The Scientific Temper (2025) Vol. 16 (6): 4487-4498
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2025.16.6.22
https://scientifictemper.com/

RESEARCH ARTICLE

Early diagnosis of cardiac disease using Xgboost ensemble voting-based feature selection, based lightweight recurrent neural network approach

S. Tamil Fathima[1*], K. Fathima Bibi[2]

Abstract

Cardiovascular disease (CVD) causes the heart and blood vessels to fail, often resulting in death or stroke. Therefore, early automatic identification of CVD can rescue many lives. CVD identification and prognosis are essential clinical tasks to ensure precise classification results, which assist cardiologists in providing suitable patient treatment. The use of deep learning (DL) in the medical field is increasing as it can determine patterns in data. Despite that, CVD prediction is a profound challenge in clinical data analysis. Conventional methods cannot handle hidden patterns, leading to less accurate model predictions. There is a critical need for a new technique that can rapidly and reliably predict future outcomes in patients with CVD. To combat this issue, this research uses a benchmark dataset to present a lightweight recurrent neural network with a long short-term memory (LRNN-LSTM) method for CVD. Initially, the min-max batch normalization (M2BN) method is used to verify the ideal margin of collected data values in the dataset. Secondly, they employed the decision tree (DT) technique to select the best gain attribute for predicting CVD. Furthermore, the XGBoost ensemble voting-based feature selection (XGB-EVFS) method determines the profound features of CVD. Then, our proposed LRNN-LSTM algorithm is used to categorize the CVD result to reduce misdiagnosis. The proposed system will develop a model that can accurately predict CVD to decrease mortality from cardiac disease. Therefore, the experiment analysis produces high classification accuracy, precision, and recall with fewer false scores than traditional methods.

Keywords: Cardiovascular disease, Deep Learning, LRNN-LSTM, Decision tree, XGBoost ensemble, Voting-based feature selection.

Introduction

The essentials for medical diagnosis of CVD have increased recently. The World Health Organization (WHO) said that in 2020, CVD accounted for the majority of deaths globally, with an increase in the death rate of 16%. Furthermore, CVD has increased the need more than ever for an accurate and fast diagnosis. Similarly, there are two types of causes of heart failure: those that have to do with the anatomy of the heart (like a heart attack) and those that have to do with how the heart works (like high blood pressure). Medications, lifestyle changes, and occasionally surgery have traditionally formed the core of heart failure treatment. Moreover, early identification of heart failure has been demonstrated to enhance quality of life and extend longevity (Iqbal, T., Soliman, O., Sultan, S., and Ullah, I., 2023, Qadri, A. M., Raza, A., Munir, K., and Almutairi, M. S.,2023).

Thus, their accuracy is crucial for predicting CVD when lives are at stake. Additionally, their data evaluation is based on data mining (DM) technology, which helps reduce clinical errors in diagnosis. Similarly, experimenters can present several DM techniques to improve accuracy based on the health sector. Distribute different departments that best use DM, including commercial banking and marketing. As far as the health sector is concerned, the more accurate prediction of diseases such as heart disease, hepatitis, etc., can be of greater advantage (Bashir, S., Almazroi, A. A., Ashfaq, S., Almazroi, A. A., and Khan, F. H. 2021).

[1]Research Scholar in Computer Science, Thanthai Periyar Govt. Arts & Science College (A), Tiruchirappalli, Affiliated to Bharathidasan University.

[2]Assistant Professor in Computer Science, Thanthai Periyar Govt. Arts & Science College (A), Tiruchirappalli, Affiliated to Bharathidasan University

***Corresponding Author:** S. Tamil Fathima, Research Scholar in Computer Science, Thanthai Periyar Govt. Arts & Science College (A), Tiruchirappalli, Affiliated to Bharathidasan University., E-Mail: tamilfathima@jmc.edu

How to cite this article: Fathima, S.T., Bibi, K.F. (2025). Early diagnosis of cardiac disease using Xgboost ensemble voting-based feature selection, based lightweight recurrent neural network approach. The Scientific Temper, **16**(6):4487-4498.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.6.22

Source of support: Nil

Conflict of interest: None.

Received: 24/06/2025 **Accepted:** 27/06/2025 **Published** : 30/06/2025

Nevertheless, CVD is a collection of disorders that impact the heart and blood arteries. These disorders can result in symptoms such as heart failure and coronary heart disease, in addition to chest pain, discomfort, dyspnea, and irregular pulse. Some cardiac conditions, such as aneurysms and strokes, can result in vision loss, excruciating headaches, and loss of eyesight. At times, the symptoms of various cardiac disorders overlap, making it difficult to distinguish between them based solely on symptoms (Ghorashi, S., *et al*., 2023).

The WHO reports that heart-related problems and HD account for almost 17.9 million deaths annually. However, some risk factors that promote heart-related problems include unhealthy diet, lack of exercise, alcohol consumption, and smoking. Nevertheless, with the increasing HD population and number of patients, providing affordable diagnoses to individuals in top medical facilities is becoming more challenging (Ahsan, M.M., and Siddique, Z.,2022).

This section presents the LRNN-LSTM method and its application on a benchmark dataset for predicting disease in patients with CVD. The M2BN method is initially employed to identify the best contour for the dataset's collected values. Similarly, DT technology can predict CVD by selecting optimal gain characteristics. Moreover, the most essential features of the CVD are chosen using the XGB-EVFS method. Then, the proposed LRNN-LSTM algorithm is used to classify CVD results to reduce misdiagnosis. The system aims to develop models that can accurately predict CVD and reduce cardiovascular mortality.

An architecture diagram for CVD illustrates how data from standard datasets can be collected and inputs generated to predict CVD patient data. Additionally, the diagram allows us to determine the margin value, and identify the best attribute, and select and classify the features shown in Figure 1.

Literature Survey

(Kartik Budholiya, Shailendra Kumar Shrivastava, Vivek Sharma, 2022) discussed the proposed diverse diagnostic systems for predicting heart disease (HD) based on bizarre techniques in the existing problem. Therefore, an optimized XGBoost technique can enhance the diagnostic process and

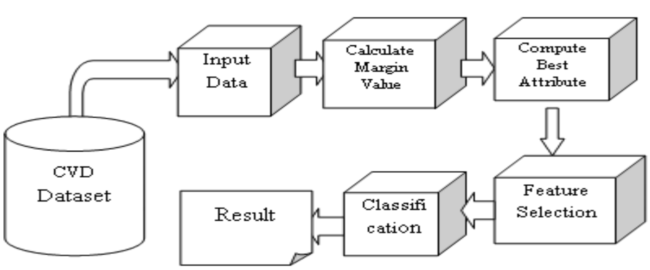

Figure 1: The architecture diagram for CVD

assist cardiologists in these areas. However, they faced a prime summons in providing quality services at inexpensive costs.

