RNN-GRU HYBRID DEEP LEARNING FRAMEWORK FOR HEART DISEASE PREDICTION

Nadikatla Chandrasekhar[1]– Samineni Peddakrishna[2*]– Sreedhar Kollem[3]
[1]Department of ECE, GMR Institute of Technology, Rajam, Vizianagaram, Andhra Pradesh, India- 532127
[2]School of Electronics Engineering, VIT-AP University, Amaravathi, Guntur, Andhra Pradesh, India- 522241
[3]Department of ECE, School of Engineering, SR University, Warangal, Telangana, India-506371

ARTICLE INFO

***Article history*:**
Received: 7. 1. 2024.
Received in revised form: 18. 3. 2025.
Accepted: 20. 5. 2025.

Keywords:
Heart disease prediction
Hybrid deep learning
Data pre-processing
RNN-GRU

DOI: 10.30765/er.2434

Abstract:

This research presents a novel hybrid deep learning approach for heart disease prediction, incorporating a Recurrent Neural Network (RNN) with multiple Gated Recurrent Units (GRU) and the Adam optimizer. The performance of the proposed model is evaluated on the IEEE Dataport heart disease dataset, achieving an accuracy of 92%. The primary focus of this research is to investigate the impact of data pre-processing, particularly outlier detection, on the accuracy of the hybrid deep learning model. The comparative analysis demonstrates that without outlier detection, the model achieves an accuracy of 89%, while incorporating outlier detection results in an increase in accuracy of 3%, reaching 92%. The study involves implementing a hybrid RNN-GRU model on a pre-processed dataset. Research tracks the accuracy and loss for each epoch throughout the training and testing phases. The results highlight the usefulness of the model in predicting heart disease and stress the importance of data pre-processing in improving its performance.

1 Introduction

The human heart plays an incredibly vital role in the body by ensuring the distribution of oxygen-rich blood to different body parts, which is a critical function. Taking care of heart health is of utmost importance. Heart disease has become a major global problem [1]. As time progresses, people's lifestyles have changed significantly, with many adopting unhealthy dietary habits, avoiding exercise, experiencing increased stress, suffering from sleep deprivation, smoking excessively, and eating junk food [2]. All of these factors significantly impact a person's heart health [3]. Due to the widespread prevalence of heart disease throughout the world, the mortality rate has risen dramatically, particularly in low-income countries [4]. This is mainly because these countries need more technological advancements and allocate more resources to the healthcare sector. Therefore, it is imperative to implement cost-effective disease prevention strategies in these regions. Detecting and addressing heart disease before they manifest is challenge, so people must be vigilant and take preventive measures [5]. These preventive measures are crucial in safeguarding human lives from heart disease [6]. Furthermore, embracing the latest technologies can help create a healthier world. Machine learning (ML) technologies are playing a significant role in early detection of heart disease [7], [8]. ML models can identify potential heart issues with a high degree of accuracy, although they require some time and effort to train, especially when dealing with large patient datasets. Moreover, ML models continually evolve, striving to uncover intricate patterns hidden within data. In this pursuit, we also turn to Deep Learning (DL) models for heart disease prediction [9]. These models are becoming increasingly adept at training to make exceedingly accurate predictions. Hybrid DL models offer several advantages over traditional DL models. They excel in combining different DL algorithms to extract and learn patterns from the data more efficiently. As a result, these models gradually improve their performance. Additionally, integrating DL with other techniques is an effective form of regularization, mitigating the risk of over-fitting and enhancing generalization. These hybrid

* Corresponding author
E-mail address: krishna.samineni@gmail.com

models can be likened to ensemble learning, where multiple models collaborate to make predictions, a strategy known for its ability to produce superior performance and more excellent stability.

This research presents the application of a hybrid DL model of recurrent neural networks and gated recurrent unit (RNN-GRU) in implementing collaborative classification approaches for heart disease prediction. We used an IEEE Dataport heart disease dataset to conduct this research.

The following noteworthy study is presented in this work:

- The article highlights the impact of data pre-processing techniques on the model's accuracy by showing the difference in accuracy between the dataset with and without outliers.
- A hybrid DL model of RNN and GRU is analysed and implemented on the IEEE data port heart disease dataset. The aim is to evaluate its performance in terms of classification values.
- To obtain the highest accuracy, the study employs gridsearchCV hyperparameter and 5-fold cross-validation methods.

2 Literature review

DL technologies are finding extensive applications in the medical field. In particular, they are instrumental in predicting heart diseases. Recent technological advancements have enabled DL techniques to yield excellent results in predicting heart diseases [10],[11]. However, due to the utilization of these techniques, a problem known as overfitting is mitigated, leading to accurate predictions of heart diseases [12].DL utilizes a hierarchy of artificial neural networks to conduct an in-depth dataset analysis. This process is modelled after the human brain, with the input being processed through multiple hidden layers, each containing neurons, before reaching the final output. The neural network's neurons, also called nodes, perform in-depth analysis of specific data using predefined algorithms and subsequently relay partially processed information to the subsequent layer [13]. DL employs a nonlinear approach to data processing, where all inputs are interconnected, resulting in favourable outcomes [14]. In a neural network, the first layer gathers input data, processes it, and passes it on to the next layer. As we progress through the layers of neurons in a DL neural network, they carefully analyse data from previous stages before making decisions and generating outcomes [15]. Deep learning (DL), due to its capacity for handling extensive datasets and recognizing intricate patterns, holds promise for enhancing heart disease prediction [16]. In recent times, several studies have explored the efficacy of DL in this context.

For instance, Tuli et al. conducted a research study centered on Health Fog, a system that integrates Internet of Things (IoT) sensors and fog computing nodes to gather and process medical data, encompassing electrocardiogram (ECG) signals and clinical information [17]. Subsequently, this system employs an ensemble of DL models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks, to analyse the data to diagnose heart diseases. The findings from this research demonstrate the considerable potential of DL in elevating the accuracy of heart disease prediction, mainly when dealing with intricate datasets. Butchi et al. propose a system that collects and processes medical data from various IoT devices [18], including wearable sensors, smartwatches, and smartphones, and applying a cascaded DL model to predict heart disease. The cascaded DL model includes a feature extraction module based on a CNN and a classification module based on a deep belief network (DBN) [19]. Surenthiran et al. discuss a range of DL methods and compare their performance with the proposed model. Their model employs multiple Gated Recurrent Units (GRU) to enhance the RNN model's performance, resulting in an impressive accuracy of 98.4%. The data used for this study is sourced from the Cleveland dataset available in the UCI Repository [20]. On the other hand, Girish et al. put forward a hybrid DL approach for classifying cardiac diseases. Their proposed system combines RNN and LSTM hybrid approaches to classify synthetic data. They employ various ML and soft computing techniques to evaluate the system's performance. These hybrid DL algorithms showcase an average classification accuracy of roughly 95.10% when subjected to multiple cross-validation assessments. The research ultimately establishes that deep hybrid learning surpasses both conventional DL and independent machine learning approaches in accuracy [21]. Table 1 includes a more detailed comparison with existing models in heart disease prediction.

Table 1. Detailed comparison of existing models.

