International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimal Squirrel Search-Gradient Decision Tree for Cardiovascular Disease Risk Prediction Using Machine Learning

R.Shanthi[1],M.Shanthi[2], P.Mohan Kumar[3], Bharath Srinivasaiah[4], N.Herald Anantha Rufus[5],S.Balaganesh[6]

Submitted: 18/01/2024 **Revised**: 27/02/2024 **Accepted**: 04/03/2024

Abstract: In recent years, big data usage in the Internet of Things (IoT) has rapidly advanced in the medical field. Furthermore, employing big data and IoT for predicting cardiovascular disease (CVD) enables more precise and timely detection of potential risks, thereby enhancing prevention and treatment approaches. Thereafter, cardiovascular diseases can be predicted by utilizing big data in IoT to create personalized and proactive interventions. Nevertheless, acquiring medical information for early disease detection and assigning timely treatment is essential. Furthermore, more prevalent forms of heart disease can impact the heart's blood flow and result in heart attacks. However, accurately predicting heart disease in medical data analysis presents a significant challenge. So to solve this problem, we proposed an Optimal Squirrel Search-Gradient Decision Tree (OSSGDT) method to classify and accurately predict CVD's high-level and low-level risk factors. Moreover, the pre-processing method utilizes a Normalized Low Pass Filter (NLPF) to eliminate noise and enhance the quality of the smooth areas in the image. Furthermore, the Fuzzy Centroid Cluster Means (FCCM) algorithm can be utilized to iteratively update cluster centres for image segmentation. Furthermore, the Relief Feature Selection Algorithm (RFSA) can be utilized to estimate the weight of each feature. Finally, the OSSGDT method based on Machine Learning (ML) techniques can predict high and low-risk factors for CVD. Moreover, using the proposed OSSGDT method, we can evaluate the performance of heart disease prediction based on accuracy, recall, precision, false ratio, and sensitivity.

Keywords: *Cardiovascular disease prediction, Optimal Squirrel Search-Gradient Decision Tree, NLPF, Big data, Internet of Things, RFSA, machine learning*

1. Introduction

Big data analytics platforms have been gradually replacing traditional warehousing paradigms in recent years, particularly when it comes to processing, analysing, and storing enormous amounts of clinical data. This shift has enabled the concurrent and effective analysis of structured, semi-structured, and unstructured data. Big Data analytics provides a centralised solution that makes it possible to expedite the processing of huge data collections. Furthermore, the integration of IoT features in medical devices has the potential to enhance service quality and efficiency, ultimately expediting the adoption of IoT in the healthcare industry. As a result, IoT technology stands to benefit the elderly, individuals with chronic diseases, and those in need of continuous treatment. This progress has the potential to significantly improve the overall quality of care and support for these specific demographics [1-2].

Moreover, heart and vascular problems or CVD, cause more than 17 million deaths worldwide each year. The timely identification and diagnosis of cardiovascular disease play a pivotal role in mitigating mortality rates and reducing the burden of illness. Within the realm of clinical procedures, cardiac imaging has evolved into an indispensable tool, providing comprehensive insights into the heart's structure and functionality. Echocardiography and Computed Tomography (CT) are popular methods of cardiac imaging, providing clinicians with valuable data for accurate diagnosis and effective management of cardiac diseases. These imaging techniques not only aid in the early detection of cardiovascular ailments but also contribute significantly to the formulation of tailored treatment strategies, ultimately enhancing patient outcomes and quality of life [3].

Heart disease presents a range of symptoms, including chest pain, shortness of breath, fatigue, irregular heartbeat, weakness or numbness on one side of the body, difficulty speaking or understanding words, blurred vision, and severe headaches. It is important to note that the symptoms of different heart diseases can overlap, making it challenging to differentiate them based solely on symptoms. Therefore, it is crucial to seek medical attention and undergo thorough diagnostic testing to accurately identify and address any potential heart conditions [4].

The arteries of the heart play a crucial role in delivering oxygen-rich blood to all parts of the body. However, heart problems can result in blockages or restrictions in the blood vessels, leading to serious health issues. It is important to note that the cardiac diagnostic applications of the past are now considered outdated

[1]*Department of Computer Applications, B.S Abdur Rahman Crescent Institute of Science and Technology , Vandalur, Tamil Nadu 600048, India. Email:shanthi@crescent.education*
[2]*Department of Information Technology, R.M.K Engineering College, Kavaraipettai, Tamil Nadu- 601206, India. Email: msi.it@rmkec.ac.in* (Corresponding Author)*
[3]*School of Computer science and Engineering VIT university Vellore, Vellore, Tamil Nadu 632014 Email:pichaimuthumohankumar@gmail.com*
[4]*Engineer Lead Sr., Elevance Health Inc, Richmond VA-23230. Email-bharathsrinivasaiah@gmail.com*
[5]*Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr Sagunthala,R & D Institute of Science and Technology, Chennai-600062, Tamil Nadu, India. Email:drrufus@veltech.edu.in*
[6]*Assistant Professor, Department of Computer Science and Engineering, P.S.R Engineering college, Sivakasi, Tamil Nadu 626140,India. Email: bala3cse@gmail.com*

and less efficient. As such, it is imperative to improve the content and technology used in cardiac diagnostics to ensure accurate and timely detection of heart-related issues [5].

This paper presents an OSSGDT method for accurately predicting both high-level and low-level risk factors for CVD. The ACDC dataset stands as the most extensive publicly accessible dataset for the estimation of cardiac MRI. The pre-processing technique contains the NLPF method to remove noise and improve the overall quality of smooth areas in the image. Furthermore, the FCM algorithm is used to iteratively update the cluster centres for image segmentation. Moreover, RFSA is employed to select the most appropriate features and estimate the weight of each feature. Similarly, the ML-based OSSGDT method can be used to predict cardiac disease and achieve a more effective and precise assessment.