Accordingly, chronic disease management is described as one of the most urgent, quickly evolving, and intricate issues healthcare providers and governments worldwide face. A comprehensive tree-boosting system known as XGBoost will also be a key component of a new integrated predictive modeling framework.

They discussed how heart centers and hospitals rely heavily on electrocardiograms to diagnose heart failure. Primitive spotting of HD is a critical issue for medical services. Therefore, the accuracy of HD diagnosis can be improved based on the XGBoost test, which analyzes different ML protocols for HD diagnosis and replaces decision tree (DT) classification methods.

Ali L. *et al*. (2019) developed an expert system that can overestimate support vector machine (SVM) models to create effective heart failure prediction. Diagnosing HF effectively is challenging due to insufficient diagnostic equipment and clinical expertise.

Similarly, they discussed the challenge of predicting important global causes of death amidst clinical data inquiry. ML classifiers are used to discover critical characteristics that improve the accuracy of HD prediction. Among the methods developed for predicting cardiovascular problems, learned vector measurement can save lives, improve health conditions, and efficiently allocate clinical assets.

Furthermore, it is difficult for cardiologists and surgeons to determine when heart failure occurs. To achieve this goal, an ML model will be deployed to augment prognostic precision for diseases such as heart failure (HF). Afterward, the RF and gradient boosting (GB) methods were examined to yield the most reliable results in predicting future heart attacks.

They described accurate prediction and timely identification of cardiac disease as critical to improving patient survival. Nevertheless, this was a significant drawback for hospitals and clinics, as no medical diagnosis could predict the disease. AI has become a valuable tool for predicting and understanding heart disease symptoms. Further, chi-square statistical tests can be applied to manipulate HD Cleveland datasets to identify specific features.

Moreover, many investigators have developed expert systems to predict HD and help improve cardiologists' diagnostic progress. Nevertheless, most machine learning (ML) techniques are complex and often rely on big data developed for an application. To overcome this issue, a simple and effective detection system is proposed using extreme gradient boosting and feature selection algorithms.

Ganie *et al*. (2023) described the novel early detection as critical to preventing disease progression. An improved ML

Table 1: Prediction of cardiovascular disease using machine learning

Author	*Year*	*Techniques*	*Methods*	*Limitations*
Alqahtani, *et al*., 2022	2022	ML and DL	RF	One of the uttermost difficult chores for doctors is to diagnose the indication of HD as early as possible.
Ambrish, *et al*., 2022	2022	DL	Logistic regression (LR)	A major challenge for the medical field is to use cheap and reliable methods to predict CVD to avoid confounding outcomes.
Ayatollahi, H., Gholamhosseini, L., & Salehi, M., 2019	2019	ML	SVM	The main challenges are a substantiated amplification in CVDs and the ponderous fund load they impose on society.
Guarneros-Nolasco, L. R., Cruz-Ramos, N. A., Alor-Hernandez, G., Rodriguez-Mazahua, L.,and Sanchez-Cervantes, J. L., 2021	2021	ML	K-FCV	The predictions could be more precise because they cannot account for differences in the data.
Hosni *et al*., 2021	2021	ML	Artificial Neural Networks (ANN)	However, predicting diseases takes time.
Hossain, *et al*., 2023	2023	Artificial Intelligence (AI)	SVM	The lack of additional discriminative feature sets and additional data sets significantly reduces the performance of Naive Bayes models.
Khandaker *et al*., 2023	2023	ML	Ada-Boost Classifier (AB), DT	Clinical perception is a complex process that is indispensable for precision and competence.
Osamah Sami, Yousef Elsheikh and Fadi Almasalha, 2021	2021	ML	RF	It has poor accuracy classifiers, and the accomplishment could be better for determining the peril of coronary HD.
Rahim, *et al*., 2021	2021	ML	K-Nearest Neighbor (K-NN)	A wrong prediction can be life-threatening and fatal.
Sai *et al*., 2021	2021	ML	DT and RF	Making a correct decision takes a long time, so it causes death.

technique was deployed to obtain the goal. Furthermore, the individual distributions are then modeled using classification and regression trees. The test results on the datasets reach assortment accuracy, which is better than the cross-layer leap gated convolution neural network (CLLG-CNN) method.

Recent studies have demonstrated the efficacy of applying diverse machine learning algorithms for medical diagnosis tasks. For instance, a comprehensive performance comparison of classifiers—decision tree, random forest, support vector machine, K-nearest neighbor, and logistic regression—was conducted in the context of diabetes prediction, highlighting the critical role of algorithm selection in enhancing diagnostic accuracy (S. Peerbasha *et al*., 2023). The manual methods used to diagnose HD are biased and do not consider inter-exam variations. Furthermore, various ML algorithms have been used to diagnose and predict human HD based on the standard K-fold cross-validation (K-FCV) technique to instruct and adjust protocols.

Based on a survey of ML methods for predicting cardiovascular heart disease, the author discusses their strategies, techniques, and limitations, as outlined in Table 1.

The novel discussed the least impact disease prone rate-densenet multi perceptron neural network (LIDPR-DNMPNN) method, in which a vast amount is collected from the healthcare industry to healthcare data, Maheswari Subburaj *et al*. (2019). Still, unfortunately, the hidden information cannot be identified for effective decision-making. An end-user upholds a system as an HD calculation application to provide practical guidance on HD to users through an intelligent prediction system window.

Similarly, addressing cardiovascular risks requires primary verdict-making and congenial treatment. The deployed CVD prediction system (CDPS) performance is estimated using different measures to determine the optimal ML technique.

Therefore, rapid and accurate patient diagnosis is one of medicine's most toilsome drawbacks. Since the objective was to predict HD using assortment techniques, a DM technique, DT, was used in disease diagnosis. Moreover, DT can reasonably accurately predict the risk of HD in people with diabetes.

After that, we discussed applying the RF method for predicting HD. They utilized an ML method and compared its performance with KNN to achieve higher prediction accuracy. Additionally, they highlighted that protocols such as RF use comparison factors to outperform KNN.

They discussed that early assessment methods for CVD uncertainty could guide risk-reducing decisions for high-risk patients. The accuracy of different algorithms can be evaluated based on the values of RF, SVM, KNN, and DT to select the most preferred prediction methods.

Proposed Methodology

This section uses a Cleveland dataset to analyze the LRNN-LSTM method and its significance in predicting cardiovascular disease (CVD) patients. The M2BN method plays a crucial role in determining the optimal marginal values collected in the dataset, ensuring accuracy and precision in the analysis. They utilized the decision tree (DT) method to determine the optimal gain attribute for predicting CVD. Similarly, the XGB-EVFS method can select the optimal features of CVD. The proposed LRNN-LSTM algorithm is then implemented to classify heart disease, which can reduce misdiagnosis, and the accuracy of this improves overall patient data.