Author	Dataset	Methodologies used	Limitations
B.Mali et.al [22]	IEEE Dataport Heart disease dataset.	Used the artificial neural network as the federated learning (FL) base model for heart disease prediction.	In the data preprocessing not used outliers detection method.
N.K Sharma et.al [23]	IEEE Dataport Heart disease dataset.	Used the voting ensemble ML method.	Not used hybrid DL models. Outliers in the dataset were not detected.
A.Ahmad et.al [24]	IEEE Dataport Heart disease dataset.	K-nearest neighbor (KNN), random forest (RF), support vector machine (SVM), decision tree (DT), and gradient boosting (GB). And neural network (NN) are used to predict the HD.	DL models and hybrid models were not used. Outlier detection is not done in the data preprocessing stage.
S. Mondal et.al [25]	IEEE Dataport Heart disease dataset.	ML classifiers are used to build the initial prediction model, with 10-fold cross-validation ensuring its robustness. RandomizedSearchCV and GridSearchCV are applied for hyperparameter tuning. The top-performing models—RF, XGBoost, and Decision Tree—are further refined using a stacking ensemble to enhance accuracy.	This work is limited to traditional ML classifiers, without using DL models or outlier detection during preprocessing.
S.M. Alanazi et.al [26]	IEEE Dataport Heart disease dataset.	ML techniques—Auto-WEKA, Decision Table/Naive Bayes (DTNB), and Multiobjective Evolutionary (MOE) fuzzy classifier—are used to develop diagnostic models.	This work includes the lack of parameter tuning, which may affect model optimization. Outlier detection is not applied during preprocessing, potentially impacting accuracy. The study relies solely on traditional ML techniques, without exploring (DL) or hybrid DL models, which could enhance predictive performance.
Navita et. al [27]	IEEE Dataport Heart disease dataset.	A stacking ensemble model with three base learners: Random Forest Tree (RFT), K-Nearest Neighbor (K-NN), and AdaBoost, along with Logistic Regression (LR) as the meta-learner, optimized using Grid Search Cross-Validation (GSCV).	It lacks DL and hybrid DL models, as well as outlier detection, which may affect accuracy.
B.P. Doppala et.al [28]	IEEE Dataport Heart disease dataset.	Used NB, SVM, RF, XGB ML models, and Ensembled voting classifier.	DL models were not used, and Outlier detection was not performed.

M. Ferdowsi et .al [29]	IEEE Dataport Heart disease dataset.	Heart disease prediction used data anonymization with Laplace noise and a differential privacy (DP) framework. Models included LR, GNB, and RF, with SHAP and LIME for interpretability.	The study has some limitations: it did not use DL or hybrid models, and outlier detection was not applied during data preprocessing.

3 Materials and methods

This section explains how heart disease predictions are made based on the IEEE dataport heart disease dataset [30]. This work comprises multiple stages, from data collection to heart disease prediction. Outlier detection, one-hot encoding, and standardization are performed in the initial data preprocessing phase. The dataset is split into two parts: 80% for training and 20% for testing. After this, a hybrid DL model is created using RNN and GRU. To train the model, a grid search CV is applied with 5-fold cross-validation. Once the best hyperparameters are found, the model is tested using these settings. Finally, the model's performance is checked using metrics like accuracy, precision, recall, and F1-score. Figure 1 shows the workflow methodology, while Figure 2 comprehensively illustrates the proposed architecture design.

Figure 1. Proposed model workflow

3.1 Description of the dataset

This study used an IEEE Dataport dataset that contains 1190 instances of heart disease. After eliminating 272 duplicate rows, the dataset now includes 918 unique samples, 508 related to patients with heart disease and 410 to healthy patients. Figure 3 illustrates the distribution of heart disease instances based on the target variable. The dataset includes continuous features such as "age," "resting bp," "cholesterol," "maxhr," and "old peak," along with categorical attributes like "chest pain type," "resting ECG," and "st_slope," and binary features such as "sex," "fasting bs," and "exercise angina." These features were used to analyse potential risk factors for heart disease, and more details can be found in the accompanying Table 2.

A correlation heat map Figure 4 illustrates the relationships in this dataset between variables related to heart disease. The correlation coefficients measure these relationships, with 1 indicating strong positive correlation, -1 indicating a strong negative correlation, and 0 indicating that there is no correlation. In particular, age is positively correlated with factors such as resting blood pressure and cholesterol levels but negatively correlated with maximum heart rate and presence of heart disease. Furthermore, the heat map highlights variables such as the type of chest pain and the maximum heart rate as highly associated with the target variable, suggesting their potential in predicting heart disease. Effective predictive use requires extensive data analysis and pre-processing, including outlier removal and data standardization, to ensure model accuracy and stability.