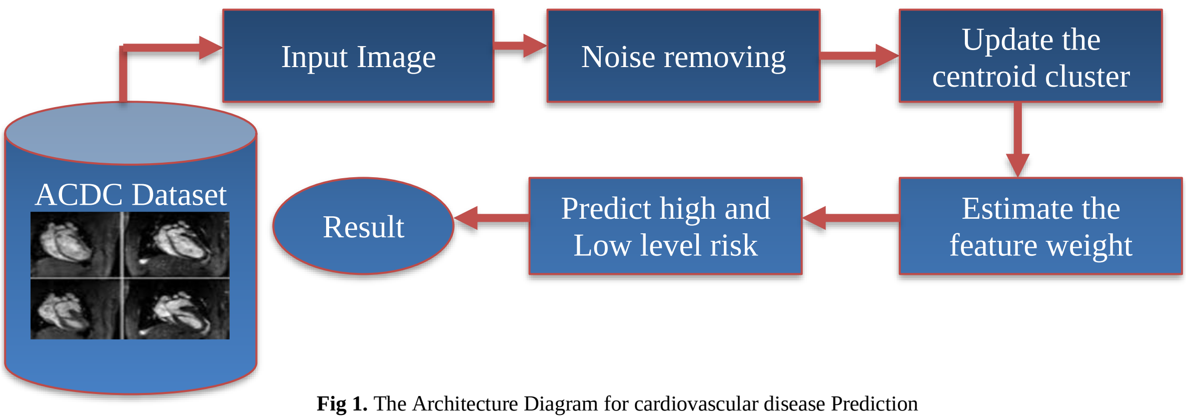


Fig 1. The Architecture Diagram for cardiovascular disease Prediction

The architecture diagram for CVD prediction, as depicted in Figure 1, offers a comprehensive framework for evaluating high-level and low-level risk prediction. The diagram facilitates the analysis of X-ray angiography data through pre-processing, segmentation, feature selection, and classification phases, ultimately leading to quick and accurate predictions.

2. Literature Survey

The enhanced Support Vector Machine (SVM) classifier program can enhance accuracy and efficiency through parameter tuning. Furthermore, the SVM method offers an effective evaluation approach for processing streaming medical data [6]. Similarly, Convolutional Neural Networks (CNNs) have been used to identify clinical images of cardiac lesion plaques by extracting different phase features from raw input Optical Coherence Tomography (OCT) images [7]. To further enhance network performance, the Bi-CLSTM (Bi-Directional Convolutional Long Short Term Memory) technique will be applied. Similarly, the inclusion of multiscale transformations permits the analysis of long-range temporal information in addition to gathering local and global features [8]. The Quantum Support Vector Classifier (QSVC) technique can be used for multi-class classification. However, examples of CVD may include heart attacks, arrhythmias, and coronary artery disease [9]. A Binomial Linear Regression (BLR) model can be used to develop a reliable CVD risk score. Furthermore, by adopting the proposed approach, its key elements can be demonstrated [10].

Table 1. Prediction of Cardiovascular Disease Using Deep Learning Approaches Based On IoT

Author	Year	Technique	Method	Limitation
Chitra Balakrishnan [11]	2023	Deep Learning (DL)	Pre-trained Recurrent Neural Network (PRCNN)	Gathering medical data for early diagnosis and prompt treatment is crucial.
Hasan, N. I [12]	2019	DL	CNN	CVD classifications are crucial for efficiently and promptly treating patients.
A. Hauptmann [13]	2019	DL	CNN	High acceleration factors are required for real-time estimation of ventricular volumes
Morales MA [14]	2022	DL	2D-CNN	However, the Electrocardiogram (ECG) signal is required to reconstruct the medical images.
N. V. MahaLakshmi [15]	2023	Machine Learning (ML)	Improved Particle Swarm Optimization (IPSO)	Prediction of CVD is a difficult task in the evaluation of clinical data.

As presented in Table 1, the performance of cardiovascular disease prediction using IoT-based DL approaches is mentioned in the literature on the technique, methodology and limitations.

Similarly, CVD prediction systems for patients can be designed using Hybrid Deep Neural Networks (HDNNs) to extract relevant features from input data [16]. Therefore, an Enhanced Deep Learning-CNN (EDCNN) method can be proposed to support the prognosis of cardiac patients. Furthermore, these include multi-layer perceptron models with supervised learning approaches [17]. A Gaussian mixture model can be integrated using an Online Infinite Echo State Gaussian Process (OIESGP) to achieve a real-time predictive blood pressure distribution based on input data [18]. Accordingly, the CRNN method can be used to automatically detect normal cardiac auscultation, as well as aortic stenosis and mitral stenosis, by processing ray Phonocardiogram (PCG) signals [19]. The significance of IoT-based CVD can be assessed by utilising a Physics-Guided Deep Learning Network (PGDLN). The clinical use of hemodynamic parameters is constrained by the associated measurement risks and the potential for significant medical costs [20].

Table 2. Deep Learning for Cardiovascular Disease Prediction using Big Data

Reference No	Year	Method	Accuracy
21	2022	Attention-Based CNN (ACNN)	91.56%
22	2022	Deep Graph CNN (DGCNN)	90%
23	2022	DNN	93%
24	2023	Naive Bayes (NB)	90.78%
25	2023	Long Term Short Memory (LSTM)	89.56%

As indicated in Table 2, the accuracy of methods for predicting CVD using big data can be assessed through a DL approach.
Similarly, features can be chosen based on pre-processed data values by utilising the Harris-Hawk Optimisation (HHO) approach. Furthermore, an improved DL-based framework can be proposed to predict CVD [26] Moreover, the SVM approach could be suggested for acquiring medical data via IoT devices and utilizing current medical sensors to predict CVD in real time [27]. The input features for the fuzzy approach can be processed utilising the Fuzzy Deep Convolutional Network (FDCN) method by implementing the transformed features. However, in cases of ambiguous data, the associated noise and uncertainty cannot be predicted [28]. They suggested that a CVD diagnosis can be performed using an IoT-based system incorporating a Recurrent Neural Network (RNN) [29]. Similarly, asymmetric data can be tested utilising a Generative Adversarial Network (GAN) model. Nevertheless, the accuracy of CVD detection deteriorates when the balance of data is lower [30].

3. Proposed Methodology

In this section, accuracy can be enhanced using a method called OSSGD to classify and predict high-level and low-level risk factors for CVD. Moreover, the ACDC dataset is the largest publicly available dataset for cardiac MRI evaluation. After that, a total of 1053 MRI images were collected, with 802 images allocated for training and 251 for validation. Furthermore, pre-processing is conducted to remove noise and enhance the quality of smooth areas in the image through the implementation of the NLPF method. Afterwards, cluster centres can be updated based on image segmentation using the FCM algorithm. Moreover, the RFSA approach can be used to select features and estimate their weights. Similarly, the proposed OSSGD method can be utilized to predict heart disease through performance evaluation. Therefore, the proposed approach offers a promising solution for improving the accuracy of CVD risk factor predictions.
The architectural diagram analysis includes the proposed OSSGDT method as a means to enhance the image quality. Figure 2 demonstrates how this method can remove noise using the NLPF method, update cluster centres with the FCM method, and estimate feature weights through the RFSA method. Similarly, the OSSGDT method enhances the accuracy of classifying and predicting CVD. By eliminating noise, the analysis and interpretation of the diagram can be conducted with greater accuracy and reliability. Updating cluster centres also leads to a more comprehensive understanding of the diagram, ultimately improving its overall quality and effectiveness.