Based on the proposed LRNN-LSTM method, it is utilized to reduce misdiagnosis and classify CVD consequences. Furthermore, the method proposed can create a model that accurately predicts CVD and reduces mortality from heart disease, as shown in the architecture diagram in Figure 2. The M2BN method was used to determine the optimal margin for data collection. This measure ensures that the dataset contains an adequate number of data points. DT techniques are used to construct a decision tree for CVD prediction. The algorithm selects the most informative features based on the information gained, creating a hierarchical structure that classifies the data points into CVD-positive or CVD-negative

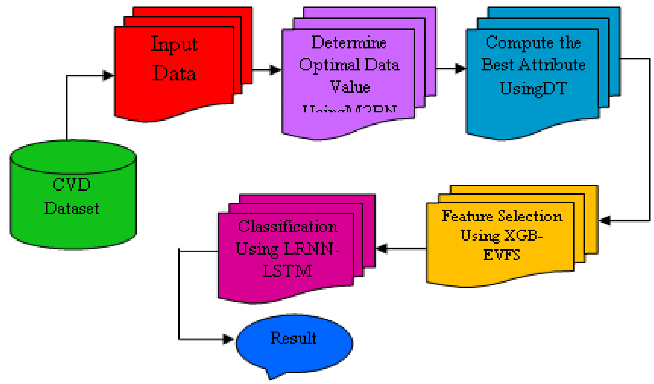

Figure 2: The proposed architecture diagram for LRNN-LSTM

categories. The XGB-EVFS method is used to identify the most important features related to CVD. Furthermore, the LRNN-LSTM algorithm can classify CVD outcomes and reduce misdiagnosis. The LRNN-LSTM architecture enables the model to learn complex patterns and relationships in the data, thereby improving the accuracy of CVD predictions.

Dataset Collection

The UCI repository is one of the most comprehensive datasets, with over 417 different datasets. Furthermore, the CVD feature dataset within the UCI machine learning repository has emerged as a key resource in heart disease research. This dataset, containing data from 1025 patients, comprises 76 features, with 14 attributes, including crucial class labels, being actively leveraged in analysis and modeling efforts.

Subsequently, the dataset is partitioned into a training set, comprising 75% of the samples, and a testing set,

Table 2: CVD feature dataset

id	*age*	*sex*	*dataset*	*cp*	*trestbps*	*chol*	*fbs*	*restecg*	*thalach*	*exang*	*oldpeak*	*slope*	*ca*	*thal*	*num*
1	63	Male	Cleveland	0	125	212	0	1	168	0	1.0	2	2	3	0
2	67	Male	Cleveland	0	140	203	1	0	155	1	3.1	0	3	3	2
3	67	Male	Cleveland	0	145	174	0	1	125	1	2.6	0	3	3	1
4	37	Male	Cleveland	0	148	203	0	1	161	0	0.0	2	2	3	0
5	41	Female	Cleveland	0	138	294	0	1	106	0	1.9	1	3	3	1
6	56	Male	Cleveland	0	100	248	0	0	122	1	1.0	1	0	3	2
7	62	Female	Cleveland	0	110	318	0	0	145	0	4.4	0	2	3	2
8	57	Female	Cleveland	0	160	289	0	1	140	1	0.8	2	0	3	3
9	63	Male	Cleveland	0	120	249	0	0	144	0	0.8	2	0	3	2
10	53	Male	Cleveland	0	122	286	0	0	116	1	3.2	1	2	2	1

comprising 25% of the samples. Of the total 1,025 patient records, 820 were utilized for training, leaving 205 samples for testing. As demonstrated in Table 2, the trained model's performance is assessed in a test phase using the validation data.

Min-max batch normalization (M2BN)

In this section, the M2BN method is used to identify the marginal data values in a dataset. After that, the original data set can be normalized by performing a linear transformation based on the M2BN method. Further, each feature weight can be transformed to a minimum feature value of 0, a maximum value of 1, and all other values to decimal numbers between 0 and 1. Therefore, the correlation between input and measured values can be calculated using this M2BN method. Furthermore, implementing the M2BN technique and normalizing each input layer is used to estimate the min-max normalization between the input and measured values. Moreover, the optimal marginal value of block normalization can be calculated by using the standard normal distribution for the input parameters provided in this.

As described in Equation 1, the feature values can be normalized, and the uniform scale can be calculated using the standard scale. Furthermore, it can be indicated using values of large dimensions such as fat, glucose, and feature weight from available datasets. Let's assume p- p-participation value, p'-standard value, μ-mean, and σ-standard deviation, M-total number of the column,

$$\begin{cases} p' = \dfrac{p-\mu}{\sigma} \\ \mu = \dfrac{\sum_{o=1}^{M} p}{M} \\ \sigma = \sqrt{\dfrac{1}{M}\sum_{o=1}^{M}\left(p^{o}-\mu\right)^{2}} \end{cases} \quad (1)$$

A min-max scale can be used for normalization to measure large eigenvalues. Furthermore, eigenvalues can be normalized to infinite intervals between minimum and maximum values. Calculate the max-min value as shown in Equation 2. Let's assume $p_{\min}$ – minimum value, $\mathcal{P}_{\max}$ – maximum value.

$$\mathrm{P}' = \frac{\mathrm{p}-\mathrm{p}_{\min}}{\mathcal{P}_{\max}-\mathrm{p}_{\min}} \quad (2)$$

Moreover, equations 3 are used to estimate the min-max normalization between input and measured values. It provides a linear transformation for the Min-Max normalization of the raw original dataset. Let's assume $theP_{o}^{m}$ – original value, $P'(o,m)$ – normalized value, and $\max^{p}(\mathcal{N}_{ew}) - \mathrm{Min}^{p}(\mathcal{N}_{ew})$ – denote as the rescaled value of the data.

$$\mathrm{p}'(0,\ \mathrm{m}) = \frac{\mathrm{P}_{o}^{\mathrm{m}} - \min\left(p_{o}^{m}\right)}{\max\left(\mathrm{P}_{o}^{\mathrm{m}}\right) - \min\left(p_{o}^{m}\right)} \left(\overset{\mathrm{p}}{\max}(\mathcal{N}_{ew}) - \mathrm{Min}^{\mathrm{p}}\left(\mathcal{N}_{ew}\right)\right) + \mathrm{m}_{in}\left(\mathrm{P}^{\mathcal{N}_{ew}}\right) \quad (3)$$

