalization across varies patient groups as well as data sources, which makes them useful tools in healthcare settings. Proposed study used hybrid model approach of CNN and LSTM with a MIT-BIH ECG signal dataset with target classes normal and abnormal. The present study's major contributions to the classification of heart disease are:

· To use a hybrid model approach with a recently updated ECG signal dataset to classify heart disease patients and normal people accurately at early stage.

· To apply hybrid models with both balanced and imbalanced datasets and then compare their performance.

· To analyze and compare the results of hybrid models with recent studies, models with the highest accuracy.

The rest of the article is structured as follows: A brief review of the published work is provided in Section 2, next to descriptions of the materials and methods in Section 3. Results and discussion are presented in Section 4, followed by a conclusion and future work in Section 5.

2 Literture Review

To tackle these challenges with traditional methods, researchers and professionals have shown a strong interest in machine learning and deep learning approaches to diagnosing heart disease. In this context, Robert et al .[15] utilized a logistic regression model for heart disease detection and achieved 77.1% accuracy. Similar to this, Ankur gupta et al. [16] utilized machine learning classifiers with hyper parameters tuning to diagnose heart disease, and obtained 93% accuracy with Cleveland heart disease dataset. Likewise, Yar Muhammad et al. [17] used various ML models with both full features and selected features of the Cleveland and Hungarian datasets, and obtained accuracy of 94.4%. In a sequence, Senthilkumar et al. [18] also employed a hybrid random forest with a linear model (HRFLM) to uncover relevant features and applied various feature classification techniques to the UCI dataset, obtaining 88.7% accuracy. In another research study, Long et al. [19] suggested a diagnosis method for heart illness based on feature optimization for rough sets utilizing the chaos firefly technique and intervals type-2 fuzzy logic, and validated it utilizing a cardiovascular disease dataset. Similarly, Das and Turkoglu

[20] suggested a neural network (ANN) ensemble-based prediction technique for heart disease diagnosis. In another research study, Alizadehsani et al. [21] have created a non-invasive framework for the early detection of heart disease and achieved promising results in terms of all performance evaluation metrics. Similarly, VijayashreeandSultana et al. [22] utilized SVM with particles warm optimization (PSO), and achieved accuracy of 83%. Likewise, Ismaeel et al. [23] applied extreme learning-based techniques to identify cardiac disease and validated their technique using the Cleveland heart disease dataset, which showed an 80% accuracy. Purushottam et al. [24] suggested a classification algorithm based on rules for predicting heart disease and obtained an accuracy score of 86.7%.

Similarly, Chintan Bhatt et al. [25] suggested a Huang-starting approach for k-mode clustering to enhance the accuracy of heart disease classification, utilized RF, DT, MP, and XGBoost models. Their proposed experiment achieved highest accuracy of 87.28%. In another research study, Shah et al.[26] used various ML models, utilized the Cleveland dataset, and obtained a maximum accuracy of 90.8%. In a sequence, Alotalibi et al.[27] used various machine learning models including DT, LR, RF, NB, and SVM with 10-fold cross-validation technique, and achieved maximum accuracy of 93.19%. Similarly, Hasan et al. [28] suggested an approach for the identification of heart disease that utilized various ML models, RF, SVM, KNN, NB, and XGBoost.and achieved the highest accuracy of 73.74%. Likewise, Seyed Matin Malakouti [29] provide a classification strategy for detecting heart disease using various ML models including Gaussian NB, RF, LR with ECG signal dataset. Their suggested approach was 96% accurate. Jafar Abdollahi et al. [30] suggested an ensemble based approach with genetic algorithms for heart disease detection and used the UCI heart disease dataset. Their proposed ensemble based approach achieved 97.57% accuracy. Jian Ping Li et al. [31] used various ML and DL classification models with some feature selection techniques to detect heart disease and utilized the Cleveland heart disease dataset, achieved 92.37% accuracy with full features. Detrano et al. [32] proposed a ML based heart disease classification system that utilized the Cleveland heart disease dataset with some feature selection techniques and achieved an accuracy of 77%. Similarly, P. Dileep et al. [33] developed a heart disease classification system based on cluster-based bi-directional LSTM classification models and achieved accuracy of 94.78%. In another research study, Krishnan et al. [34] suggested an RNN-GRU-LSTM-Adam

Optimizer hybrid deep learning model for cardiovascular disease diagnosis, and achieved the greatest accuracy of 98.6876%. Likewise, Liu et al. [35] developed a system for classifying HD utilizing relief along with rough set methods. The suggested classification approach obtained 92.32 % accuracy. Mohan et al. [36] developed a system for heart disease prediction utilizing hybrid machine learning models. In addition, he presented an innovative approach for selecting important features for the training and validation of ML models. Their proposed hybrid approach achieved 88.07% accuracy. Nafisa et al. [37] used a hybrid of CNN, LSTM, and BI-LSTM with an ECG signal dataset and achieved 99.2% accuracy. Some previous studies with high accuracy and their limitations are summarized in Table 1.

Table 1. Summary of various studies and their contributions.

Reference	Contribution	Model Applied	Dataset Used	Accuracy	Weakness
Yar Muhammad et al. [17]	Performance with selected and full features.	KNN, GB, LR, NB, SVM, AB, ET, ANN, DT, RF	Cleveland and Hungarian	94.4%	Small datasets Accuracy can be improved.
S Mohan et al. [18]	Performance with selected and full features.	HRFLM	UCI	88.7%	Small datasets Accuracy can be improved.
Das and Turkoglu [20]	Ensemble model using neural networks.	ANN	Cleveland	91%	Small dataset No proper feature selection Accuracy can be improved.
Alizadehsani et al. [21]	New model for features engineering.	SVM, NB, C4.5	Z-Alizadeh Sani	96.4%	More computational time Small dataset.
Repaka et al. [27]	Mobile application for new dataset collection.	NB	Own dataset	86%	Tested with small data Less accuracy.
Jian Ping Li et al. [31]	Rapid conditional mutual data feature selection.	FCMIM, Relief, MRMR, LASSO	Cleveland	92.37%	Novel approach, but accuracy not satisfactory.
Liu et al. [35]	Combined ReliefF and RS algorithm.	ReliefF, RFRS	UCI	92.59%	Small dataset Accuracy can be improved.
Nafisa et al. [37]	Hybrid model and dataset approach.	CNN, LSTM, BI-LSTM	MIT-BIH	99.2%	Imbalanced dataset.