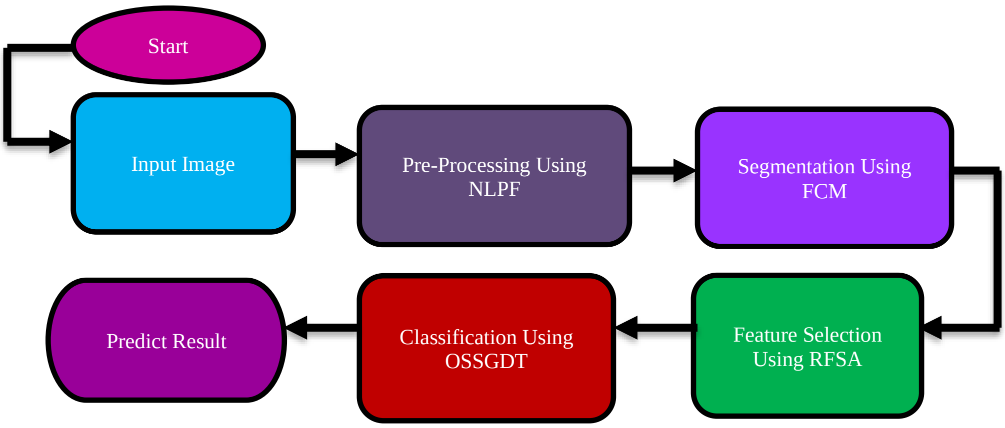


Fig 2. The Proposed OSSGDT Method for Architecture Diagram

3.1 Image Dataset Collection

In this regard, the Automated Cardiac Diagnostic Challenge dataset (ACDC) is the largest publicly available cardiac MRI evaluation dataset. The ACDC dataset derives from various well-defined cardiac pathologies and allows for effective ML model training and estimation of physiological parameters obtained from cine Magnetic Resonance Imaging (MRI). The dataset consists of 150 patients equally distributed in 5 different categories, each with specific physiological parameters.

Moreover, additional data such as weight, height, diastole and systole is provided for each patient.

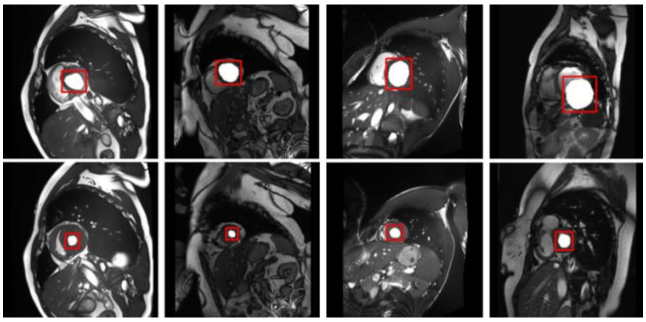


Fig 3. Cardiovascular Disease ACDC Dataset

As illustrated in Figure 3, the ACDC dataset is derived from a variety of well-defined cardiac pathologies and provides additional information for each patient, such as weight, height, diastole, and systole.

3.2 Normalized Low Pass Filter (NLPF)

In this section, pre-processing can be performed using the NLPF method to eliminate noise and enhance the overall quality of smooth areas in the image. Similarly, the analysis of Internet of Things (IoT) technology can effectively reduce the noise present in images and evaluate the accuracy of image performance. Moreover, filters like median, low-pass, and Gaussian filters are used to remove noise in CVD images. Furthermore, the de-noising level represents an effective filter for removing noise and improving the image. After that, spatial noise can also be eliminated by applying the NLPF method to adjust each pixel based on the average estimate of neighbouring pixels from the median filter. The use of IoT based on the NLPF method enables the de-blurring of a low-pass filter to eliminate sharp contrast changes and preserve smooth areas. Furthermore, the process convergence and image normalization can be identified using Min-max normalization.

The Gaussian probability distribution function describes an image affected by Gaussian filter. Moreover, the grey value noise function can be estimated using Equation 1. Where $q_{(e)}$ −noise function, e-grey scale value, μ −average value, e- Gaussian distribution curve, σ −standard deviation.

$$q_{(e)} = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-0.5\left(\frac{e-\mu}{\sigma}\right)^2\right] \quad (1)$$

Speckle noise degrades the quality of medical images and significantly reduces image contrast. It is calculated using a gamma distribution probability function as described in Equation 2. Where e^{xp} −expression, γ −gamma distribution.

$$q_e = \frac{(e^{s-1})e^{xp}\left(\frac{-e}{i}\right)}{(\gamma-1!)e^{xp}(\gamma)} \quad (2)$$

A median filter reduces spatial noise by replacing each pixel with the average value of the surrounding pixels. A low-pass filter eliminates sharp contrast changes by blurring and preserving smooth areas in the image. Calculate the simple averaging mask. Where N-denote order filter, i_p −average mask, X_p −unit matrix.

$$3\times3 i_p\begin{bmatrix}1 & 1 & 1\\1 & 1 & 1\\1 & 1 & 1\end{bmatrix}$$
$$q\times q, i_p \frac{1}{q^2}(q\times q, X_p) \quad (3)$$

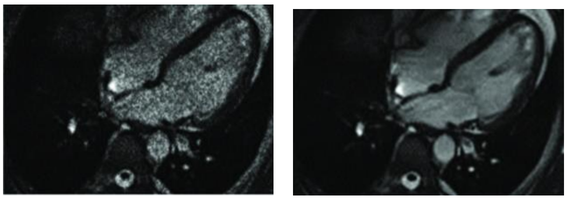
Median filtering allows to reduction and dispersal of the disruptive noise caused by feature localization. To further minimize these imperfections, a Butterworth low-pass filter is employed. Butterworth low-pass filters are utilized for smoothing images in the frequency domain, removing high-frequency noise from digital images while preserving low-frequency components. Calculate the system function of the Butterworth low-pass filter. Let's assume p −mean, C_0 −cut off frequency, f −butterworth filter, q-order of filter.

$$f(p,q) = \frac{1}{1+\left[\sqrt{\frac{p^2+q^2}{C_0}}\right]} \quad (4)$$

Each eigenvalue with a minimum value is replaced by and the maximum value is replaced to calculate the normalized process as described in Equation 5. Let's assume i-data point, x-value, X_{min} −minimum value of data point, U_q −normalized value, U_{ma} −maximum value of data point.

$$U_q = \frac{u_a - U_{mi}}{U_{ma} - U_{mi}} \quad (5)$$

The overall quality of smooth areas in normal values images can be improved by using a Butterworth low-pass filter to reduce and disperse the disturbance noise caused by the localization of the mean filtered features.



(A) Noise image (B) De-Noise Image

Fig 4. Image Pre-Processing Based on Cardiovascular Disease

Figure 4 demonstrates the possible to generate noise and de-noise images using the CVD images obtained from the dataset through pre-processing.