Z-score normalization involves rescaling and adjusting the variance and mean to normalize each input value of the feature vectors and the standard deviation. The normalized value is calculated as the square root of the variance, as illustrated in Equations 4 and 5. Where $P'_{\mathrm{o,them}}$ -new value of the feature, the P_{o}^{m} -old value of the feature, the μ_{i} -mean and standard deviation of the feature value,

$$\mathrm{P}'_{\mathrm{o,\ m}} = \frac{\mathrm{P}_{\mathrm{o}}^{\mathrm{m}} - \mu_{\mathrm{o}}}{\sigma_{\mathrm{o}}} \quad (4)$$

$$\mathrm{P}'_{\mathrm{o,\ m}} = \frac{1}{\mathrm{m}}\left[\mathrm{P}_{\mathrm{o}}(1) + \mathrm{p}_{\mathrm{o}}(2) + \mathcal{P}_{\mathrm{o}}(3) + \cdots + \sigma_{\sigma} \ldots \mathcal{P}_{\mathrm{o}}^{\mathrm{m}}\right](5)$$

Equations 6 and 7 describe the mean deviation of a small data set from the normal marginal value and calculate the optimal marginal value of batch normalization. Let's assume V-batch data, n-number of the input value, μ_{v} – mean value, σ_{v}^{2} – sample variance, γ,β – parameter.

$$\mathrm{V}_{\mathrm{M}(\tilde{a},\hat{a})} : \mathrm{P}_{1\ldots n} \rightarrow \mathrm{q}_{1\ldots n} \quad (6)$$

$$\begin{cases} \hat{\mathrm{p}}_{\mathrm{o}} \leftarrow \dfrac{\grave{\mathrm{p}}^{\mathrm{o}} - {}_{\mathrm{B}}}{\sqrt{\acute{o}_{\mathrm{V}}^{2} + \mathring{a}}} \\ \mathrm{V}_{\mathrm{o}} \leftarrow \gamma\hat{\mathrm{P}}_{\mathrm{o}} + \beta \equiv \mathrm{V}_{\mathrm{M}(\tilde{a},\hat{a})}\left[\mathrm{P}_{\mathrm{o}}\right] \end{cases} \quad (7)$$

In this section, the mean of the standard deviation is removed from the variance of the distribution and then divided by the standard deviation to determine the threshold based on the activation function.

As shown in Figure 3, the M2BN flow chart can be used to calculate the optimal limit value normalized by size from the normal limit value.

Decision Tree (DT)

In this section, the CVD can be predicted using the DT technique to select the weight attributes. Similarly, the DT technique can achieve efficiency and reliability by starting at

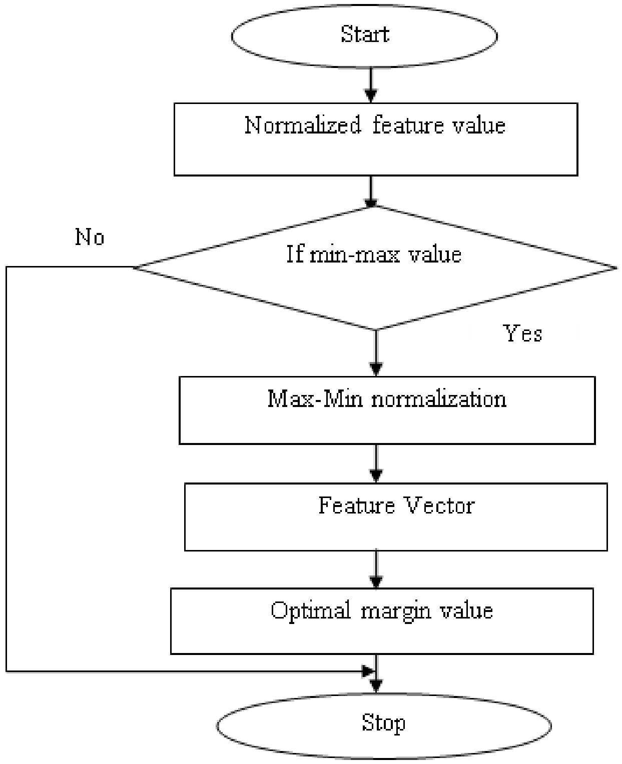

Figure 3: Flowchart diagram for M2BN

the root of the class labels tree. The decision rule proceeds by selecting additional nodes from the root node until they reach the guide node. The different types of leaf nodes can be categorized to choose the weight attribute for prediction. Furthermore, DT rules can be employed to forecast divergence in the mathematical model for selecting the attribute value for splitting gain. This decision tree method evaluated three commonly used categories: information gain, the Gini index, and the rate of increase.

Algorithm 1:DT

Input: Normalized value of V_o

Output: Weight attribute

Start

Step 1

Calculate the entropy of the entire dataset

$\mathcal{C}_t\_R_q(t)$

For each $R_{q(U)}$

Step 2

Calculate the entropy of the remaining attributes and get the result.

Step 3

Estimate the average information entropy of the considered attributes.

Step 4

Evaluate the gains attributes

Step 5

Compute the information gained from each attribute

If $R = \sum_{o=1}^{j} Z_o \log_2 z_o$ (8) End if

Step 6

Estimate the Gini index for each attribute in the dataset

If $F_o = 1 - \sum_{o=1}^{j} zo^2$ (9) End if

Step 7

Compute the information gain ratio for each attribute considering the breadth and uniformity of best attribute values.

If $F(u) = F_e(O_f|T_o)$ (10) End if

End for each

Step 8

Select the optimal weight

Return $F(u)$

End

It estimates the information gain of each attribute and selects the one with the largest gain to evaluate the optimal attribute. Let's assume R-entropy, F_o – Gini Index, F_E – gain ratio, O_f – information gain, T_u – split information, j-number of classes of target attribute, X_o – number of occurrences class, and o-instance class.

XGBoost ensemble voting-based feature selection (XGB-EVFS)

In this section, the XGB-EVFS method identifies the significant feature selection of CVD. Gradient boosting and XGBoost, both decision tree-based algorithms, have significantly enhanced oversampling performance. This process can be used to improve the current model's performance without altering the previous classifier by altering the invalid classifier. Implementing the XGB-EVFS method can help smooth out the loss function and prevent overfitting. The tree structure can be optimized by calculating each layer's leaf score, regularization, and objective function. To enhance the XGBoost function, smaller weights can be used when adding and normalizing leaf weights using gradient descent.

This tree structure is reused in subsequent iterations, and the gains for each feature are computed as nodes are split. XGBoost hyperparameter tuning to improve the prediction of heart disease.