All of the above-mentioned approaches are used for the early diagnosis of heart disease. However, all of these methods are not efficient in terms of accuracy and computation time. To overcome these challenges in previous studies, a lot of work is needed in this area. Enhancing prediction accuracy and reducing computation time is a challenge in automated heart disease diagnosis. The proposed research study uses a novel hybrid approach with both balanced and unbalanced ECG data to enhance the performance of existing methods. Furthermore, compared with previous approaches, the proposed study achieved better performance for all performance evaluation matrices.

3 Proposed system architecture

The main objective of the suggested approach is to detect heart disease accurately in its early stages. This study uses hybrid model of CNN and LSTM with MIT-BIH ECG signal datasets. Data preprocessing, normalization, outliers, feature selection, and data balancing are some crucial strategies in ML and DL related research, since these strategies ensure the data passed to the classifier is in error free and standardized form. All materials and methods used in the proposed research study are depicted in Figure 1.

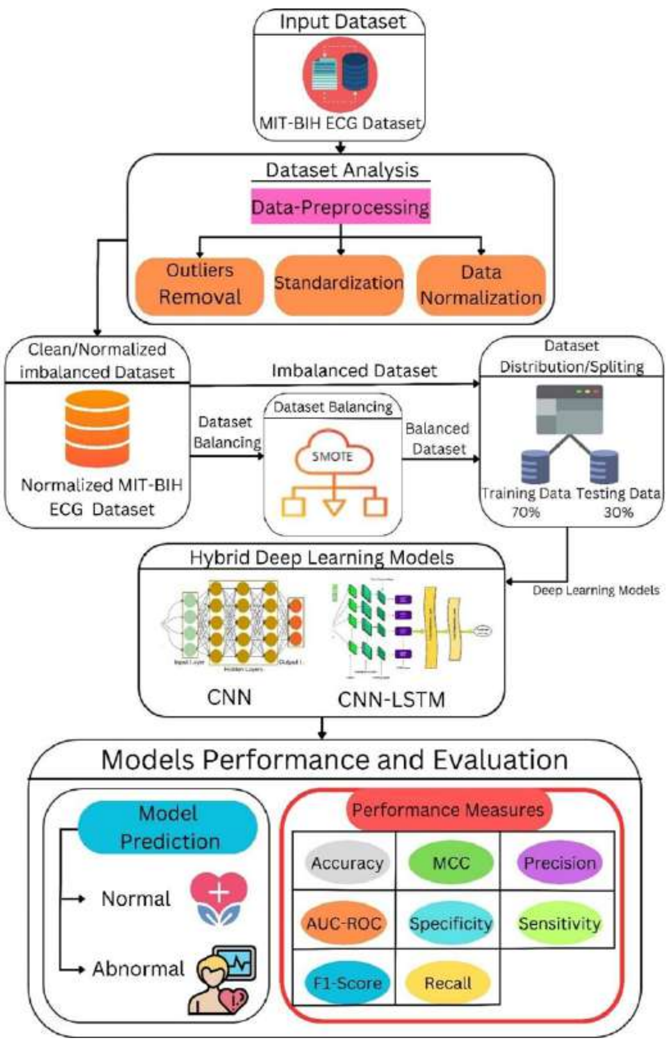

Figure 1. Proposed methodology

3.1 Dataset description

The collection of appropriate data for the specific scenario is the initial and critical stage in developing a smart computational system. The presence of an organized set of data for a particular scenario significantly affects the model's efficiency. The proposed study uses the MIH-BIT ECG signal dataset created by the Massachusetts Institute

of Technology (MIT) and Beth Israel Hospital, which comprises 48 half-hour ECG readings taken from 47 people, representing both typical sinus pulses and other arrhythmia forms. Cardiologists added annotations to all the recordings at a sample rate of 360 Hz and provided specific labels for all types of arrhythmia. Table 1 depicts the detailed description of the MIH-BIT dataset.

Table 2. Details of the MIT-BIH Arrhythmia Dataset

Attribute	Description
Dataset Name	MIT-BIH Arrhythmia Dataset.
Data Source	PhysioNet - The MIT-BIH Arrhythmia Database.
Purpose	Used for studying cardiac arrhythmias.
Sampling Frequency	Typically 360 Hz or 128 Hz.
Duration per Record	Varies, usually around 10 minutes per record.
Number of Channels	Typically two channels, representing two ECG leads.
License	Freely available for non-commercial use.

3.2 Preprocessing

Preparation of the dataset required for effective representation. It is a crucial step before feeding data to any classifier [38]. Normalization and data balancing are two important aspects of data preprocessing [39, 40]. Normalization ensures all features are contributing consistently to learning and also prevents scale-based features from prevailing, while data balancing avoids bias in Ml models. The MIT-BIH dataset does not contain any missing values; however, the dataset is unbalanced. The dataset contains 5 different classes: class 0 (normal), class 1 (fusion), class 2 (premature ventricular contraction), class 3 (atrial premature), and class 4 (fusion of ventricular and normal). Initially, the outliers in the dataset were handled and reduced into two classes: 0 (normal) and 1 (abnormal). Four different approaches are used: the first two use unbalanced data, and the remaining two use balanced data.