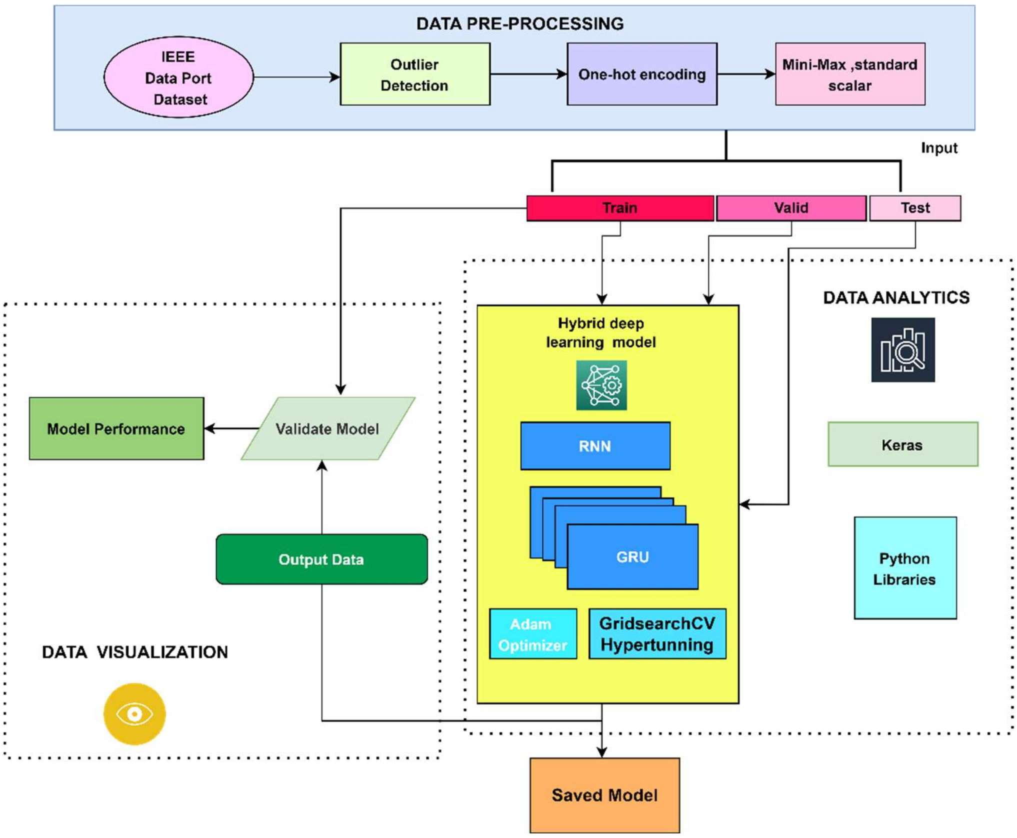

Figure 2. Proposed work architecture

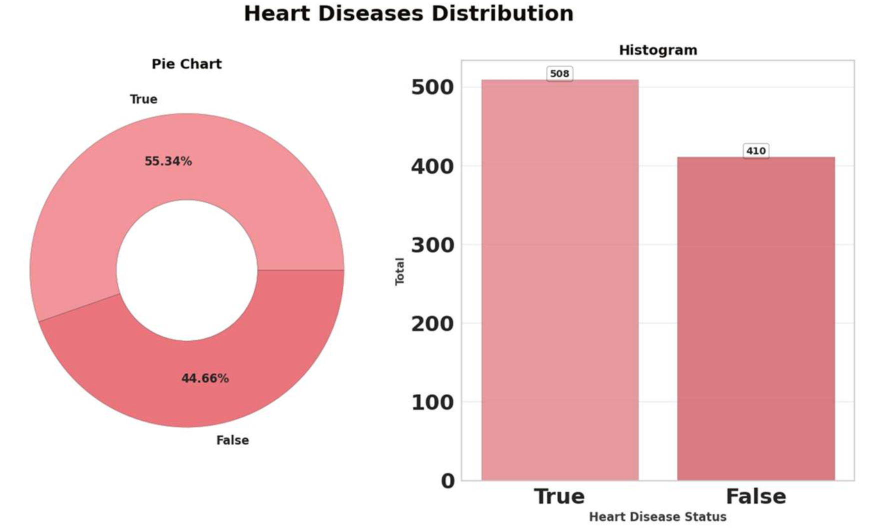

Figure 3. Heart disease distribution

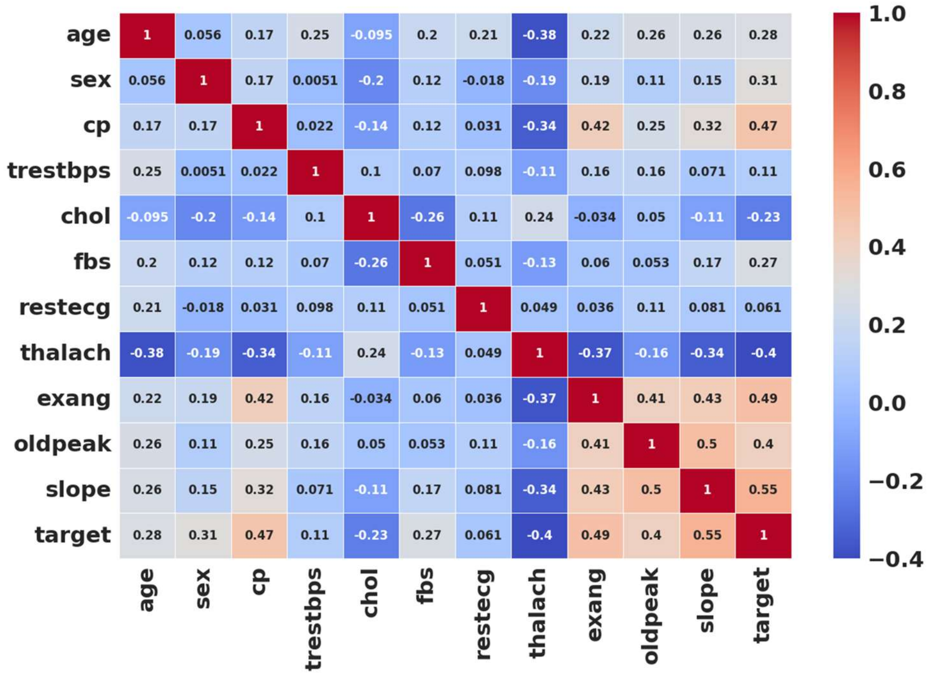

Figure 4. Heart disease attributes heat map

Table 2. Description of the IEEE dataport heart disease dataset.

Sr.No	Attribute Icon	Attribute Name	Description
1	Age	Age	Age of the patient
2	Sex	Gender	To denote males, assign a value of 1, and for females, use 0
3	Cp	Chest pain type	Chest discomfort can be categorized into four groups:0-angina,-1-atypical angina, -2-non-anginal pain, and -3-asymptomatic
4	Trestbps	Rest state blood	Resting BP upon hospital admission, measured in mm/Hg

		pressure(mm/Hg)	
5	Chol	Serum cholesterol (Fat)	Blood cholesterol level measured in mg/dL.
6	Fbs	Fasting blood sugar (not eating)	A fasting blood sugar level above 120 mg/dl after an overnight fasting period is classified as elevated (1 - true). Conversely, if it registers below 120 mg/dl, it is within the normal range (0 - false)
7	Restecg	Rest ECG test	The Resting ECG test and its associated findings can be categorized into three groups: a score of 0 denotes a normal result, a score of 1 indicates the detection of ST-T wave abnormalities, and a score of 2 signifies the presence of left ventricular hypertrophy
8	Thalach	Max Heart rate achieved	The peak heart rate experienced during physical activity
9	Exang	Exercise-induced angina	Angina onset during exercise is indicated with 0 representing the absence of angina and 1 denoting its presence
10	Oldpeak	ST depression (ECG test)	During an ECG test, ST depression is typically observed when comparing the results obtained during exercise to those during relaxation
11	Slope	Slope (ST depression)	The highest degree of slope (ST depression) during exercise testing is classified as follows: 1 for an upward slope, 2 for a level slope, and 3 for a downward slope
12	Target	Heart failure class attribute	A label of "0" indicates the absence of heart disease, while a label of "1" indicates the presence of heart disease

3.2 Data pre-processing

3.2.1 Outlier's removal

Data quality and organization play an essential role in the performance of predictive models. Outliers and unreasonable values in raw datasets can hinder the accuracy and reliability of predictive models. In applications like heart disease dataset analysis, this is especially important. We addressed this issue using the interquartile range (IQR) method, a robust statistical technique [31]. The IQR method allows outliers to be detected and handled systematically and objectively, improving the reliability and accuracy of our heart disease prediction model. Physiological indicators for healthy individuals generally show values within a range for healthy individuals. However, significant deviations in biological markers can serve as crucial indicators of underlying illnesses. Therefore, the heart disease prediction model should be able to identify specific outliers for further investigation rather than simply eliminating all outliers when addressing outliers. This study explicitly uses the IQR method to identify outliers in two important columns in cholesterol and resting blood pressure. These columns typically follow a typical distribution pattern, but we noticed conspicuous deviations from the expected range when examining their boxplots.