Equation 11 shows that the sum of prediction scores in slope increments during dataset estimation can be calculated using the additive function. Where o-instance, $\hat{q}_o$ – represents the prediction, j-boost, p_u – sample of the training dataset, G-feature, G_j – value of the tree,

$$\hat{q}_o = F(u)\sum_{j=1}^{j} g_j(p_u),\ \ g_j \in G \tag{11}$$

XGBoost objective function combines the loss function and the regularization term. The model's predictive power is determined by the simplicity of the loss function and the regularization period. To minimize the loss function defined in equation 12, XGBoost computes the objective function. Let's assume the h_j – losss function, $\hat{q}_o$, q_o – predicted and actual label, $E(g_o)$ – responsible for penalizing the complexity of the functions.

$$\begin{cases} h_j = \sum_{o=1}^{m} h\left(\hat{q}_o, q_o\right) \\ \sigma_{bj} = \sum_{o=1}^{m} h\left(\hat{q}_o, q_o\right) + \sum_{j=1}^{j} \varepsilon(g_o) \end{cases} \tag{12}$$

Furthermore, it is described in Equation 13 as an estimate tree function to handle the overfitting problem. Where g_p – function tree.

$$g_p = z_y(p), z \in E^s, y : E^n \to \{1,2,..,s\} \tag{13}$$

Estimate the hyper parameter or constant factor for the total number of leaves in the tree. As shown in Equation 14, calculate the number of leaves associated with the data instances in the mapping process. Let's assume γ *and* λ – hyperparameter, γ – each leaf value, S- total number of leaves in the tree, z – denote the weight, α -leaf control,

$$E(g) = \gamma^s + \alpha(z) + \frac{1}{2}\lambda\left(\| z \|^2\right) \tag{14}$$

A tree addition model can be chosen to minimize the objective function as long as the predicted result is equal to the previous tree combined with the new tree. Moreover, the purpose of each step in the objective function is determined, and then a distinct quadratic Taylor approximation for each objective function is calculated, as shown in Equations 15 and 16. Where d_i – constant, O_{bj} – objective, g_s – objective function,

$$\begin{cases} \sigma_{bj}(s) = \sum_{o=1}^{m} h\left(q_o, \hat{q}_\sigma^{s-1} + g_s(p_o)\right) + E(g_s) + D_i \\ \sigma_{bj}(s) = \sum_{o=1}^{m} h\left(q_o, \hat{q}_\sigma^{s-1} + f_o\, g_s(p_o) + g_s(p_o) + \frac{1}{2} l_o\, g_\sigma^2(p_\sigma)\right) \\ \qquad + E(g_s) + \mathcal{D}_i \end{cases} \tag{15}$$

$$\begin{cases} f_o = \partial_{\hat{q}_o^{(s-1)}} h\left(q_o, \hat{q}_\sigma^{[s-1]}\right) \\ h_i = \partial_{\hat{q}_o^{(s-1)}} L\left(q_\sigma, q_o^{[s-1]}\right) \end{cases} \tag{16}$$

Calculate detailed voting features by removing the standard terms to connect their thresholds, as described in Equation 17.

$$\sigma_{bj}(s) = \sum_{O=1}^{m}\left[f_o\, g_s(p_o) + g_s(p_o) + \frac{1}{2} lo_\sigma g_\sigma(p_\sigma)\right] + \gamma^s + \alpha\sum_{k=1}^{\delta} z_k + \frac{1}{2}\lambda\sum_{k=1}^{s} z_k^2 \tag{17}$$

In this category, the results are determined by a combinatorial voting function using the specified hyper parameter values.

Evaluate W1 by fitting A value into a decision tree, as depicted in Figure 4. Subsequently, the second tree, B value, can estimate the remaining W-W1 and continue with the C value W-W1, W2 based on this iterative approach. By illustrating inferences from an instance, it is possible to reduce misclassification effectively. This approach selects 6 features like fps, xang, oldpeak, slope, ca and thal.

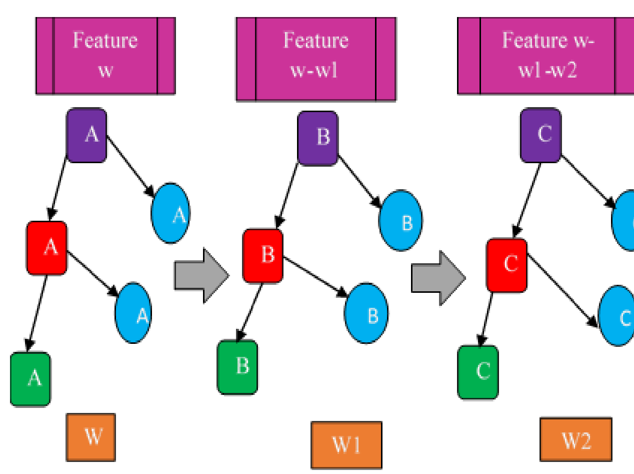

Figure 4: The Architecture Diagram for Gradient Boosting

Lightweight Recurrent Neural Network with a Long Short-Term Memory (LRNN-LSTM)

This section utilizes the LRNN-LSTM algorithm to classify heart diseases and reduce false diagnoses. Furthermore, cardiac patient data can be categorized using input and output gates. Similarly, the LRNN method generates a memory unit's hidden and current states. Moreover, the LRNN-LSTM method is assessed by processing the hidden states of the input and output gates in the memory cells. Similarly, each node in a neural network is assigned a weight to enhance its performance. The overall fitness function is determined by calculating the predicted probability of heart disease generated by the hidden layer of sigmoid activation functions.

A neuron mapping function associates inputs with selected outputs. Furthermore, the neuron's weight vector of the input bias weights is evaluated as given by Equations 18 and 19. Where E-real number, Z-weight, Z_0 – weight vector, $\mathcal{O}_s$ – input vector, p and q-input and output element, N-mapping function.

$$O_{bj}(\mathcal{S}) = E^{N+!} \rightarrow \mathcal{E} \quad (18)$$

$$\begin{cases} \mathcal{I}_x = \phi\left(\sum_{s-1}^{N} z_d o_s + Z_i\right) \\ E_q \rightarrow E_p \end{cases} \quad (19)$$

Compute the hidden state of the nonlinear mapping as illustrated in Equation 20. Let's assume $\mathcal{L}_s$ – hidden state, $\mathcal{O}_s$ – current input, ℓ_{s-1} – preceding input state, g-denotes nonlinear mapping, x_s – normalized exponential function, T_N – softmax function, S-time, $\tan_h$ – softmax expression.

$$\begin{cases} \mathcal{L}_s = g\left(\ell_{s-1}, \mathcal{O}_s\right) \\ \mathcal{L}_s = \tan_h\left(z_{1\ell}L^{s-1}, z_o L^{o_s} + V_l\right) \\ y_s = T_N\left(Z_{ly}L^s + V_y\right) \end{cases} \quad (20)$$

As shown in Equation 21, estimate the objective function's min-max probability and generate the output by increasing the time value. Let's assume $(s+1)-l$ time value, M-objective function, $\log x_s \,\&\, log\mathcal{X}_{s+1}$ – maximum minimum log.

$$\begin{cases} M(u,v) = -\sum_{s} h_s \log x_s \\ \mathcal{M}(s+1) = -l_{s-1}\log\mathcal{X}_{s+1} \end{cases} \quad (21)$$