3.3 Dataset balancing using SMOTE

Synthetic Minority Oversampling Technique (SMOTE) is a statistical method used for increasing the number of instances in dataset in a balanced way. This technique generates additional training data by modifying real data. SMOTE oversampling technique has several advantages over other techniques because SMOTE not creating duplicate data, it generates the synthetic data that are closely similar to real data. Using an imbalanced dataset in machine learning increases the bias of the model toward majority classes. Additionally, the imbalanced data may lead to changes in evaluation measures, which makes it challenging to determine the model's actual performance. In converse using balanced dataset make the model training easy and help to prevent the model from bias. Balancing strategies include either oversampling the minority class by duplicating or generating new data points or reducing the majority class by eliminating some of its records. In our case, the MIB-BIH ECG dataset was unbalanced, with 36236 records belonging to 0 and 7548 records belonging to 1. The Synthetic Minority Oversampling Technique (SMOTE) dataset balancing technique is used to balance the dataset.

3.4 Utilized Deep Learning models for the diagnosis of heart disease

3.4.1 Convolutional neural networks architecture (CNN)

This section provides a detailed explanation of the suggested structure, including all of its levels and the methods utilized to enhance the architecture. CNN has gained popularity in finding patterns challenges such as classifying images and recognition of objects [41]. Similar to ANNs, CNNs are made that consists of self-optimizing neurons that have been trained to carry out a certain task. The suggested CNN model architecture consists of a feed-forward network with a sequential structure in which the various layers are linked together in a single-input and single-output

Input: X_{nor}, X_{per}, K
Output: X_{mote}

$X_{mote} \leftarrow \{\}$
for $i = 1$ to $Len(X_{nor})$ **do**
  $mm \leftarrow KNN(mm)$
  $mm \leftarrow KNN(X_i, X_{nor}, K)$
  $q \leftarrow \frac{X_{per}}{100}$
  while $q \neq 0$ **do**
    $X_{neighb} \leftarrow$ select random(mm)
    $X_{mote} \leftarrow X(mm)$
    $X_{mote} \leftarrow X_i + \text{rand}(0,1) \times |X_{neighb} - X_i|$
    $q \leftarrow q - 1$
  end while
end for
return X_{mote}

---

mode. The framework takes as input shape tensors with batch sizes of 187, 2, and 1, which represent ECG readings and 187 data points with two channels of data (potentially lead data) and a single channel (depth of 1). The architecture uses a number of Conv2D layers, with batch normalization, ReLU activation, and dropout following each layer at an average dropout rate of 0.5. This helps to prevent overfitting. The first Conv2D layer contains 256 filters with kernel sizes of (10, 2) and strides of (5, 1), respectively. Following the convolutional layers, a Global Average Pooling 2D layer is used to minimize the overall dimension of the feature features. Lastly, a dense layer with two units and a softmax activation function is used to get a result that shows the expected probability for each class (normal and abnormal). The model has been trained with the Adam optimizer and cross entropy loss function. Figure 2. shows the architecture of CNN model.

CNN is made up of several layers intended for pattern detection and image processing [42]. Convolutional layers are used to find basic characteristics like edges. The next pooling layer works on all feature maps separately to shrink the dimension, which in turn reduces the number of parameters and computations required by the network. At the end fully connected layers classify the objects based on learned

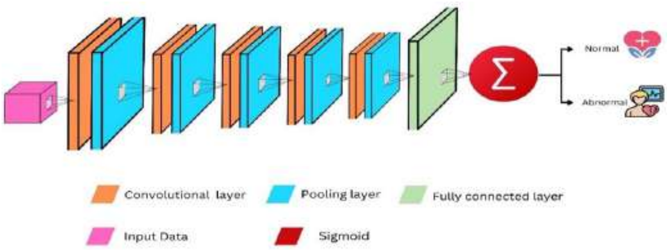

Figure 2. shows the architecture of CNN model

features. Various feature in the data and convolution are represented in equation (1), max pooling in equation (2) and ReLu is explained in equation (3).

$$(f * g)(n) = \int f(x)\,dx \quad \text{(continuous)}$$

or

$$(f * g)(p) = \sum f(q)g(p-q) \tag{1}$$

Where f is the input data, g is the filter, and $*$ denotes the convolution operation.

$$p_{max}(l,m) = \max\left(\text{input}(l,m), //\text{input}(l+1,m), \text{input}(l,m+1), \text{input}(l+1,m+1)\right) \tag{2}$$

Where p_{max} is the pooled value, and input(l,m) represents the value at position (l,m) in the input.

$$f(n) = \max(0,n) \tag{3}$$

3.5 CNN-LSTM architecture.

CNN-LSTM is hybrid model of CNN and LSTM models. The proposed CNN-LSTM model begins with a sequence of convolutional layers, where each layer uses filters of variable sizes to extract various levels of information from the source data. After each layer of convolutions, batch normalization and ReLU activation are used to make learning more stable and non-linear [43]. By introducing dropout layers that randomly deactivate specific neurons throughout learning, overfitting is reduced. The output of the convolutional layers is then modified to provide an LSTM layer with data in the correct dimensions. The LSTM layer analyses the features that the preceding layers collected and records the temporal relationships in the data [44]. The output possibilities of the model are then produced by applying a dense layer and an activation function based on Softmax. The Adam optimizer, sparse categorical cross-entropy loss, and accuracy are used to build the model. The built model is returned along with a printout of the model's structure. Convolutional and RNN elements are used in this model to handle sequential information as well as generate predictions using learned features and temporal patterns.