As part of the IQR process, the dataset is sorted in ascending order; then the median is determined; after that, the dataset is divided into two equal parts at the median, making up the lower and upper halves. Using equation (1), the IQR is calculated by subtracting the first quartile Q1 from the third quartile Q3.

$$IQR = Q3 - Q1 \tag{1}$$

Where Q1 and Q3 are the medians of the lower and upper halves of the dataset, respectively. When data points are significantly below Q1 or above Q3, they can be identified as outliers since they lie outside of the following ranges, as shown in equations (2) and (3).

$$\text{Lower bound} = Q1 - 1.5 * IQR \tag{2}$$

$$\text{Upper bound} = Q3 + 1.5 * IQR \tag{3}$$

Data points outside this range are deemed potential outliers, and the handling of data points outside of this range is critical for ensuring the integrity of the dataset and the accuracy of subsequent analyses. Using these bounds, the data points outside these ranges are removed from the dataset, a common practice to minimize the influence of outliers. To visualize the effects of outlier processing on data distribution, an analysis of the differences in data distribution before and after outlier removal was performed. Figure 5, Figure 6, Figure 7, and Figure 8 illustrate these changes, providing a visual confirmation of the effectiveness of outlier removal. According to the dataset, a substantial portion of the cholesterol and resting blood pressure columns contained outliers (19.93% and 4.29%, respectively). These outliers could significantly skew the results, resulting in a less accurate predictive model. To address this issue, the outliers have been removed the outliers. Additionally, it has been identified some records that contain zero cholesterol values. There may have been human errors in these records, which have been removed. Finally, any missing values in the cholesterol and resting blood pressure columns are credited with the mean value. Considering that it is relatively robust to outliers and improved the accuracy and reliability of heart disease prediction.

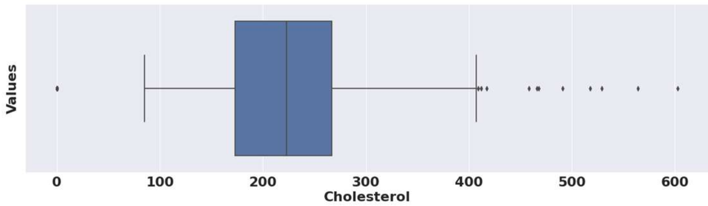

Figure 5. Cholesterol box plot with contain outliers

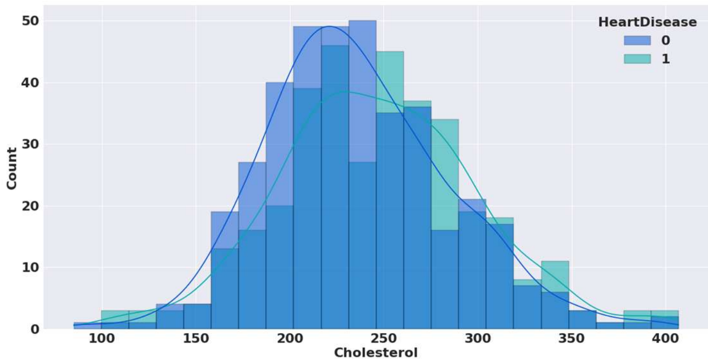

Figure 6. Cholesterol histogram plot without outliers

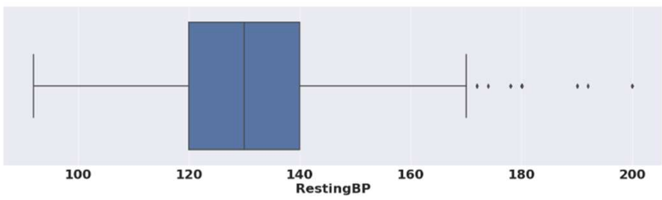

Figure 7. RestingBP box plot with contain outliers

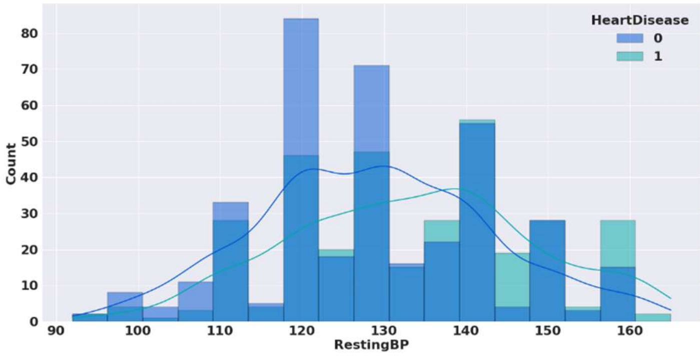

Figure 8. Resting BP histogram plot without outliers

3.3 Data Encoding and Standardization

Following removing outliers, handling categorical data through encoding, and standardizing numerical data is often performed as pre-processing steps. Data encoding is typically used to handle categorical variables, which represent discrete labels or categories. One-hot encoding changes categorical variables, such as gender, chest pain type, and exercise-induced angina, into a numerical format. Each category in the variable gets its unique number. This helps DL models use clear information while keeping the data simple and easy to understand. On the other hand, data standardization scales numerical features so they all have the same range. This is done by adjusting the data so that each feature has a mean of 0 and a standard deviation of 1. To achieve this, the mean is subtracted from each value, and then it is divided by the standard deviation. This makes sure all features are equally balanced for the model. Such an approach converges faster, making them less sensitive to the scale of input features. Because the feature size is consistent, the algorithm can update its parameters more evenly and efficiently. In the context of the present dataset, different variables like age, chol, etc., have been standardized. The summary statistics for the standardized features are presented in Table 3. These measures include mean such an approach converges faster, making, standard deviation, minimum, and quartile.

Table 3. Description of the IEEE Dataport heart disease dataset.

Feature	Count	Mean	Std	Min	25%	50%	75%
age	918.00	-0.00	1.0005	-2.706	-0.690	0.051	0.688
trestbps	918.00	0.00	1.0005	-7.154	-0.669	-0.129	0.410
chol	918.00	0.00	1.0005	-1.818	-0.233	0.221	0.623
thalach	918.00	0.00	1.0005	-3.018	-0.660	0.046	0.754
oldpeak	918.00	0.00	1.0005	-3.271	-0.832	-0.269	0.574

3.4 Performance measuring matrices

Performance evaluation metrics play a crucial role in assessing the efficiency and effectiveness of DL models. These metrics help identify a model's strengths and weaknesses and determine the areas that need improvement. Here are some of the most common performance measuring metrics for DL models:

Accuracy: accuracy is a common practice for assessing classification models, as highlighted in Rapaka's work [32]. It quantifies the proportion of correctly classified samples, which is a fundamental measure. It is important to note that accuracy may lead to misleading conclusions, particularly in imbalanced datasets. Equation (4) shows accuracy.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \tag{4}$$

Precision: Precision measures the proportion of true positives accurately identified among all the instances predicted as positive. This metric holds particular importance in situations where the implications of false positives come with substantial consequences, which is frequently observed in the field of medical diagnosis. Equation (5) shows the precision of the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

Recall: Recall measures the proportion of true positive cases out of all positive cases. This metric is handy when the potential repercussions of missing true positives are substantial, as seen in applications like fraud detection. Equation (6) represents the recall of the model.

$$\text{Re}call = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

F1-score: The F1-score functions as a metric that strikes a balance between precision and recall by taking their harmonic mean. This makes it a valuable tool for assessing performance in scenarios where false positives and negatives have substantial consequences. Equation (7) represents the F1-score of the model.

$$\text{F1-Score=} \frac{2 \times (\text{Precision} \times \text{recall})}{\text{Precision+Recall}} \tag{7}$$

ROC-AUC: The ROC-AUC metric evaluates a model's ability to distinguish between positive and negative samples by measuring how well it correctly identifies true positives and true negatives while also minimizing false positives and false negatives.