Equation 22 describes an estimate that resets the memory cells using a sigmoid function forget gate. Let's assume i_t – memory unit or cell, $\mathbb{O}_s$ – current observation, g_{fs} – forget gate, i_s – output gate,

$$\begin{cases} g_{fs} = \sigma\left(z_{ogf}o_s + z_{lgf}L_{s-1} + V_{gf}\right) \\ o_{ns} = \tanh\left(z_{\mathbb{O}\mu\mathbb{O}_s} + Z_{l\mu}L_{s-1} + V_\mu\right) \\ \mu_s = g_{fs} i \mu_{s-1} + w_s Io_{ns} \\ o_t = \sigma(Z_{oi}\mathbb{O}_s + h_{li}L_s + V_i) \\ L_s = I_s i \tanh(\mu_s) \\ X_s = T_N(z_{lx}L^s + V_x) \end{cases} \quad (22)$$

Calculate the predicted probability of HD output by the hidden layer in a sigmoid activation function as described in Equation 23. Similarly, implementing a system-wide loss function can modify the weighted sum of the maximum and minimum time-slice expected loss. Where $\hat{x}$ – probability value, $\hat{x}_n$ – output probability of the HD, x, and x_o – class label, D-dimensional value,

$$h_{(\hat{x},x)} = \frac{1}{|C|}\sum_{n=1}^{N=|\mathcal{C}|} -\left(x_n.\log(\hat{x}_n) + (1 - x_n).\log\left(1 - \hat{\mathcal{X}}_n\right)\right) \quad (23)$$

Equation 24 demonstrates the calculation of the ratio of the vector labels in the overall fitness function. Where $x^{(s)}$ – original classification label, s-time frame, α – hyperparameter, I_G – overall fitness.

$$I_G = \alpha\frac{1}{S}\sum_{s=1}^{\mathcal{S}|} \ell oss\left(\hat{x}^s, x^{(s)}\right) + (1-\alpha).\,\mathbb{l}oss(\hat{x}^s, \mathbb{X}^{\mathcal{S}}) \quad (24)$$

Furthermore, memory cells are used in LRNN-LSTM to alter the basic tanh function. The sigmoid system is used as an activation function to predict heart disease and can classify heart disease data by fitness function.

Result and Discussion

This section suggests that the performance of the proposed model can be evaluated through training and testing. The LRNN-LSTM method can be used for CVD prediction classification, and various performance metrics such as accuracy, precision, recall, and false score should be used to estimate its performance. Therefore, these evaluation metrics should be utilized to ensure a reliable performance evaluation. Similarly, the proposed LRNN-LSTM method improves the accuracy of the classification system and provides analytics such as confusion matrices and performance measures to describe its performance despite the highly variable CVD datasets.

Table 3 describes the approximate accuracy achieved by training and testing the dataset for CVD prediction using the simulation parameters and verifying the prediction for CVD based on Python language using Jupyter Notebook.

Table 3: Simulation Parameter

Parameters	*Values*
Dataset name	CVD dataset
Tool	Anaconda
Language	Python
No of Records	1000
Training	700
Testing	300

Table 4: Performance Matrices

Metric	*Scoring*
$P_{recision}$	$\sum Tp / \sum Tp + \sum Fp$
Re_{call}	$\sum Tp / \sum Tp + \sum Fn$
Acc_{uracy}	$\frac{\sum T_p + \sum T_n}{\sum T_p + \sum T_n + \sum F_p + \sum F_n}$
$F1_{score}$	$2*(\sum recall * \sum precision)/(\sum Recall + \sum Precision)$
$True_{Positive\ Rate}$	$\frac{\sum T_p}{\sum T_n + \sum F_p}$

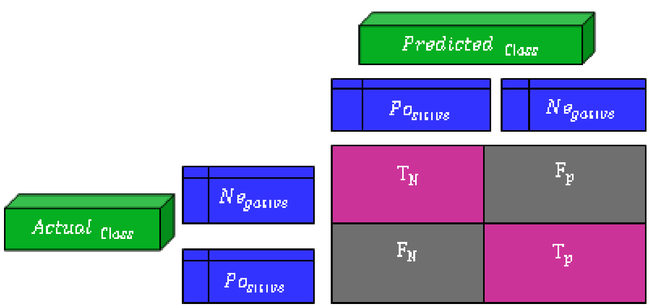

Figure 5: Confusion Matrix

Table 4 describes the performance measures TP and FP, which are implemented to correctly and incorrectly categorize the number of heart disease subjects. TN and FN represent the number of issues correctly and incorrectly classified as not having CVD.

Figure 5 shows the confusion matrix, which was used to evaluate the model's performance on both training and test data.

As shown in Figure 6, the CVD dataset can be used to obtain precision analysis and accurate estimates. Additionally, the DL model results are tested and trained to get a precise score for performing well on large amounts of data. However, when the proposed system's LRNN-LSTM method isimplemented for CVD prediction, the precision analysis of the estimate increases to 96.51%.

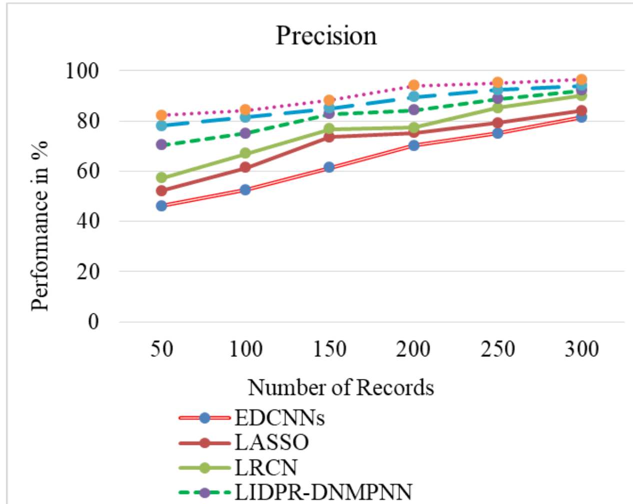

Figure 6: Analysis of Precision

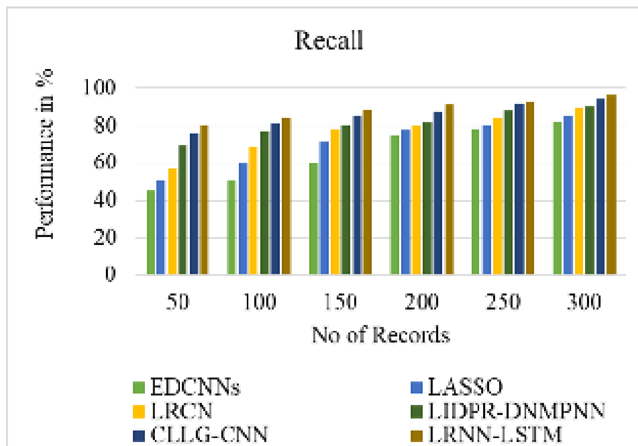

Figure 7: Analysis of Recall

In this section, the DL model results are tested and trained to analyze the performance in recall and obtain an accurate score that can be performed well on a large amount of data. Figure 7 illustrates how the CVD dataset can be used to obtain precise estimates of recall performance. Furthermore, when allocating with methods derived from the literature, such as EDCNN, LASSO, LRCN, LIDPR-DNMPNN and CLLG-CNN methods the recall performance produced as low outcome. Furthermore, this precision result is improved to 96.45% when using the proposed LRNN-LSTM method on the test and training datasets.