3.3 Fuzzy Centroid Cluster Means (FCCM)

In this category, the FCCM algorithm is capable of iteratively updating the cluster centres for image segmentation. By dynamically updating the cluster centres, the FCCM algorithm can effectively improve the quality and accuracy of the image segmentation, leading to more precise and meaningful segmentation results. The FCCM method uses an unsupervised iterative process to optimize the objective function by minimizing the least squares error through iteration. The FCCM method enables the segmentation of the CVD image into a determined number of clusters. Moreover, centres of clusters can be found by minimizing the variance function. Similarly, optimal segmentation can be achieved by minimizing the dissimilarity function if the weighting metric has high fuzzy values. The enhancement of the FCCM method can be achieved through the incorporation of a fuzzy membership matrix and multidimensional matrix-based cluster centres, which can be distributed to improve its performance.

Algorithm 1: FCCM

Input: Normalized Value X_N
Output: Update of clustering centresX_{ab}
Start
1. The fuzzy membership sequence is approximated by satisfying the limitations.

$$\sum_{a=1}^{d} X_{ab} = \left(U_q\right) 1,1 \leq b \leq q,\ X_{ab} \in [0,1] \tag{6}$$

For each $a = 1,2,\ldots, q,\ b = 1,2,\ldots, p$
Calculate the updated cluster centroids

$$Y_a = \frac{\sum_{a=1}^{q}\left(X_{a_b}^{p}\right) u_a}{\sum_{a=1}^{q}\left(x_{a_b}^{p}\right)} \tag{7}$$

End for each
2. Calculations are made and constant based on differences between data points and centroids until the improvement over the previous iteration is less than the threshold.

$$B = \sum_{a=1}^{q} \sum_{b=1}^{d} X_{a_b}^{p} \left|u_a - y_b\right|^2 \tag{8}$$

3. Evaluate the fuzzy membership matrix and continue.

$$X_{ab} = \left[\sum_{w=1}^{d} \frac{\|u_b - y_b\|}{\|u_1 - y_w\|}\right]^{\frac{2}{p-1}} \tag{9}$$

Return X_{ab}
Stop

As discussed in Algorithm 1, the FCCM method offers the ability to iteratively enhance the membership ranks and cluster centres. This allows for the precise positioning of cluster centres within the CVD image dataset. Let's assume X_{ab} −fuzzy membership matrix, d-cluster, q-number of collection sample set, p-higher weighting index value, i and q- multidimensional matrix, Y_a −centroid cluster, U-set of sample collection, B- function, y_w −data points, a and b-matrix.

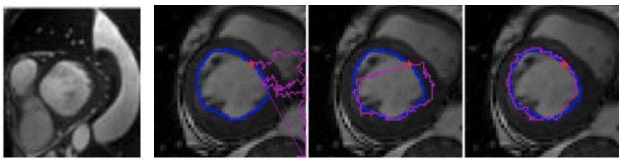


Fig (5) Cardiovascular Disease Segment Image

Figure 5 shows the results of the original and segmented images for CVD during the model training process for images in the ACDC 2017 dataset at different levels. This image shows the evolution of the segmentation results as a sample for different iterations during training.

3.4 Relief Feature Selection Algorithm (RFSA)

In this section, the utilization of RFSA techniques can significantly enhance the estimation of feature weights. By employing RFSA techniques, a more accurate and precise assessment of the importance of each feature can be achieved. The RFSA approach is designed to assign a weight to each data set feature and then automatically update those weights. Features with higher weight values are prioritized for selection, while those with lower weight values are denied. The process for determining feature weights within the RFSA remains consistent, as the algorithm iteratively recovers with random training samples without selection permutation and parameters. This RFSA approach ensures that the algorithm can effectively identify and prioritize the most relevant features within the CVD image data set.

Algorithm 2: RFSA
Input: Update the cluster centroid X_{ab}
Output: Each feature weight K
Begin
1. Estimate the total number of cluster centres← $\mathcal{Q}$
2. Calculate the quantity of feature dimensions ← $\mathcal{C}$
3. Set of feature weight $K_I \leftarrow 0.0$;
For each $w \leftarrow 1\ to\ p$ do
The target samples can be selected randomly S_w
Find the nearest hit-and-miss F_P
For each $I \leftarrow 1\ to\ i$ do

$$K[I] \leftarrow \mathcal{K}[i] - D_{iff}(\mathcal{J}, \mathcal{S}_{\mathcal{W}}, \mathcal{F})/\mathcal{P} + \mathcal{D}_{iff}(\mathbb{I}, s_w, p)/\mathcal{P} \tag{10}$$

End for each
End for each
Return K;
End

As described in Algorithm 2, random training samples are utilized to calculate the updated weight vector for each feature in every target sample via parameter iteration. Let's assume K-weight, Q-number of cluster, C-feature dimension, p-parameter, s_w −randomely training sample, F_P −hit and miss, $K[I]$ −weight set, W-target.

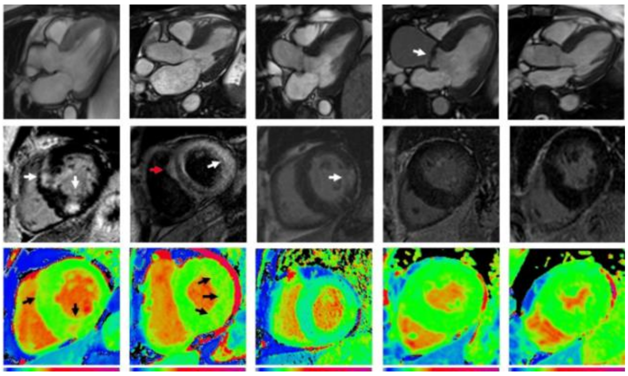


Fig 6. CVD Feature Selection

As illustrated in Figure 6, the variances in feature selection of characteristics in the MRI images can be computed using the feature weights among the similar appearances.

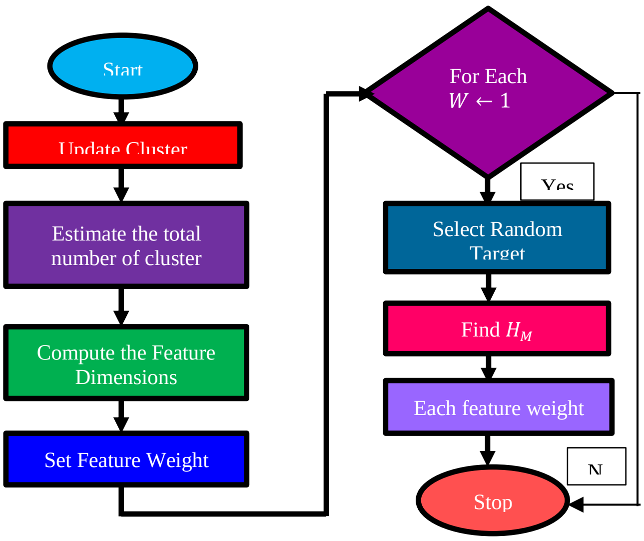


Fig 7. Flowchart Diagram for RFSA

Figure 7 illustrates the process of determining feature weights by iterating random training samples and updating cluster centroids using the RFSA flowchart diagram.