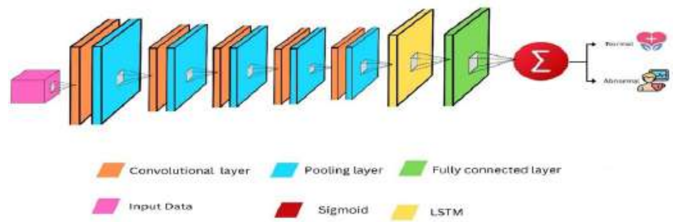

Figure 3. shows the CNN-LSTM architecture

In the proposed architecture the CNN layers extract the spatial features form input data and then pass these features to LSTM layers for classification purposes. LSTM units consist of a cell with an input gate, output gate, and forget gate as shown in equations (3-9).

$$f_b = \sigma(X_b * U_f + H_{b-1} * w_f) \tag{4}$$

$$\overline{cb} = \tanh(X_b * U_d + H_{b-1} * W_d) \tag{5}$$

$$l_b = \sigma(X_b * U_{\dot{y}} + H_{b-1} * w_j) \tag{6}$$

$$o_b = \sigma(X_b * U_m + H_{b-1} * w_m) \tag{7}$$

$$c_b = f_b * c_{b-1} + l_b * \overline{cb} \tag{8}$$

$$H_b = o_b * \tanh(c_b) \tag{9}$$

Here, X_b = Input vector, H_{b-1} = Previous Cell Output, c_{b-1} = Previous Cell Memory, H_b = Current cell Output, c_b = Current cell Memory, $*$ = Element-wise multiplication, + = Element-wise Addition, W, U = Weight vectors for forget gate (f), candidate (c), input gate (l), and output gate (o).

3.6 Performance evaluation process

The proposed study uses a confusion matrix, accuracy score, precision, F1-score, recall, and AUC-ROC curve to measure the performance of the proposed CNN, CNN-LSTM models. In machine learning, a confusion matrix is used to evaluate the effectiveness of a classification algorithm [45, 46]. It offers a tabular view of expected versus actual class labels for a given set of data, allowing easy inspection of the model's capacity to correctly categorize cases into various classes. The confusion matrix is divided into four different categories: true positives, true negatives, false positives, and false negatives, as shown in figure 4.

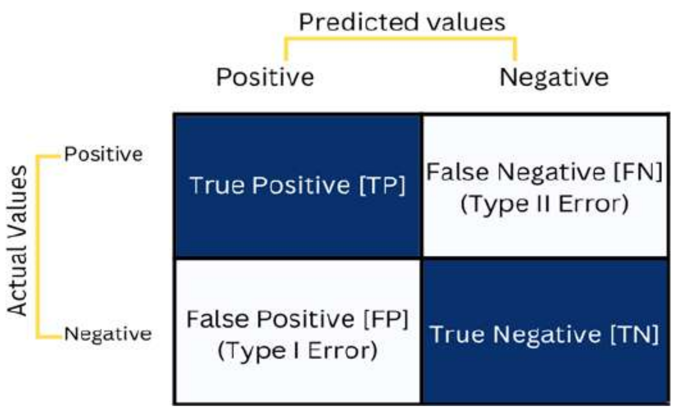

Figure 4. Show the confusion matrix sample

The following evaluation measure in equation (10) to (14) are applied to measure the performance of both models: Accuracy is defined as the percentage of accurately predicted instances among all instances:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Instances}} \tag{10}$$

Precision measures the capacity of the algorithm to prevent false positives by comparing true positive predictions with the total positive predictions:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{11}$$

Recall measures the percentage of positive cases the model accurately predicted:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{12}$$

The F1-score is a measure of how well a model achieves an appropriate balance between accuracy and recall.

$$F1\ \text{Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{13}$$

Matthews Correlation Coefficient (MCC) measures binary and multiclass classification model quality. The model's MCC score is calculated from true positives, true negatives, false positives, and false negatives:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{14}$$

4 Result and discussion.

In this section, we discuss the experimental setup, the simulation environment, and the experimental outcomes of the various classifiers used. Firstly, we discuss the experimental setup and tool used for conducting this experiment. Secondly, we analyzed the experimental results of our proposed models with an imbalanced dataset. Thirdly, we checked the experimental results of the proposed models with a balanced dataset. Fourthly, we compared the experimental results of the proposed models with both a balanced dataset and an imbalanced dataset. After analyzing performance, we compared the results of both imbalanced, balanced datasets. Finally, we compared the performance of the proposed models with previous studies.

4.1 Experimental Setup and parameters configuration

The proposed experiment was carried out on an HP G4 7th Generation system with 8 GB of RAM, an Intel (R) CoreTM i5, and a 256 SSD. The Jupyter Notebook IDE was used for the experimental and computational outcomes. Python packages such as Pandas, TensorFlow, Scikit-Learn, and Numpy are used to generate models and classify data. In addition, parameter configurations play a crucial role in improving the efficiency of ML models. Considering the importance of parameter configurations, we have summarized the parameter configurations of both models in Table 1.

Table 3. Parameters configuration of proposed models

Classifier	Parameters Configuration
CNN	Input shape=187x2x1, Kernel-size=5x1, Dropout rate=0.5, Strides=5x1, Optimizer=Adam, Filters=156x512
CNN-LSTM	Input shape=187x2x1, Kernel-size=10x2, Dropout rate=0.5, Strides=5x1, Optimizer=Adam, Filters=156x512

4.2 Experimental outcomes of CNN and CNN-LSTM with imbalanced data.

Propose study uses two different approaches with imbalanced dataset. Firstly, we utilized imbalanced dataset with ratio of 70/30 for training and testing of proposed CNN model, secondly the same imbalanced dataset was utilized for CNN-LSTM model. Various evaluation matrices are used to measure the performance of CNN and CNN-LSTM classifiers with imbalanced data is shown in Table 4.

Table 4. Performance metrics of CNN and CNN-LSTM models

Model	Type	Sample No.	Accuracy	Precision	F1-Score	Recall	MCC
CNN	Training	76,612	99.38	99.83	1.00	99.43	0.978
	Testing	32,834	98.62	99.65	0.99	98.84	0.951
CNN-LSTM	Training	76,612	98.67	99.74	0.99	97.63	0.974
	Testing	32,834	98.54	99.66	0.99	97.45	0.973

The data presented in Table 3 shows that CNN achieved 99.38 training accuracy, 98.6 testing accuracy, precision of 99.83 in training and 99.65 in testing, F1-Score of 1.00 and 0.99, and recall of 99.43 and 98.84 during training and testing, respectively. Similarly, CNN-LSTM obtained 98.67 training and 98.54 testing accuracy, precision of 99.74 in training, 99.66 in testing, F1-Score of 0.99, and recall of 97.63, 97.45 during training and testing phases with imbalanced data. To measure the performance of proposed classifiers more accurately confusion matric is calculated shown in figure 5.