Utilizing the false score for performance analysis can evaluate the DL model's output and train it to provide accurate results that work well with large data sets. Figure 8 demonstrates how the CVD dataset may be used to produce precise estimates of false scores. Moreover, EDCNN is 36.23%, LASSO is 34.2%, LRCN is 31.1%, LIDPR-DNMPNN is 28.4%,

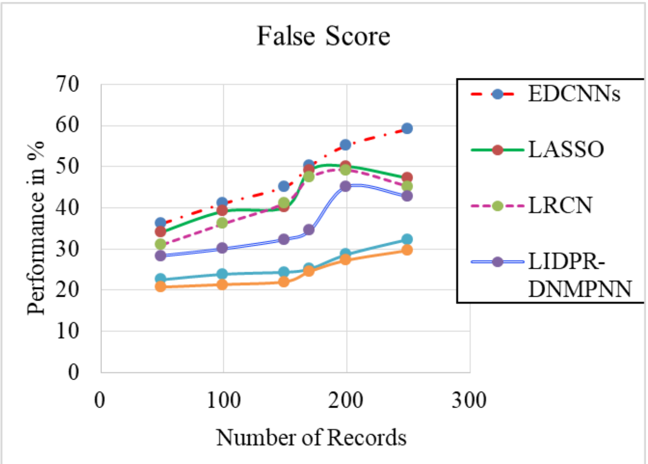

Figure 8: Analysis of False Score

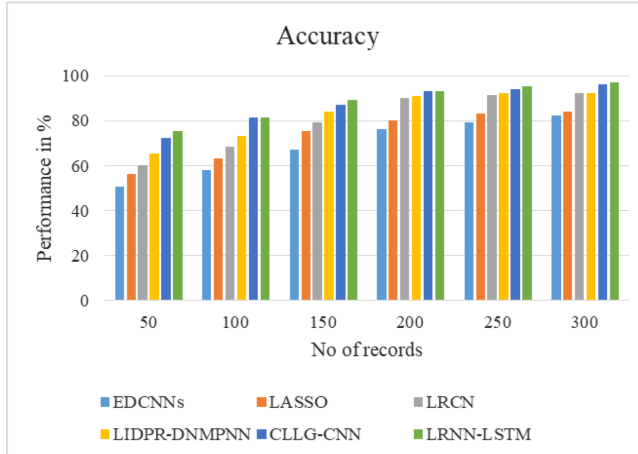

Figure 9: Analysis of Accuracy

Table 5: Overall performance for CVD prediction

Performance in %			
Parameters/ Performances	*Accuracy*	*Precision*	*Recall*
EDCNNs	82.14	81.43	82.15
LASSO	84.16	84.16	85.36
LRCN	92.2	90.18	89.65
LIDPR-DNMPNN	95.17	92.2	90.32
CLLG-CNN	96.01	94.03	94.38
LRNN-LSTM	97.16	96.51	96.45

CLLG-CNN is22.6%, and the proposed method LRNN-LSTM is 20.8% for 50 records.

The CVD dataset may be used to determine accuracy rates, as shown in Figure 9. Similarly, the accuracy performance drops to 81% when using literature-based approaches like EDCNNs, LASSO, LRCN, LIDPR-DNMPNN and CLLG-CNN. Moreover, the accuracy of the suggested LRNN-LSTM technique rises to 97.16% when it is used with the test and training datasets. Table 5 describes the overall performance for CVD prediction.

Conclusion

This paper proposes an LRNN-LSTM technique to predict heart disease analysis of the CVD feature dataset obtained from Kaggle, with both test and training phases. The proposed method uses LRNN-LSTM to classify the CVD results, thereby reducing misdiagnosis. Furthermore, by offering the M2BN technique, we can determine the marginal value of the dataset based on the collected data values. Moreover, DT techniques are used to identify and reduce the dimensions of long-range features. In addition, the XGB-EVFS method selects the optimal features of the CVD. Similarly, processing can classify heart disease data by analyzing the confusion matrix, performance evaluation, and other factors. The accuracy achieved by several approaches found through the literature review increased to 97.16% over the proposed LRNN-LSTM methods.

Acknowledgements

The authors would like to express their sincere gratitude to Thanthai Periyar Govt. Arts & Science College (A), Tiruchirappalli for providing the necessary facilities and support to carry out this research work. We also thank our college management, heads, colleagues and reviewers for their valuable feedback and constructive suggestions, which greatly improved the quality of this manuscript.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

Abdul Saboor, Muhammad Usman, Sikandar Ali, Ali Samad, Muhammad Faisal Abrar, Najeeb Ullah,(2022) "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms," Mobile Information Systems, vol. 2022, Article ID 1410169, 9 pages. https://doi.org/10.1155/2022/1410169.

Ahsan., M.M., and Siddique, Z.(2022) "Machine learning-based heart disease diagnosis: A systematic literature review," Artif. Intell. Med., vol. 128, Jun. 2022.

Ali., L.,*et al*., (2019) "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," in IEEE Access, vol. 7, pp. 54007-54014, doi: 10.1109/ACCESS.2019.2909969.

Alqahtani, A., Alsubai, S., Sha, M., Vilcekova, L., Javed, T.(2022 Aug 16) Cardiovascular Disease Detection using Ensemble Learning. Comput Intell Neurosci.; 2022:5267498. Doi: 10.1155/2022/5267498. PMID: 36017452; PMCID: PMC9398727.

Anju Ambika, Adline Freeda, Krithikaa Venket, Dinesh Ram, Logesh

Tanu, Praveen Kumar.(2023 November) Machine learning based heart disease prediction system. *AIP Conf. Proc*. 14 November 2023; 2963 (1): 020009. https://doi.org/10.1063/5.0184383.