3.5 Optimal Squirrel Search-Gradient Decision Tree (OSSGDT)

The proposed OSSGDT method can predict both high-level and low-level risk factors for cardiovascular disease with high accuracy. Furthermore, using the proposed OSSGDT method to enhance the diagnosis of heart disease in IoT leads to improved effectiveness. The squirrel's location is considered a separate decision variable, and the distance between the food and individual squirrels is similar to the fitness of the objective function. Furthermore, the fitness of each flying squirrel can be evaluated by entering the decision variable values into a custom fitness function and computing the related values. The optimal parameters obtained through squirrel search can predict the values of the gradient decision tree model. In the model, decision trees serve as weak learners, with each new tree added sequentially from the previous tree with minimal residual loss. The model integration is achieved by observing the direction of the negative gradient rather than using weighted data. Similarly, using big data analysis can predict high and low-risk factors for CVD and improve the high accuracy.

Algorithm 3: OSSGDT

Input: Each feature weights K

Output: Classify high and low-risk prediction -$H_P(u)$

Start

1. Initialize the squirrel search process to a random variable.

For each $a(S_c) = (1, q_m, +1)$

For each $B(s) = (1, q_c + 1)$

$$H_{ac} = \ell^c + S * (X_c - \ell_c) \quad a = 1,2,..,Q_m; b = 1,2,\ldots,Q_c \quad (11)$$

End for each

End for each

2. Evaluate Rate the fitness of each flying squirrel

If $a(s) = (1, Q_m, +1)$

$$h_a = \hbar_a\left(\mathcal{H}_{a,1}, \mathbb{H}_{a,2,\ldots,h_{a,q}}\right) \quad (12)$$

End if

3. Evaluate the quality of the food available to the flying squirrel according to its location and improve its fitness value

$$[R_h, \mathcal{R}_a] = R^H \quad (13)$$

4. Create new spaces with gliding or random walks

$$H_{iz}^{Q} = \begin{cases} H_{iz}^{O} + C_e D_e\left(H_{fz}^{\sigma} - \mathcal{H}_{iz}^{\mathfrak{o}}\right), if\ s_1 \geq m \\ S_l \quad Otherwise \end{cases} \quad (14)$$

For each H_r

If $S_1 > m$

Else

$h_{iz}^{q} = S_l$

End if

End for each

5. Calculate the optimization parameters obtained by the squirrel search

6. Calculate the initial constant values

$$h_0(u) = I_p(\gamma) \sum_{a=1}^{Q} l\,(v_a - \gamma) \quad (15)$$

7. Average the weight of different trees and determine step size and minimum loss reduction.

$$\gamma^p, \eta^{\mathscr{p}} = I_p(\gamma, \eta) \sum_{a=1}^{q} L\left(v h_{p-1}(U_{w,a}) + \eta_j(\mathcal{U}_{w,a}; \gamma) + y_z + \frac{1}{2}\beta\|\gamma\|^2\right) \quad (16)$$

8. Update the fitness model of high and low-risk prediction

$$\mathcal{H}_{\mathcal{P}}(u) = \mathbb{H}_{p-1}(x) + \eta_p^j \eta_{j}(U_{w,a}; \gamma) \quad (17)$$

Return ← $H_p(u)$

Stop

As illustrated in Algorithm 3, the minimum loss reduction can be calculated by selecting the average weight of different trees. Additionally, the fitness models can be improved to predict both high and low risks of CVD accurately. By integrating these enhancements, the overall effectiveness and precision of the models can be significantly improved. Let's assume F-flying squirrel, Q_m −population size, q_c −decision variable, H_r −scalling factor, M-predator presence, s-range, s_c −random variable, D_e −gliding constant, X_j, l_j −upper and lower bound decision variable, p-additive function, U-determine data, Z-tree, l-leaf count, $\eta_j(U_{w,a}; \gamma)$ −decision model, I_P −mminimum argument variable, S_1 −return function, H_{fz}^{O} −Flying squirrel halting constraint, Q-feature value, R_h −sorted flying, R_a −sorted index, h_p −fitness model.

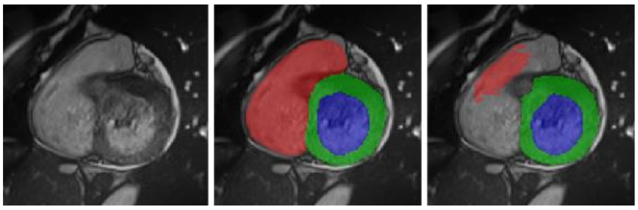


Fig 8. Classifying the CVD Prediction

As depicted in figure 8, which categorizes CVD prediction, it demonstrates a typical outcome in the classification prediction. Additionally, the accuracy of heart disease diagnosis classification can be enhanced.