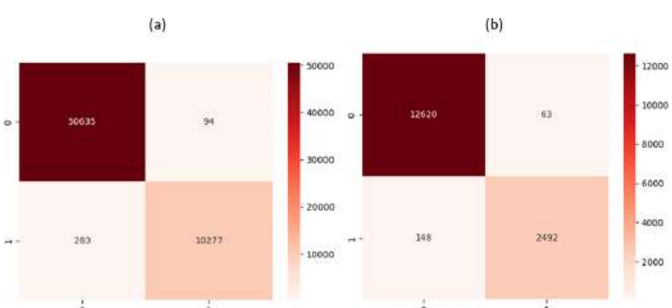

Figure 5. (a) shows the confusion matrix CNN with imbalanced dataset during training and figure 5 (b) show confusion matrix of CNN with imbalanced dataset during testing.

According to Figure 5(a), CNN predicted 50635 records correctly and 94 records incorrectly; out of 10,560 records, it successfully predicted 10277 records during the training phase. Similarly, according to the confusion matrix in 5(b), CNN predicted 12620 records correctly and 63 records incorrectly, out of 2,640 records, it successfully predicted 2492 records during the testing phase. The confusion matrices of CNN-LSTM model with imbalanced dataset is depicted in figure 6.

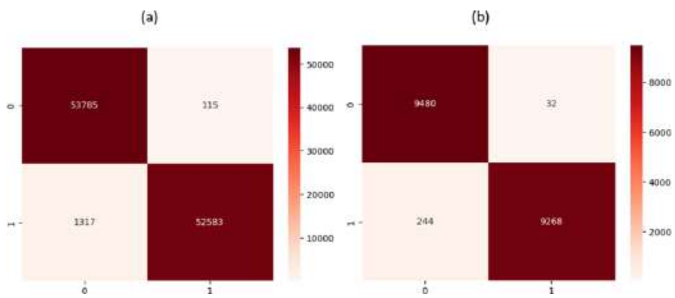

Figure 6. (a)Shows the confusion matrix CNN-LSTM with imbalanced dataset during training and figure 6 (b) show confusion matrix of CNN-LSTM with imbalanced dataset during testing

Similarly , in 6(a), CNN-LSTM predicted 53,785 records correctly and 115 records incorrectly, out of 53,900 successfully predicted 52,583 during training, while according to figure 6(b), it predicted 9480 records correctly and 32 records incorrectly, out of 9,512 successfully predicted 9268 records.

Learning curves play a crucial role in evaluating the performance of a model during training and validation phases. Figure 7 (a) show the accuracy graphs of CNN, and Figure 7 (b) show the accuracy graphs of CNN-LSTM with an imbalanced MIT-BIH ECG dataset.

The figure 7(a) depicts CNN achieved 98.62% accuracy, while 7(b) shows CNN-LSTM achieved 98.54% accuracy with MIT-BIH ECG signal Imbalanced dataset. In addition, it is observed that the accuracy plots obtained from the utilization of CNN and CNN-LSTM show variations during the first period. This discrepancy arises due to a substantial decline in the validation accuracy level in the case of CNN. However, as the training progresses, the accuracy plots of CNN eventually match with those of CNN-LSTM.

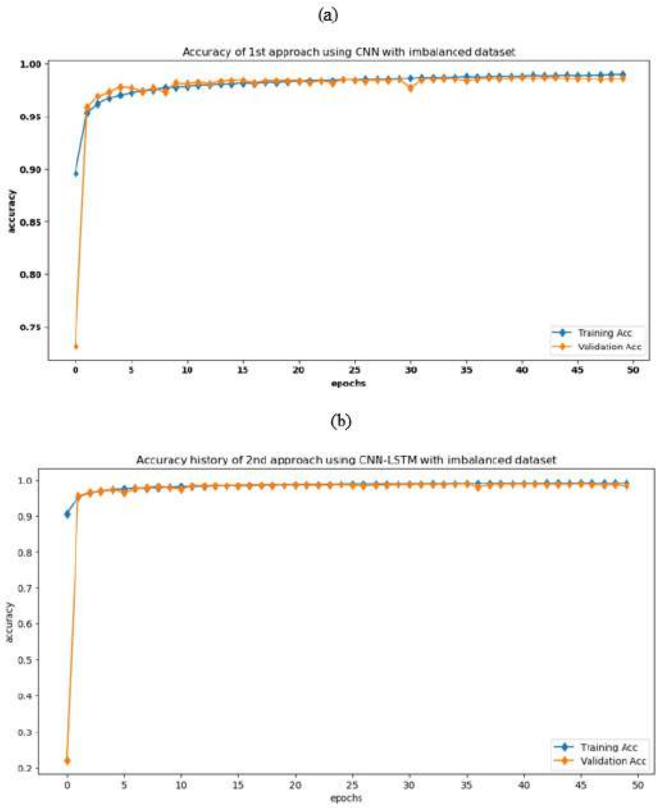

Figure 7. (a) shows the accuracy history of CNN and figure 7(b) shows the accuracy history of CNN –LSTM with imbalanced dataset.

The ROC curves of both CNN and CNN-LSTM models using MIT-BIH imbalanced dataset are also obtained and compared in figure 8(a) and 8(b).

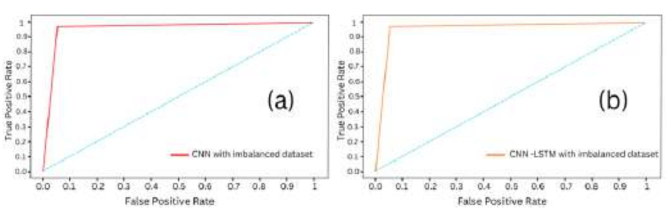

Figure 8. (a) Shown ROC of CNN, figure 8(b) shows the ROC of CNN-LSTM with imbalanced dataset

Figure 8(a) shows the ROC value of CNN which is 1.00 while figure 8(b) shows the receiver operating characteristic (ROC) value of CNN-LSTM with imbalanced dataset which is also 1.ROC values of both models are equal to one another, which shows that both the models are able to make decisions perfectly well.