Ambrish, G., Bharathi Ganesh, Anitha Ganesh, Chetana Srinivas, Dhanraj, Kiran Mensinkal. (2022) Logistic regression technique for prediction of cardiovascular disease, Global Transitions Proceedings, Volume 3, Issue 1, Pages 127-130, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2022.04.008

Ayatollahi, H., Gholamhosseini, L., & Salehi, M.(2019) Predicting coronary artery disease: comparing two data mining algorithms. BMC Public Health 19, 448 . https://doi.org/10.1186/s12889-019-6721-5

Bashir, S., Almazroi, A. A., Ashfaq, S., Almazroi, A. A., and Khan, F. H. (2021) "A Knowledge-Based Clinical Decision Support System Utilizing an Intelligent Ensemble Voting Scheme for Improved Cardiovascular Disease Prediction," in IEEE Access, vol. 9, pp. 130805-130822, doi: 10.1109/ACCESS.2021.3110604.

Ganie, S. M., Pramanik, P. K. D., Malik, M. B., Nayyar, A., and K. Kwak, S. (2023) "An improved ensemble learning approach for heart disease prediction using boosting algorithms," Computer Systems Science and Engineering, vol. 46, no.3, pp. 3993–4006.

Ghorashi, S., *et al*., (2023) "Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases," in IEEE Access, vol. 11, pp. 60254-60266, doi: 10.1109/ACCESS.2023.3286311.

Guarneros-Nolasco, L. R., Cruz-Ramos, N. A., Alor-Hernandez, G., Rodriguez-Mazahua, L.,and Sanchez-Cervantes, J. L. (2021 Oct) "Identifying the Main Risk Factors for Cardiovascular Diseases Prediction Using Machine Learning Algorithms," Mathematics, vol. 9, no. 20, pp. 2537.

Hossain, M.I., Maruf, M.H., Khan, M.A.R. *et al*.(2023) Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. Iran J Comput Sci 6, 397–417 (2023). https://doi.org/10.1007/s42044-023-00148-7.

Hosni, M., Carrillo de Gea, J.M., Idri, A., *et al*(2021) A systematic mapping study for ensemble classification methods in cardiovascular disease. Artif Intell Rev 54, 2827–2861. https://doi.org/10.1007/s10462-020-09914-6.

Iqbal,T., Soliman, O., Sultan, S., and Ullah, I. (2023) "Machine Learning Approaches for Segmentation of Cardiovascular Neurocristopathy Related Images," in IEEE Access, vol. 11, pp. 118301-118317, 2023, doi: 10.1109/ACCESS.2023.3325960.

Kartik Budholiya, Shailendra Kumar Shrivastava, Vivek Sharma,. (2022) An optimized XGBoost-based diagnostic system for effective prediction of heart disease, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 7, Pages 4514-4523, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2020.10.013.

Karthick, K., Aruna, S. K., Ravi Samikannu, Ramya Kuppusamy, Yuvaraja Teekaraman, Amruth Ramesh Thelkar, (2022) "Implementation of a Heart Disease Risk Prediction Model Using Machine Learning," Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 6517716, 14 pages. https://doi.org/10.1155/2022/6517716

Karthikeyan.M, Chaitanya Rajeev Myakala, Sai Chaitanya Chappidi. (2020). Heart Attack Prediction Using XGBoost. International Journal of Advanced Science and Technology, 29(06), 2392 - 2399. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/13543

Khandaker Mohammad Mohi Uddin, Rokaiya Ripa, Nilufar Yeasmin, Nitish Biswas, Samrat Kumar Dey, (2023) Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset, Intelligence-Based Medicine, Volume 7,2023, 100100, ISSN 2666-5212, https://doi.org/10.1016/j.ibmed.2023.100100.

Maheswari Subburaj*, Pitchai Ramu,(2019) Heart Disease Prediction System Using Decision Tree and Naive Bayes Algorithm, Current Medical Imaging 2019; 15 (8) . https://dx.doi.org/10.2174/1573405614666180322141259

Nadakinamani, R. G., Reyana, A., Kautish, S., Vibith, A. S., Gupta, Y., Abdelwahab, S. F., & Mohamed, A. W. (2022). Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques. Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/2973324

Nagavelli U., Samanta, D., Chakraborty, P.(2022 Feb) Machine Learning Technology-Based Heart Disease Detection Models. J Healthc Eng. 2022 Feb 27; 2022:7351061. Doi: 10.1155/2022/7351061. PMID: 35265303; PMCID: PMC8898839.

Osamah Sami, Yousef Elsheikh and Fadi Almasalha,(2021) "The Role of Data Pre-processing Techniques in Improving Machine Learning Accuracy for Predicting Coronary Heart Disease" International Journal of Advanced Computer Science and Applications (IJACSA), 12(6), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120695

S. Peerbasha, Y. Mohammed Iqbal, Praveen K.P, M. Mohamed Surputheen, & A Saleem Raja. (2023). Diabetes Prediction using Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression Classifiers. *Journal Of Advanced Applied Scientific Research*, *5*(4), 42–54. https://doi.org/10.46947/joaasr542023680

Rajliwall, N. S. Davey, R. and Chetty, G. (2018) "Cardiovascular Risk Prediction Based on XGBoost," 2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Nadi, Fiji, 2018, pp. 246-252, doi: 10.1109/APWConCSE.2018.00047.

Rahim, A., Rasheed, Y., Azam, F.,. Anwar, M. W., Rahim, M. A., and Muzaffar, A. W. (2021) "An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases," in IEEE Access, vol. 9, pp. 106575-106588, doi: 10.1109/ACCESS.2021.3098688.

Reddy, R., Vijaya Kumar, K., Prudvi Raju, M., Jogendra Kumar, Chandrasekaran Sujatha and Ravi Prakash, P.(2016) "Prediction of Heart Disease Using Decision Tree Approach.".

Reddy Thummala, G. S. and R. Baskar,(2022) "Prediction of Heart Disease using Decision Tree in Comparison with KNN to Improve Accuracy," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, pp. 1-5, doi: 10.1109/ICSES55317.2022.9914044.

Sai Krishna Reddy1,V., Meghana1, P., Subba Reddy2, N .V., and Ashwath Rao3,B.(2021) Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers, Journal of Physics: Conference Series, Volume 2161, 1st International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS 2021) 28-30 October 2021, Manipal, India, Doi: 10.1088/1742-6596/2161/1/012015.

Srinivasan, S., Gunasekaran, S., Mathivanan, S.K. *et al*.(2023) An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database. Sci Rep 13, 13588 (2023). https://doi.org/10.1038/s41598-023-40717-1

Sumaira Ahmed, Salahuddin Shaikh, Farwa Ikram, Muhammad Fayaz, Hathal Salamah Alwageed, Faheem Khan, Fawwad Hassan Jaskani,(2022) "Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models", Journal of Sensors, vol. 2022, Article ID 3730303, 21 pages, 2022. https://doi.org/10.1155/2022/3730303

Qadri, A. M., Raza, A., Munir, K., and Almutairi, M. S.(2023) "Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning," in IEEE Access, vol. 11, pp. 56214-56224, doi: 10.1109/ACCESS.2023.3281484.