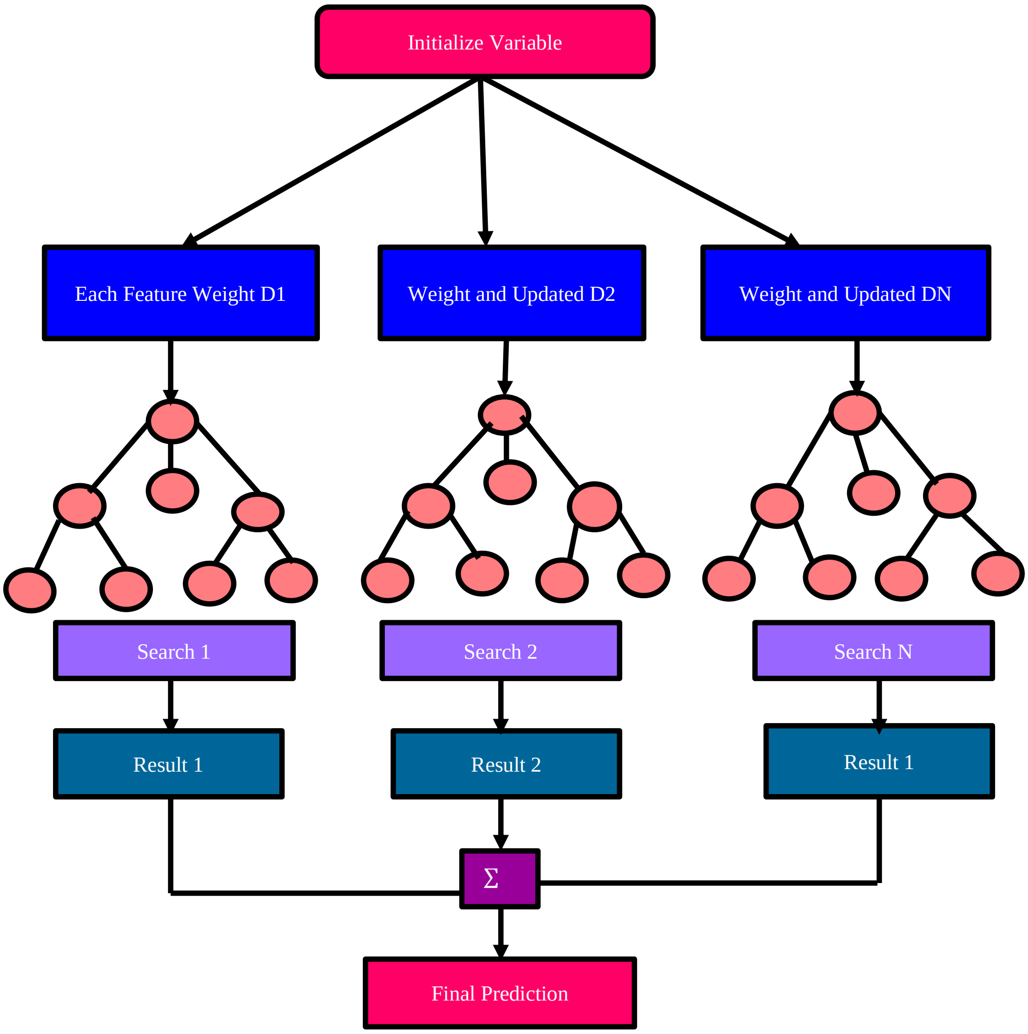


Fig 9. The Architecture Diagram for OSSGDT

As shown in Figure 9, based on the architectural diagram of the proposed OSSGDT, big data analysis provides estimates of high and low-risk factors for CVD to estimate the appropriate step size for tree weight determination.

4. Result and Discussion

As described in this paper, the OSSGDT technique is presented via an ML approach to analyse extensive datasets within the IoT domain. Moreover, the proposed OSSGDT method is implemented and allows a comparative analysis of its performance against recent methods. At present, measurements, comprising sensitivity, precision, accuracy, recall, and false ratio, are available in the literature for evaluation approaches such as SVM, Bi-CLSTM, and CNN to create an assessment for CVD.

$$Sim^{Parameter} = \begin{cases} A_{cc} = \frac{[\mathcal{T}^N - \mathcal{T}^P]}{[\mathcal{T}^P + \mathcal{F}^P + \mathcal{T}^N + \mathcal{F}^N]} \\ P_{rec} = \frac{[\mathcal{T}^P]}{[\mathcal{T}^P + \mathcal{F}^P]} \\ R_{ec} = \frac{[\mathcal{T}^P]}{[\mathcal{T}^P + \mathcal{F}^N]} \\ F1_M = \frac{[2 \times P^{re} \times R^{ec}]}{[\mathcal{P}^{re} + \mathcal{R}^{ec}]} \\ S_{en} = \frac{[\mathcal{T}^P]}{[\mathcal{T}^P + \mathcal{F}^N]} \end{cases} \quad (18)$$

Table 3. Simulation Parameter

Simulation	Value
Dataset Name	ACDC Dataset
No of Images	1053
Training	802
Testing	221
Tool	Jupyter
Language	Python

As described in Table 3, the enhancement of content can be achieved through the utilization of Python language within the Jupyter tool for experiment and training, employing image processing techniques based on the IoT.

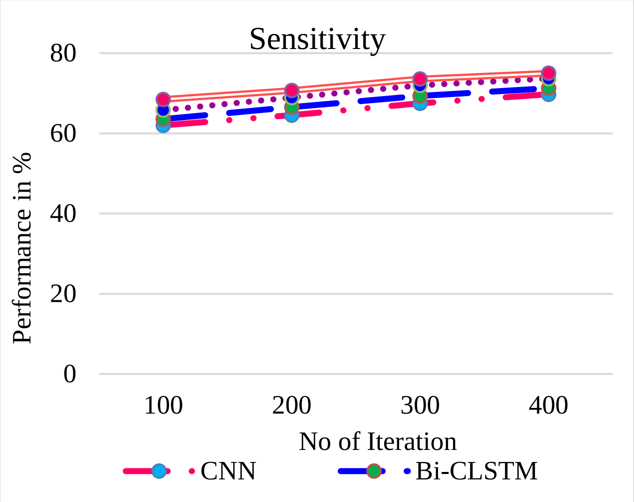


Fig 10. Performance of Sensitivity

As depicted in Figure 10, sensitivity analysis shows a crucial role in forecasting both high and low accuracy levels for CVD through the utilization of image processing techniques on extensive datasets. Specifically as declared in the existing literature the accuracy of SVM, CNN and Bi-CLSTM models is as low as 62%, through which they can establish the sensitivity performance. However, with testing and training, the accuracy of sensitivity analysis can be significantly improved, reaching an impressive 74.98% accuracy rate. The accuracy of the proposed OSSGDT method has significantly improved the efficiency in the field of CVD detection and analysis.

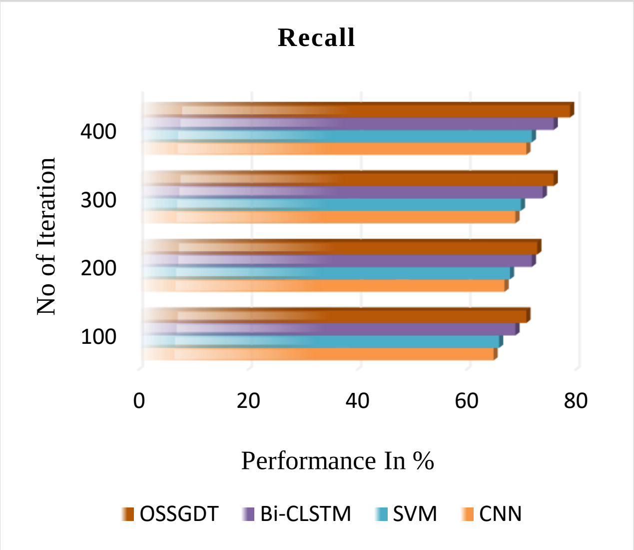


Fig 11. Analysis of Recall

Figure 11 demonstrates the significance of recall performance analysis in predicting CVD accuracy levels using image processing techniques on diverse datasets. Prior research indicates that Bi-CLSTM SVM and CNN models have low accuracy rates, reaching only 65%, highlighting their sensitivity. Nevertheless, with further testing and experience, reproduction analysis accuracy can increase substantially, up to 79%. The OSSGDT method has notably enhanced CVD detection efficiency by significantly improving accuracy rates.