4 Proposed Methodology

The GRU marks a notable progression within the domain of RNNs, aimed at addressing a persistent challenge called the vanishing gradient problem, a common issue encountered in conventional RNNs. This innovation was introduced by Cho et al. [33]. In their seminal work, it has gained considerable popularity for its effectiveness in handling sequence-to-sequence learning tasks. The GRU architecture comprises two key components: the updating gate denoted as "z" and the reset gate labelled as "r." These gates are instrumental in determining how much the model should retain or forget information from the previous hidden state at each time step [34].In GRU notation, "x_t" is the input at time step "t," and "h_t" is the hidden state at the same time. The weight matrices are written as "W_z," "W_r," "W_h," "U_z," "U_r," and "U_h," and the bias terms are "b_z," "b_r," and "b_h." The sigmoid function is shown as "sigma," and the tanh function is written as "tanh." The update gate, "z_t," decides how much of the previous hidden state should be updated with the new candidate, as shown in equation (8). The reset gate, "r_t," in equation (9), controls how much of the previous hidden state should be ignored when calculating the new candidate. The new hidden state candidate, "h_tilde," is calculated using the reset gate and the input, as in equation (10). Finally, the updated hidden state, "h_t," is a mix of the previous hidden state and the new candidate, with the update gate deciding the balance, as shown in equation (11). GRUs are a type of RNN architecture that can be used to model sequential data. GRUs leverage activation functions to enable learning of long-term dependencies in data and control the information flow within the network. Two main activation functions are used in GRUs: the sigmoid function and the hyperbolic tangent (tanh) function. GRUs are a type of RNN architecture designed for modelling sequential data. They utilize activation functions to facilitate learning of long-term dependencies within the data and regulate the flow of information throughout the network. GRUs employ two primary activation functions: the sigmoid function and the hyperbolic tangent (tanh) function. GRUs employ the sigmoid function to regulate information flow via their update and reset gates, while they utilize the tanh function to calculate the veiled state of the candidate. These activation functions are crucial in the GRU's ability to learn and model sequential data.

$$z\_t = \sigma\left(W\_z * x\_t + U\_z * h\_(t-1) + b\_z\right) \tag{8}$$

$$r\_t=\sigma\left(W\_r*x\_t+U\_r*h\_(t-1) +b\_r\right) \tag{9}$$

$$h\_tilde = \tanh\left(W\_h * x\_t + U\_h * \left(r\_t\, e\, h\_(t-1)\right) + b\_h\right) \tag{10}$$

$$h\_t = \left(1 - z\_t\right) e\, h\_(t-1) + z\_t\, e\, h\_tilde \tag{11}$$

In this study's proposed method, a GRU-based RNN is implemented for the binary classification of a heart disease dataset (IEEE dataport dataset) using the Keras library, as illustrated in Figure 9. The methodology involves several stages: Initially, the required libraries and modules are imported, such as the Sequential model, GRU and Dense layers, Adam optimizer, and tools for model evaluation. Subsequently, the input data is modified into a 3-dimensional array to meet the GRU layer's requirements. The training dataset is divided into training and validation sets, using an 80:20 split. This division is essential for evaluating the model's performance during training. A custom function is established to construct and compile the model, which includes a GRU layer followed by a Dense layer that utilizes a sigmoid activation function [35]. The configuration of the model involves the utilization of the Adam optimizer alongside the binary cross-entropy loss function. It is then encapsulated within a Keras Classifier to enable seamless integration with Scikit-learn tools. A parameter grid is defined to fine-tune hyperparameters, covering aspects like the number of hidden units, learning rate, and batch size. A GridSearchCV object is created, which includes the underlying model, the parameter grid, and a 5-fold cross-validation approach. This grid search process entails training the model using the training dataset and assessing its performance on the validation set. After completion, the training history is extracted from the model that performs the best identified during the grid search. Lastly, the best hyperparameters and the performance of all combinations of hyperparameter are displayed in Table 4. This approach enables the optimization and evaluating of the GRU-based RNN model for heart disease classification.

Table 4. Best hyperparameter tuning values.

Hybrid deep learning model	Best hyperparameters provide by using grid searchcv 5-fold cross-validation
RNN+GRU	'batch_size': 64, 'lr': 0.001, 'units': 128

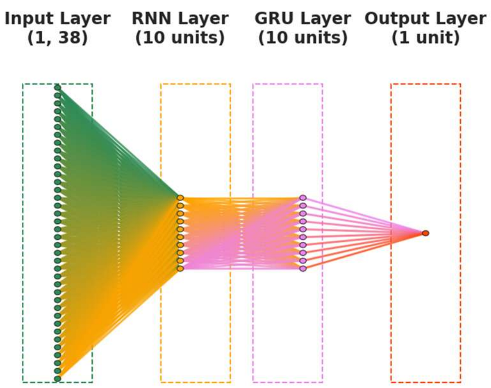

Figure 9. Proposed model neural network architecture

5 Results and discussion

The proposed method for predicting heart disease was evaluated using an IEEE Dataport heart disease dataset. Before analysis, the dataset was pre-processed with the removal of outliers, and the impact of this pre-processing step was evaluated both with and without outlier removal. Accuracy was used to assess how the model was affected by the outlier removal process and data standardization at the onset of analysis. To optimize the performance and stability of the model, a wide range of gridsearch CV 5-fold hyperparameter tuning techniques were employed.

5.1 Proposed model performance with outliers in the dataset

A classification report assesses the performance of a model by detailing precision, recall, F1-score, and support for each dataset class [36]. Precision gauges correct positive class identification, while recall measures positive class detection. The F1-score balances precision and recall, crucial when false positives/negatives matter. Support quantifies class observations, aiding the identification of imbalanced classes. The report features a table of these metrics per class, plus an overall model summary with macro, micro, and weighted averages. The macro-average treats all classes equally. The micro-average considers all observations equally. The weighted average prioritizes larger classes. In the dataset, without considering the outliers the model produced a classification report shown in Table 5.

Table 5. Classification report with outlier's model performance.

	Precision	Recall	F1-Score	Support
Class 0	0.90	0.84	0.87	88
Class 1	0.86	0.92	0.89	96
Accuracy	-	-	0.88	184
Macro Avg	0.88	0.88	0.88	184
Weighted Avg	0.88	0.88	0.88	184

In this evaluation, we evaluate the performance of a model using a binary dataset comprising two categories: Class 0 (representing individuals without heart disease) and Class 1 (representing individuals with heart disease). The assessment involves various metrics, including precision, where we observe a precision of 90% for Class 0 and 86% for Class 1. Additionally, we calculate recall, yielding values of 84% for Class 0 and 92% for Class 1. The F1-score, a combined measure of precision and recall, is computed at 87% for Class 0 and 89% for Class 1. The support values, representing the number of instances in each class, are 88 for Class 0 and 96 for Class 1. Overall, the model achieves an accuracy rate of 88%. The report encompasses macro-average, micro-average, and weighted-average precision, recall, and F1-score, all of which are documented as 88%. A confusion matrix is also included, depicted in Figure 10.