4.3 Experimental outcomes of CNN and CNN-LSTM with balanced data

Using the second approach, we balanced the dataset using SMOTE and divided 70/30 for training and testing. Similar to first approach with imbalanced dataset the second approach also used two different models with balanced dataset. Firstly, we applied the CNN model and secondly we applied CNN-LSTM with balanced dataset. Using a balanced dataset, the results of the CNN and CNN-LSTM models are shown in Table 5.

Table 5. Performance metrics of CNN and CNN-LSTM models

Model	Type	Sample No.	Accuracy	Precision	F1-Score	Recall	MCC
CNN	Training	107,800	99.75	99.78	1.00	99.73	0.995
	Testing	19,024	99.24	99.20	0.99	99.28	0.985
CNN-LSTM	Training	107,800	99.75	99.58	1.00	97.63	0.995
	Testing	19,024	99.31	98.88	0.99	99.92	0.986

The result shows that CNN achieved 99.75 training and 99.24 testing accuracy, precision of 99.78 in training and 99.20 in testing, F1-Score of 1.00 and 0.99, recall of 99.73 for training and 99.28 testing, respectively. Similarly, CNN-LSTM achieved 99.75 training and 99.31 testing accuracy, precision of 99.58 in training and 98.88 in testing, F1-Score of 1.00 and 0.99, recall of 97.63 for training and 99.92 for testing, respectively. Figure 9 shows the confusion metrics of the proposed CNN model with a balanced dataset.

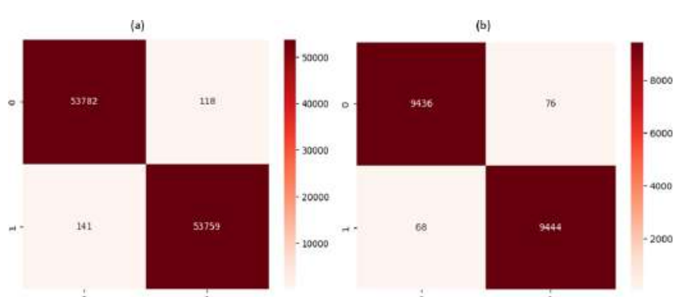

Figure 9. (a) shows the confusion matrix CNN with balanced dataset during training and figure 9(b) show confusion matrix of CNN with balanced dataset during testing.

According to Figure 9(a), CNN predicted 53782 records correctly and 110 records incorrectly. Out of 53,900 records, CNN successfully predicted 53759 records during the training phase. Similarly, according to the confusion matrix in 9(b), CNN predicted 9436 records correctly and 76 records incorrectly, out of 9,512 records, CNN successfully predicted 9444 records during the testing phase. Figure 10 shows the confusion metrics of the proposed CNN-LSTM model with a balanced dataset.

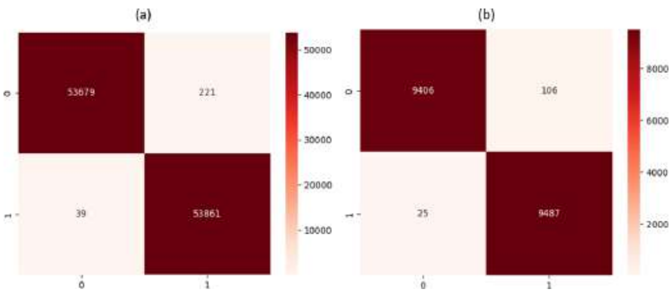

Figure 10. (a) shows the confusion matrix CNN-LSTM with balanced dataset during training and figure 10(b) show confusion matrix of CNN-LSTM with balanced dataset during testing

Likewise, 10(a) shows that CNN-LSTM predicted 53679 records correctly and 221 records incorrectly, out of 553,900 records, CNN-LSTM successfully predicted 53861 records during the training phase. Similarly, according to the confusion matrix in 10(b), CNN-LSTM predicted 9406 records correctly and 106 records incorrectly, out of 9,512 records, CNN-LSTM successfully predicted 9487 records during the testing phase. Figure 11 (a)) show the accuracy graphs of CNN, and Figure 11 (b) shows the accuracy graphs of CNN-LSTM with an imbalanced MIT-BIH ECG dataset.

As shown in figure 11(a) CNN achieved accuracy of 99.75% during training, and 99.24% accuracy during testing phase. Similarly, CNN-LSTM achieved accuracy of 99.75% during training, and accuracy of 99.31% during testing phase. Based on the result shown in table 4,5 and accuracy graph the CNN-LSTM outperformed with balanced dataset compared to the result with imbalanced dataset. The ROC curves of both classifiers using MIT-BIH balanced dataset is depicted and compared in figure 11(a) and 11(b).

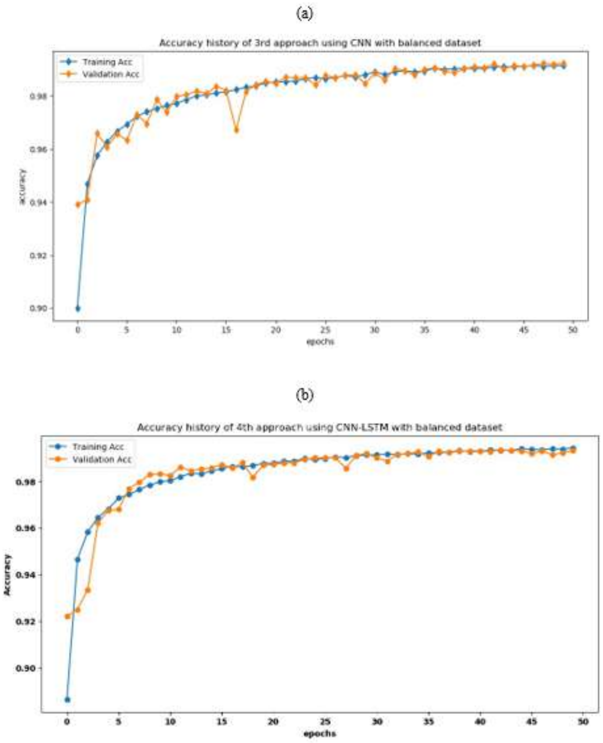

Figure 11. (a) shows the accuracy history of CNN and figure 11(b) shows the accuracy history of CNN –LSTM with imbalanced dataset.