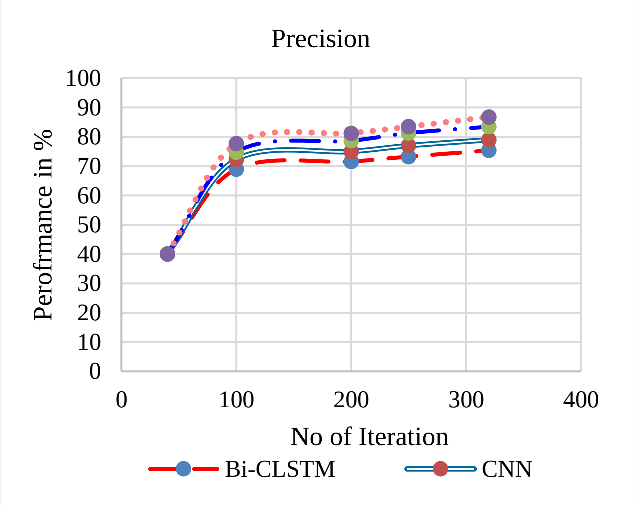


Fig 12. Analysis of Precision

Figure 12 demonstrates the importance of precision performance analysis in predicting CVD accuracy levels through image processing on various datasets. The literature review shows that CNN, Bi-CLSTM, and SVM methods have a low accuracy of 69% in their F1-measure. However, through continued testing and training, the F1-Measure analysis can achieve an accuracy of up to 86.79%. The OSSGDT method especially enhances CVD detection efficiency and improves accuracy significantly.

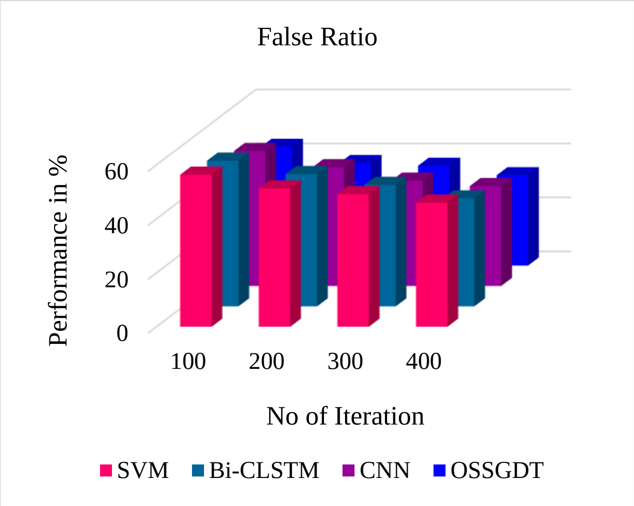


Fig 13. Analysis of False Ratio

The significance of false ratio performance in image processing-based CVD accuracy level prediction is illustrated in Figure 13. According to the literature review, the F1-measure accuracy of SVM, Bi-CLSTM, and CNN approaches is only 56.32%. Nevertheless, the false ratio performance indicates that the proposed OSSGDT method achieves an accuracy of 33.47% with testing and training.

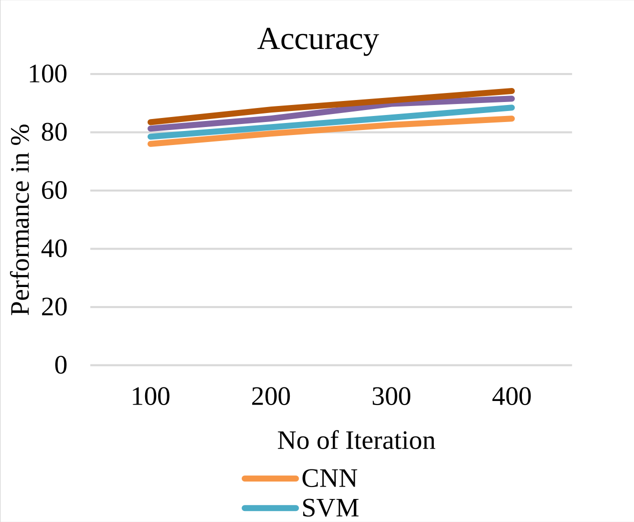


Fig 14. Analysis of Accuracy

Figure 14 demonstrates the analysis of accuracy for predicting high and low-risk factors of CVD in big data. The proposed OSSGDT method resulted in an accuracy of 94.18%, which is a significant improvement. In contrast, methods such as CNN, Bi-CLSTM, and SVM obtained from the literature had lower accuracy ratings of only 76%.

5. Conclusion

In the healthcare field, the use of big data from IoT analytics has become crucial in many important areas. Big data analysis also has the potential to improve CVD prognosis by providing better opportunities for predicting health parameters and producing optimal results. Furthermore, the ACDC dataset, with its extensive collection of MRI image data, has the potential to improve the quality and precision of cardiac disease. In the pre-processing phase, the NLPF technique is used to remove noise and enhance the overall quality of smooth areas in the image. Similarly, the FCCM algorithm is used to iteratively update cluster centres for image segmentation. Moreover, the RFSA approach is used to determine feature weights and select the most relevant features. Furthermore, the ML-based proposed OSSGDT methods can achieve more effective accuracy in predicting cardiovascular diseases. The proposed OSSGDT approach achieves an accuracy of 94.18%, describing a significant improvement.

Reference

[1] Sulagna Mohapatra, Prasan Kumar Sahoo and Suvendu Kumar Mohapatra, "Healthcare Big Data Analysis with Artificial Neural Network for Cardiac Disease Prediction", Electronics 2024, 13(1), 163; https://doi.org/10.3390/electronics13010163.

[2] Mohammad Ayoub Khan, "An IoT Framework for Heart Disease Prediction based on MDCNN Classifier", Thu, 10 Dec 2020 22:00:56 UTC (583 KB), DOI: https://doi.org/10.1109/ACCESS.2020. 2974687.

[3] Rachana Pandey, Monika Choudhary, "Cardiovascular Imaging using Machine Learning: A Review," International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-11 Issue-6, March 2023.

[4] S. Ghorashi et al., "Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases," in IEEE Access, vol. 11, pp. 60254-60266, 2023, doi: 10.1109/ACCESS.2023 .3286311.

[5] K. Saikumar, V. Rajesh, "Heart disease detection based on internet of things data using linear quadratic discriminant analysis and a deep graph convolutional neural network," Front. Comput. Neurosci., 07 October 2022, Volume 16 - 2022 | https://doi.org/10.3389/fncom. 2022. 964686.

[6] Ahmed Ismail, Samir Abdlerazek, I. M. El-Henawy, "Big Data Analytics In Heart Disleases Prediction", Journal of Theoretical and Applied Information Technology, 15th June 2020. Vol.98. No 11.