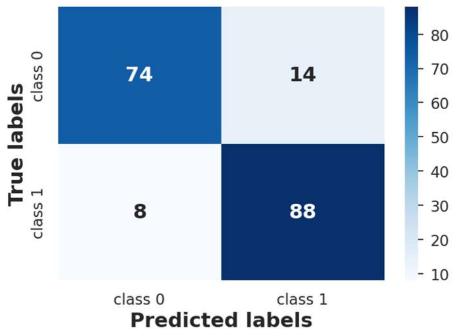

Figure 10. Model confusion matrix values with outliers

5.2 Proposed model performance without outliers in the dataset

Table 6 summarizes the model performance in a two-class dataset, including precision, recall, F1-score, and support for Class 0 and Class 1. Class 0 achieved 93% precision, 92% recall, and a 93% F1-score with 76 samples, while Class 1 had 91% precision, 93% recall, and a 92% F1-score with 67 samples.

Table 6. Classification report without outlier's model performance.

	Precision	Recall	F1-Score	Support
Class 0	0.93	0.92	0.93	76
Class 1	0.91	0.93	0.92	67
Accuracy	-	-	0.92	143
Macro Avg	0.92	0.92	0.92	143
Weighted Avg	0.92	0.92	0.92	143

The overall accuracy is 92%, matching macro-average and weighted-average scores. Figure 11 presents the outlier-detected confusion matrix, while Figure 12 displays the ROC curve for the hybrid RNN-GRU model. In the implementation, data is prepossessed for GRU and RNN models, using the Keras framework to create the neural network with various configurations explored through GridSearchCV for optimization. Figures 13 and 14 show epoch-wise accuracy and loss for the proposed hybrid DL model. Figure 15 shows the accuracy of training and validation.

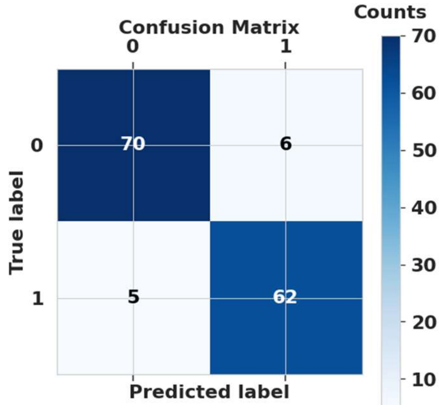

Figure 11. Model confusion matrix values without outliers

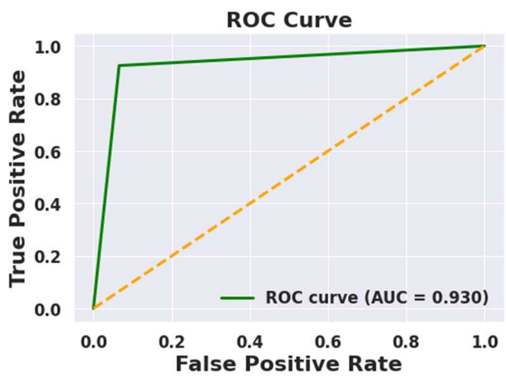

Figure 12. Model ROC plot without outliers in the dataset

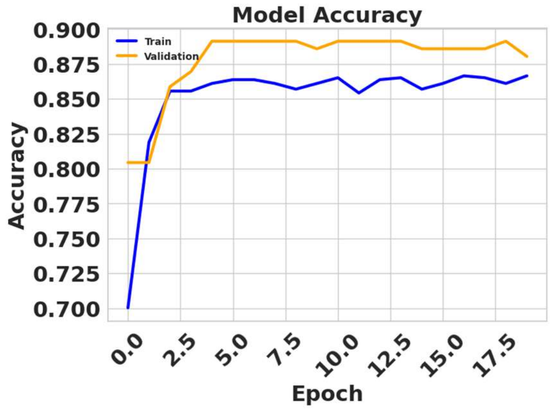

Figure 13. Model accuracy for each epoch

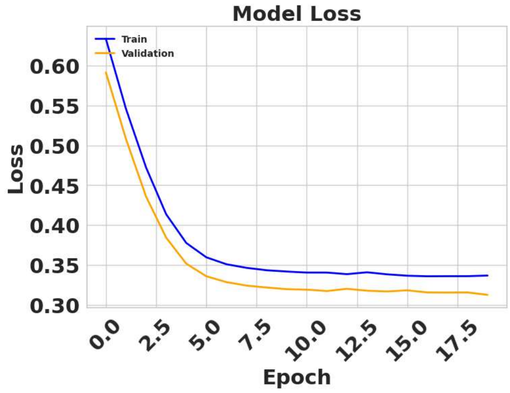

Figure 14. Model loss for each epoch

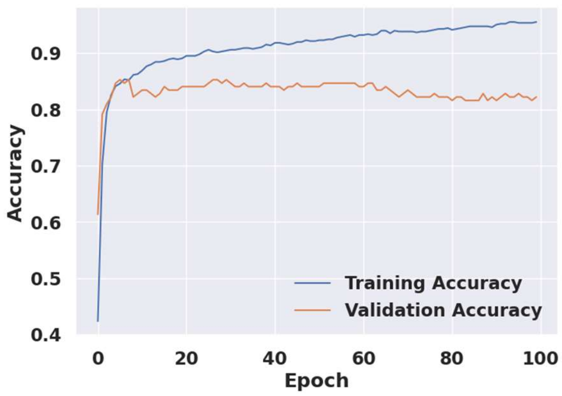

Figure 15. Model training and validation accuracy

5.3 *Comparison of model performance metrics with and without outlier detection*

Figure 16 shows the comparison of the performance of an ML model trained on the IEEE dataset with and without outlier detection. We analysed the model's effectiveness using standard evaluation parameters such as precision (correct positive predictions), recall (ability to detect true positives), F1-score (balance of precision and recall), support (sample counts), and overall accuracy. Class-specific results (for healthy and diseased cases) were summarized using macro-average (equal class weighting) and weighted-average (class-size-adjusted) metrics. The results show that incorporating outlier detection significantly enhances the model's performance. For instance, precision for class 0 increases from 90% to 93%, and class 1, it rises from 86% to 91% with outlier detection.