According to Figure 11(a) and 11(b) the ROC value of 0.99 for both classifiers with balanced dataset indicates that both classification models have an extremely high predictive power and is doing an excellent job of separating the two groups (normal and abnormal)

4.4 Comparative analysis

We compared the proposed model results with previous studies. Our proposed study used CNN, and a hybrid of CNN and LSTM with the MIT-BIH arrhythmia dataset and achieved an accuracy of 99.31%, an F1-score of 0.99, and a recall of 99.92%. Table 6 depicts the comparison of previous studies with proposed research work.

4.5 Conclusion

This study suggests a hybrid model-based technique with the MIT-BIH arrhythmia dataset for heart disease diagnosis. The proposed study uses CNN and a hybrid of CNN and LSTM with both balanced and imbalanced

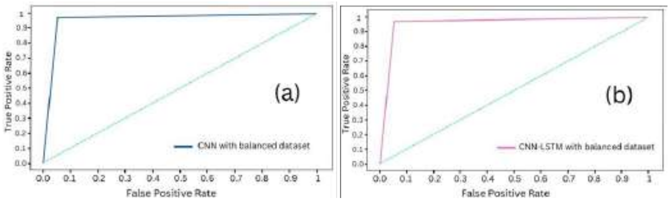

Figure 12. (a) shown ROC of CNN, figure 12(b) shows the ROC of CNN-LSTM with balanced dataset.

Table 6. Comparison of Proposed Work with Previous Studies

Author	Dataset Used	Accuracy	Year
Yar Muhammad et al. [17]	Cleveland and Hungarian dataset	94.4%	Nov 2020
S Mohan et al. [18]	UCI dataset	88.7%	JUN 2019
Repaka et al. [27]	Own collected dataset	86%	June 2019
Jian Ping Li et al. [31]	Cleveland dataset	92.37%	June 2020
Nafisa et al. [37]	MIT-BIH arrhythmia dataset	99.2%	March 2023
Proposed	**MIT-BIH arrhythmia dataset**	**99.31%**	—

datasets. Additionally, the dataset is standardized, normalized, and outliers are removed. Firstly, the CNN and CNN-LSTM models are trained with an imbalanced dataset. The outcomes with an imbalanced dataset were good but not satisfactory enough. Secondly, we balanced the dataset using the SMOTE technique, and again trained the CNN, CNN-LSTM models with a balanced dataset. Furthermore, various evaluation matrices are used to analyze the performance of classifiers with both balanced and imbalanced datasets. The proposed CNN model achieved 98.62% accuracy, precision of 99.65, f1-score of 0.99, MCC of 0.95, and recall of 98.84; CNN-LSTM achieved 98.54% accuracy, precision of 99.66, f1-score of 0.99, MCC of 0.97, and recall of 97.4 with an imbalanced dataset. Similarly, CNN achieved accuracy of 99.24%, precision of 99.20%, f1-score of 0.99, MCC of 0.985, recall of 99.28, and CNN-LSTM achieved accuracy of 99.31%, precision of 98.88, f1-score of 0.99, MCC of 0.98, and recall of 99.92 with a balanced dataset. It has been concluded through comparison with a recently published study that the suggested system can enhance accuracy performance by 0.1%. In the future, we will use various hybrid models and hybrid datasets (MIT-BIH and PTB-ECG) with some feature selection techniques to enhance the performance of the proposed study. Furthermore, the early diagnosis of chronic diseases such as Alzheimer, brain tumors, and lung cancer is crucial; therefore, we will work on these diseases too.

Author Contributions

Areebah: Conceptualization, Methodology, and Software,Investigation. **Muhammad Rabbi Butt**: Data curation, writing, and original draft preparation Supervision **Faheem Aslam**: Visualization, . Software, Validation. Writing- Reviewing and Editing

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

Funding Information

This paper has not received any funding.

References

[1] A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," *Nature Reviews Cardiology*, vol. 8, no. 1, pp. 30–41, 2010.

[2] K. Polat and S. Güneş, "Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and fft method based new hybrid automated identification system for classification of eeg signals," *Expert Systems with Applications*, vol. 34, no. 3, pp. 2039–2048, 2008.

[3] World Population Review, "Heart disease rates by country 2023," aug 2023. Accessed: 2023-08-04.

[4] M. Durairaj and N. Ramasamy, "A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate," *JOUR*, pp. 255–260, Jan 2016.

[5] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.

[6] L. A. Allen *et al.*, "Decision making in advanced heart failure," *Circulation*, vol. 125, no. 15, pp. 1928–1952, 2012.

[7] H. Yang and J. M. Garibaldi, “A hybrid model for automatic identification of risk factors for heart disease,” *Journal of Biomedical Informatics*, vol. 58, 2015.

[8] H. R. Al-Absi *et al.*, “Cardiovascular disease diagnosis from dxa scan and retinal images using deep learning,” *Sensors*, vol. 22, no. 12, p. 4310, 2022.

[9] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, “Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm,” *Computer Methods and Programs in Biomedicine*, vol. 141, pp. 19–26, 2017.

[10] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, “An integrated decision support system based on ann and fuzzy ahp for heart failure risk prediction,” *Expert Systems with Applications*, vol. 68, pp. 163–172, 2017.

[11] W. Song, “A new method for refined recognition for heart disease diagnosis based on deep learning,” *Information*, vol. 11, no. 12, p. 556, 2020.