[7] J. Wang, "OCT Image Recognition of Cardiovascular Vulnerable Plaque Based on CNN," in IEEE Access, vol. 8, pp. 140767-140776, 2020, doi: 10.1109/ACCESS.2020.3007599.

[8] Q. Lyu et al., "Cine Cardiac MRI Motion Artifact Reduction Using a Recurrent Neural Network," in IEEE Transactions on Medical Imaging, vol. 40, no. 8, pp. 2170-2181, Aug. 2021, doi: 10.1109/TMI.2021.3073381.

[9] S. Prabhu, S. Gupta, G. M. Prabhu, A. V. Dhanuka and K. V. Bhat, "QuCardio: Application of Quantum Machine Learning for Detection of Cardiovascular Diseases," in IEEE Access, vol. 11, pp. 136122-136135, 2023, doi: 10.1109/ACCESS.2023.3338145.

[10] K. Zarkogianni, M. Athanasiou, A. C. Thanopoulou and K. S. Nikita, "Comparison of Machine Learning Approaches Toward Assessing the Risk of Developing Cardiovascular Disease as a Long-Term Diabetes Complication," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1637-1647, Sept. 2018, doi: 10.1109/JBHI.2017.2765639.

[11] Chitra Balakrishnan and V. D. Ambeth Kumar, "IoT-Enabled Classification of Echocardiogram Images for Cardiovascular Disease Risk Prediction with Pre-Trained Recurrent Convolutional Neural Networks", Diagnostics 2023, 13(4), 775; https://doi.org/10.3390/ diagnostics13040775.

[12] Hasan, N. I., and Bhattacharjee, A. (2019). Deep learning approach to cardiovascular disease classification employing modified ECG signal from empirical mode decomposition. Biomed. Signal Process. Control 52, 128–140. doi: 10.1016/j.bspc.2019.04.005

[13] A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu and J. A. Steeden, "Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease", *Magn. Reson. Med.*, vol. 81, no. 2, pp. 1143-1156, Feb. 2019.

[14] Morales MA, Assana S, Cai X, Chow K, Haji-Valizadeh H, Sai E, Tsao C, Matos J, Rodriguez J, Berg S, Whitehead N, Pierce P, Goddu B, Manning WJ, Nezafat R. "An inline deep learning based free-breathing ECG-free cine for exercise cardiovascular magnetic resonance," J Cardiovasc Magn Reson. 2022 Aug 11;24(1):47. doi: 10.1186/s12968-022-00879-9. PMID: 35948936; PMCID: PMC9367083.

[15] N. V. MahaLakshmi and R. K. Rout, "Effective heart disease prediction using improved particle swarm optimization algorithm and ensemble classification technique", *Soft Comput.*, vol. 27, no. 15, pp. 11027-11040, Aug. 2023.

[16] M. S.s A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," in IEEE Access, vol. 11, pp. 121574-121591, 2023, doi: 10.1109/ACCESS.2023.3328909.

[17] Y. Pan, M. Fu, B. Cheng, X. Tao and J. Guo, "Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 189503-189512, 2020, doi: 10.1109/ACCESS. 2020.3026214.

[18] S. Abrar, C. K. Loo and N. Kubota, "A Multi-Agent Approach for Personalized Hypertension Risk Prediction," in IEEE Access, vol. 9, pp. 75090-75106, 2021, doi: 10.1109/ACCESS.2021.3074791.

[19] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami and A. Gumaei, "CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings," in IEEE Access, vol. 9, pp. 36955-36967, 2021, doi: 10.1109/ACCESS.2021.3063129.

[20] D. Zhang, X. Liu, J. Xia, Z. Gao, H. Zhang and V. H. C. de Albuquerque, "A Physics-Guided Deep Learning Approach for Functional Assessment of Cardiovascular Disease in IoT-Based Smart Health," in IEEE Internet of Things Journal, vol. 10, no. 21, pp. 18505-18516, 1 Nov.1, 2023, doi: 10.1109/JIOT.2023.3240536.

[21] Dwarakanath B., Latha M., Annamalai R., Jagadish S. Kallimani, Ranjan Walia, Birhanu Belete, "A Novel Feature Selection with Hybrid Deep Learning Based Heart Disease Detection and Classification in the e-Healthcare Environment", Computational Intelligence and Neuroscience, vol. 2022, Article ID 1167494, 12 pages, 2022. https://doi.org/10.1155/2022/1167494.

[22] K. Saikumar, V. Rajesh, Gautam Srivastava, "Heart disease detection based on internet of things data using linear quadratic discriminant analysis and a deep graph convolutional neural network, "Volume 16 - 2022 https://doi.org/10.3389/fncom.2022.964686.

[23] Seung-Jae Lee, Sung-Ho Lee, Jong-Young Lee, "Deep Learning Improves Prediction of Cardiovascular Disease-Related Mortality and Admission in Patients with Hypertension: Analysis of the Korean National Health Information Database", J. Clin. Med. 2022, 11(22), 6677; https://doi.org/10.3390/jcm1122667.

[24] Srinivasan, S., Gunasekaran, S., Mathivanan, S.K. et al. 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

[25] Golande, A.L., Pavankumar, T. Optical electrocardiogram based heart disease prediction using hybrid deep learning. J Big Data 10, 139 (2023). https://doi.org/10.1186/s40537-023-00820-6.

[26] G. Rajkumar, T. Gayathri Devi, A. Srinivasan, "Heart disease prediction using IoT based framework and improved deep learning approach: Medical application," Medical Engineering & Physics, Volume 111, 2023, 103937, ISSN 1350-4533, https://doi.org/10.1016/j.medengphy.2022.103937.

[27] Jameel Ahamed, Abdul Manan Koli, "CDPS-IoT: Cardiovascular Disease Prediction System Based on IoT using Machine Learning" September 2021International Journal of Interactive Multimedia and Artificial Intelligence In Press(In Press):1-9, DOI:10.9781/ijimai.2021.09.002.

[28] Manohar Manur, Alok Kumar Pani, Pankaj Kumar, "A Big Data Analysis Using Fuzzy Deep Convolution Network-Based Model for Heart Disease Classification", International Journal of Intelligent Engineering and Systems, Vol.14, No.2, 2021 DOI: 10.22266/ijies2021.0430.13.

[29] Ashif Newaz Shihab, Miftahul Jannat Mokarrama, "An IoT-Based Heart Disease Detection System Using RNN", January 2021, DOI: 10.1007/978-3-030-51859-2_49.

[30] Adyasha Rath, Debahuti Mishra, Ganapati Panda, Suresh Chandra Satapathy, "Heart disease detection using deep learning methods from imbalanced ECG samples, Biomedical Signal Processing and Control," Volume 68, 2021, 102820, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2021.102820.