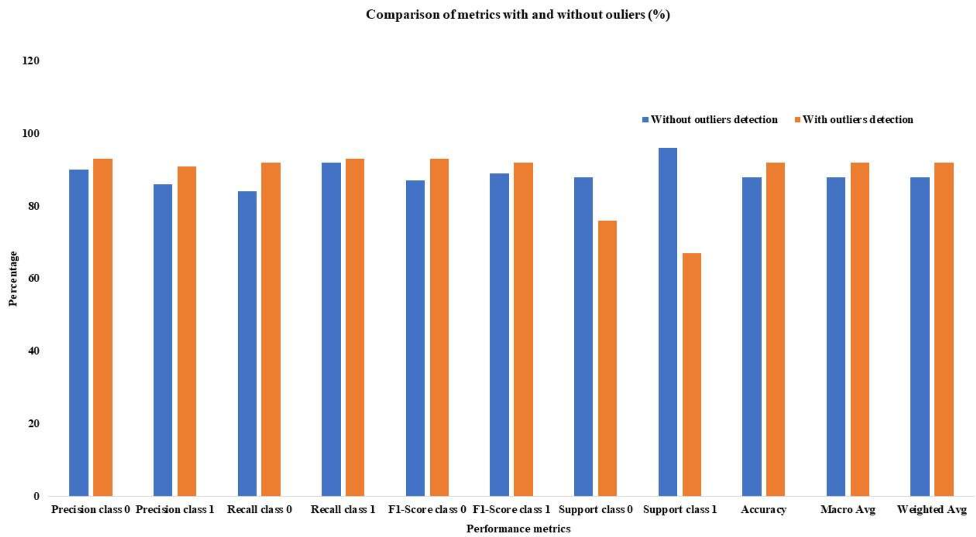

Figure 16. Model performance metric comparison with and without outliers in the dataset

Similarly, recall for class 0 improves from 84% to 92%, and for class 1, it increases from 92% to 93%. F1 scores also show improvements for both classes, with class 0 going from 87% to 93%, and class 1 from 89% to 92%. Support values decrease for both classes, indicating fewer instances classified into each class with outlier detection. Additionally, accuracy improves from 88% to 92%, and macro-average and weighted-average metrics demonstrate enhancements, highlighting overall model improvement without bias toward one class. These findings underscore the significant enhancement in the efficacy of DL models, especially in terms of precision, recall, F1-score, and accuracy, when the utilization of outlier detection techniques is incorporated. For further proposed model accuracy comparisons with existing literature, refer to Table 7.

Table 7. Comparison of the proposed system with existing heart disease prediction systems.

S. No	Author(s) and Year	Methodology/Algorithm	Data Set	Key Findings/Results
1	Mert Özcan et al.(2023) [37]	Classification and Regression Tree Algorithm (CART)	IEEE Dataport Heart Disease Dataset (Comprehensive)	Achieved 87% accuracy in heart disease prediction using CART.
2	Justin et al. (2022) [38]	SVM	IEEE Dataport Heart Disease Dataset (Comprehensive)	Improved prediction accuracy to 90.12% by using recursive feature elimination with ML models.
3	Mohammad et al. (2022) [39]	DT, SVM, LR	IEEE Dataport Heart Disease Dataset (Comprehensive)	Improved prediction accuracy to 87.57% by using DT, SVM, LR.
4	Sudipta Modak et al. (2022) [40]	MLP	IEEE Dataport Heart Disease Dataset (Comprehensive)	Improved prediction accuracy to 87.70% by using a modified variation of infinite feature selection and multilayer perceptron.

5	Proposed	RNN-GRU	IEEE Dataport Heart Disease Dataset (Comprehensive)	Achieved 92% accuracy in heart disease prediction using RNN-GRU.

6 Conclusion

The research work has successfully developed a hybrid deep learning model for the prediction of heart disease. This model combines RNN with GRU architecture, utilizing the Adam optimizer. Impressively, the model exhibited strong performance when evaluated on the IEEE Dataport dataset, achieving an accuracy rate of 92%. Notably, the research underscores the critical role of data pre-processing, specifically in outlier detection, to enhance the model's accuracy. Through comparative analysis, it was evident that the incorporation of outlier detection led to a notable improvement in model performance, resulting in a boost in accuracy 3% accuracy. This study has a constraint in terms of dataset size, which is relatively small. Subsequent research efforts will prioritize the use of more extensive datasets. These findings underscore the significance of employing effective data pre-processing techniques to optimize DL models for heart disease prediction, ultimately leading to more precise and dependable diagnostic tools in cardiovascular health. The developed model can be very useful in hospitals and clinics. It acts like a smart tool to help doctors spot heart problems sooner, so patients can get treatment faster. It could also work with home health devices (like wearables) to check heart health daily and send alerts if risks are found. Since our model is 92% accurate and works quickly, it could be used for free or low-cost heart check-ups in communities. This would help find people at high risk early, saving hospitals time and money.

Declaration

Conflict of Interest: The authors declare that there are no conflicts of interest.
Funding Information: No financial assistance or grants were provided for this research.
Author contribution: All authors were involved in the conceptualization and design of the study.

References

[1] "World health statistics. Cardiovascular Diseases, Key Facts. 2021." https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed Dec. 10, 2022).

[2] L. Yahaya, N. David Oye, and E. Joshua Garba, "A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques," *Am. J. Artif. Intell.*, vol. 4, no. 1, p. 20, 2020, doi: 10.11648/j.ajai.20200401.12.

[3] A. K. Shanmugaraj, A. Nagaradjane, A. Candassamy, and K. Dhamodharan, "Smart Heart Disease Prediction and Amalgamation Tracking System," *Eng. Rev.*, vol. 44, no. 1, pp. 14–24, 2024, doi: 10.30765/er.2110.

[4] B. Sai Bharadwaj and C. Sumanth Kumar, "A Novel Digital Phonocardiography Method to Identify the Cardiac Sounds through Intrinsic Time Scale Decomposition and Inter time space Measurement Between Cardiac Sounds," *SSRG Int. J. Electr. Electron. Eng.*, vol. 9, no. 12, pp. 22–29, 2022, doi: 10.14445/23488379/IJEEE-V9I12P102.

[5] N. Chandrasekhar, M. Krishnaiah, S. Peddakrishna, S. P. Kancharla, N. Gummadavelly, and U. Nuvvula, "Heartfelt AI: Analyzing Emotions Through ECG Signals for Affective Computing," in *Intelligent Computing and Optimization*, 2024, pp. 120–129.

[6] N. Chandrasekhar and S. Peddakrishna, "Unveiling the Contributing Features in Heart Disease Occurrence: An Explainable AI Approach," in *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, 2023, pp. 1–6, doi: 10.1109/ASIANCON58793.2023.10270542.

[7] A. Alanazi, "Using machine learning for healthcare challenges and opportunities," *Informatics Med. Unlocked*, vol. 30, no. March, p. 100924, 2022, doi: 10.1016/j.imu.2022.100924.

[8] N. Chandrasekhar, "Enhancing Heart Disease Prediction Accuracy through Machine Learning

Techniques and Optimization," *Processes*, vol. 11, no. 4, p. 1210, 2023, doi: https://doi.org/10.3390/pr11041210.

[9] A. De Santo, A. Galli, M. Gravina, V. Moscato, and G. Sperli, "Deep Learning for HDD Health Assessment: An Application Based on LSTM," *IEEE Trans. Comput.*, vol. 71, no. 1, pp. 69–80, Jan. 2022, doi: 10.1109/TC.2020.3042053.

[10] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Brief. Bioinform.*, vol. 19, no. 6, 2017, doi: 10.1093/bib/bbx044.

[11] S. Kollem, P. Krishna, S. Peddakrishna, and B. D. Rao, "Innovative Approach: CNNs for Heart Disease Prediction," in *2024 IEEE Region 10 Symposium (TENSYMP)*, 2024, pp. 1–6, doi: 10.1109/TENSYMP61132.2024.10752241.