[12] S. Arooj *et al.*, “A deep convolutional neural network for the early detection of heart disease,” *Biomedicines*, vol. 10, no. 11, p. 2796, 2022.

[13] M. T. García-Ordás *et al.*, “Heart disease risk prediction using deep learning techniques with feature augmentation,” *Multimedia Tools and Applications*, vol. 82, no. 20, pp. 31759–31773, 2023.

[14] A. Methaila, P. Kansal, H. Arya, and P. Kumar, “Early heart disease prediction using data mining techniques,” *Computer Science & Information Technology (CS & IT)*, 2014.

[15] R. Detrano *et al.*, “International application of a new probability algorithm for the diagnosis of coronary artery disease,” *The American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, 1989.

[16] M. Gudadhe, K. Wankhade, and S. Dongre, “Decision support system for heart disease based on support vector machine and artificial neural network,” in *2010 International Conference on Computer and Communication Technology (ICCCT)*, 2010.

[17] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, “Early and accurate detection and diagnosis of heart disease using intelligent computational model,” *Scientific Reports*, vol. 10, no. 1, 2020.

[18] S. Mohan, C. Thirumalai, and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” *IEEE Access*, vol. 7, pp. 81542–81554, 2019.

[19] N. C. Long, P. Meesad, and H. Unger, “A highly accurate firefly based algorithm for heart disease prediction,” *Expert Systems with Applications*, vol. 42, no. 21, pp. 8221–8231, 2015.

[20] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.

[21] R. Alizadehsani *et al.*, “Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries,” *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 119–127, 2018.

[22] J. Vijayashree and H. P. Sultana, “A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier,” *Programming and Computer Software*, vol. 44, no. 6, pp. 388–397, 2018.

[23] A. M. S. Ismaeel and D. Chourishi, “Using the extreme learning machine (elm) technique for heart disease diagnosis,” in *2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015)*, 2015.

[24] K. S. Purushottam and R. Sharma, “Efficient heart disease prediction system,” *Procedia Computer Science*, vol. 85, pp. 962–969, 2016.

[25] C. M. B. et al., “Effective heart disease prediction using machine learning techniques,” *Algorithms*, vol. 16, no. 2, p. 88, 2023.

[26] J. Ramos, “Ieee editorial style manual.” *IEEE Publishing Operations*, 2007. [Online]. Available: https://www.ieee.org/content/dam/ieee-org/ieee/web/org/pubs/ieee/editorial/style/manual.pdf.

[27] F. S. Alotaibi, “Implementation of machine learning model to predict heart failure disease,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.

[28] N. Hasan and Y. Bao, “Comparing different feature selection algorithms for cardiovascular disease prediction,” *Health and Technology*, vol. 11, no. 1, pp. 49–62, 2020.

[29] S. M. Malakouti, “Heart disease classification based on ecg using machine learning models,” *Biomedical Signal Processing and Control*, vol. 84, p. 104796, 2023.

[30] J. Abdollahi and B. Nouri-Moghaddam, “A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation,” *Iran Journal of Computer Science*, vol. 5, no. 3, pp. 229–246, 2022.

[31] J. P. L. et al., “Heart disease identification method using machine learning classification in e-healthcare,” *IEEE Access*, vol. 8, pp. 107562–107582, 2020.

[32] R. D. et al., “International application of a new probability algorithm for the diagnosis of coronary artery disease,” *The American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, 1989.

[33] P. D. et al., “An automatic heart disease prediction using cluster-based bi-directional lstm (c-bilstm) algorithm,” *Neural Computing and Applications*, vol. 35, no. 10, pp. 7253–7266, 2022.

[34] P. M. S. Krishnan and R. Ibrahim, “Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5467–5476, 2021.

[35] X. L. et al., “A hybrid classification system for heart disease diagnosis based on the rfrs method,” *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–11, 2017.

[36] C. T. S. Mohan and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” *IEEE Access*, vol. 7, pp. 81542–81554, 2019.

[37] M. A. H. N. Anjum, K. A. Sathi and M. A. Dewan, “A temporal transformer-based fusion framework for morphological arrhythmia classification,” *Computers*, vol. 12, no. 3, p. 68, 2023.

[38] A. B. et al., “A supervised classification phenotyping approach using machine learning for patients diagnosed with primary breast cancer,” *i-manager’s Journal on Computer Science*, vol. 11, no. 1, p. 1, 2023.

[39] I. A. S. U. G. Z. Khan, Ikram Ulhaq and I. Ullah, “A privacy-preserving-based technique for customer churn prediction in telecom industry,” *VFAST Transactions on Software Engineering*, vol. 11, no. 3, pp. 73–80, 2023.

[40] G. Z. K. et al., “An efficient deep learning model based diagnosis system for lung cancer disease,” in *2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2023.

[41] S. I. U. N. K. G. Zaman, R., “A robust deep learning-based approach for breast cancer detection from histopathological images,” in *Proceedings of the 4th International Electronic Conference on Applied Sciences*, 2023.

[42] S. U. A. Spogmay Yousafzai, G. Z. Khan and M. R. Butt, “Improved neural network-based system for early and accurate diagnosis of alzheimer disease,” *Journal of Computer Science and Technology Studies*, vol. 5, no. 1, pp. 58–69, 2023.

[43] A. S. et al., “Predicting the growth of the heart disease mortality rate in the united states using recurrent neural network-based long short-term memory,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 16, p. 5730, 2020.

[44] F. M. Hosseini and M. Salimi, “A hybrid evolutionary model for heart disease prediction based on gaussian mixture models,” *Journal of Biomedical Informatics*, vol. 119, p. 103833, 2021.

[45] H. J. P. et al., “Using deep learning and semantic segmentation to extract disease relevant information from chest x-rays for improved diagnosis and treatment,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2403–2413, 2021.

[46] C. Y. L. et al., “An explainable ai approach for personalized treatment of heart failure using multi-modal data,” *Artificial Intelligence in Medicine*, vol. 114, p. 102049, 2021.