[12] H. Wang, Q. Zu, J. Chen, Z. Yang, and M. A. Ahmed, "Application of Artificial Intelligence in Acute Coronary Syndrome: A Brief Literature Review," *Advances in Therapy*, vol. 38, no. 10. 2021, doi: 10.1007/s12325-021-01908-2.

[13] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 5, 2018, doi: 10.1109/JBHI.2017.2767063.

[14] X. Bai *et al.*, "Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments," *Pattern Recognit.*, vol. 120, p. 108102, Dec. 2021, doi: 10.1016/J.PATCOG.2021.108102.

[15] S. Pouyanfar *et al.*, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys*, vol. 51, no. 5. 2018, doi: 10.1145/3234150.

[16] N. Chandrasekhar, S. Canavoy Narahari, S. Kollem, S. Peddakrishna, A. Penchala, and B. Prasad Chapa, "Heart abnormality classification using ECG and PCG recordings with novel PJM-DJRNN," *Results Eng.*, vol. 25, p. 104032, 2025, doi: https://doi.org/10.1016/j.rineng.2025.104032.

[17] S. Tuli *et al.*, "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments," *Futur. Gener. Comput. Syst.*, vol. 104, pp. 187–200, Mar. 2020, doi: 10.1016/j.future.2019.10.043.

[18] G. Surange and P. Khatri, "Integrated intelligent IOT forensic framework for data acquisition through open-source tools," *Int. J. Inf. Technol.*, vol. 14, no. 6, pp. 3011–3018, 2022, doi: 10.1007/s41870-022-01025-5.

[19] K. B. Raju, S. Dara, A. Vidyarthi, V. M. Gupta, and B. Khan, "Smart Heart Disease Prediction System with IoT and Fog Computing Sectors Enabled by Cascaded Deep Learning Model," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/1070697.

[20] S. Krishnan, P. Magalingam, and R. B. Ibrahim, "Advanced recurrent neural network with tensorflow for heart disease prediction," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 5 Special Issue, pp. 966–977, 2020.

[21] G. S. Bhavekar and A. Das Goswami, "A hybrid model for heart disease prediction using recurrent neural network and long short term memory," *Int. J. Inf. Technol.*, vol. 14, no. 4, pp. 1781–1789, Jun. 2022, doi: 10.1007/s41870-022-00896-y.

[22] B. Mali, S. Saha, D. Brahma, R. Pinninti, and P. K. Singh, "Towards Building a Global Robust Model for Heart Disease Detection," *SN Comput. Sci.*, vol. 4, no. 5, p. 596, 2023, doi: 10.1007/s42979-023-02083-7.

[23] N. K. Sharma, A. S. Chauhan, S. Fatima, and S. Saxena, "Enhancing heart disease diagnosis: Leveraging classification and ensemble machine learning techniques in healthcare decision-making," *J. Integr. Sci. Technol.*, vol. 13, no. 1, pp. 1–8, 2025, doi: 10.62110/sciencein.jist.2025.v13.1015.

[24] M. Ahmed and I. Husien, "Hybrid Machine Learning Approach for Accurate Heart Disease Prediction," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 4, pp. 728–737, 2024, doi: 10.22266/IJIES2024.0831.55.

[25] S. Mondal, R. Maity, Y. Omo, S. Ghosh, and A. Nag, "An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches," *IEEE Access*, vol. 12, no. January, pp. 7255–7270, 2024, doi: 10.1109/ACCESS.2024.3350996.

[26] S. M. Alanazi and G. S. M. Khamis, "Optimizing Machine Learning Classifiers for Enhanced Cardiovascular Disease Prediction," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 1, pp. 12911–12917, 2024, doi: 10.48084/etasr.6684.

[27] Navita *et al.*, "Advanced Hybrid Machine Learning Model for Accurate Detection of Cardiovascular Disease," *Int. J. Comput. Intell. Syst.*, vol. 18, no. 1, pp. 1–20, 2025, doi: 10.1007/s44196-025-00771-1.

[28] B. P. Doppala, D. Bhattacharyya, M. Janarthanan, and N. Baik, "A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques," *J. Healthc. Eng.*, vol. 2022, no. 1, p. 2585235, 2022, doi: https://doi.org/10.1155/2022/2585235.

[29] M. Ferdowsi, M. M. Hasan, and W. Habib, "Responsible AI for cardiovascular disease detection: Towards a privacy-preserving and interpretable model," *Comput. Methods Programs Biomed.*, vol. 254, p. 108289, 2024, doi: https://doi.org/10.1016/j.cmpb.2024.108289.

[30] IEEE dataport, "Heart Disease Dataset," *IEEE datport*, 2020. https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive (accessed Nov. 12, 2022).

[31] A. Tiwari, A. Chugh, and A. Sharma, "Ensemble framework for cardiovascular disease prediction," *Comput. Biol. Med.*, vol. 146, no. May, p. 105624, 2022, doi: 10.1016/j.compbiomed.2022.105624.

[32] A. Rapaka and A. C. Kanmani, "An Optimized Hyper Parameter Tuned Convolution Neural Frame for Potato Leaves Disease Prediction," vol. 10, no. 2, pp. 180–195, 2023.

[33] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," pp. 1–9, 2014, [Online]. Available: http://arxiv.org/abs/1412.3555.

[34] D. M. Kotambkar and P. M. Wankhede, "Hybrid LSTM / GRU-based Domain Adaptation Model for Correlation Analysis to Detect Glaucoma," vol. 10, no. 1, pp. 168–175, 2023.

[35] S. Bodiwala and N. Nanavati, "An efficient stochastic computing based deep neural network accelerator with optimized activation functions," *Int. J. Inf. Technol.*, vol. 13, no. 3, pp. 1179–1192, 2021, doi: 10.1007/s41870-021-00682-2.

[36] R. Ahuja and S. C. Sharma, "Stacking and voting ensemble methods fusion to evaluate instructor performance in higher education," *Int. J. Inf. Technol.*, vol. 13, no. 5, pp. 1721–1731, 2021, doi: 10.1007/s41870-021-00729-4.

[37] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthc. Anal.*, vol. 3, no. December 2022, p. 100130, 2023, doi: 10.1016/j.health.2022.100130.

[38] J. A. P. Denopol, M. S. A. Magboo, and V. P. C. Magboo, "Heart Alert: A heart disease prediction system using machine learning approach and optimization techniques," *Philipp. J. Heal. Res. Dev.*, vol. 26, no. 3, pp. 83–92, 2022.

[39] M. Wazid, J. Singh, A. K. Das, S. Shetty, M. K. Khan, and J. J. P. C. Rodrigues, "ASCP-IoMT: AI-Enabled Lightweight Secure Communication Protocol for Internet of Medical Things," *IEEE Access*, vol. 10, pp. 57990–58004, 2022, doi: 10.1109/ACCESS.2022.3179418.

[40] S. Modak, E. Abdel-Raheem, and L. Rueda, "Heart Disease Prediction Using Adaptive Infinite Feature Selection and Deep Neural Networks," in *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2022, pp. 235–240, doi: 10.1109/ICAIIC54071.2022.9722652.