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

Early Risk Identification of Cardiac Disease Prediction using Data Mining and Deep Learning Technique

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

Submitted: 03/02/2024 **Revised**: 11/03/2024 **Accepted**: 17/03/2024

Abstract: Data Mining (DM) in cardiovascular data prediction is a rapidly growing field of research. As the amount of cardiovascular data available continues to grow, new and more sophisticated data mining. Most peoples affected by the disease without knowing the feature dependencies to make proper treatment leads more deaths. Artificial intelligence techniques make intelligence feature analysis to predict the disease earlier to support treatment. But most of the prevailing techniques are failed to analyses the disease margins and feature release related to disease factor affects the precision rate, so the prediction accuracy is low and to make improper suggestion. To resolve tis properly a novel optimization is need to improve the prediction accuracy based on the deep learning techniques. DM is utilized to extract valuable information from cardiovascular dataset. In this paper, to propose an enhanced deep featured neural network is designed to analyses the cardiac risk to make efficient prediction and recommendation. The Cross Layer Leap Gated Convolution Neural Network (CLLG-CNN) using Recursive Random Forest Feature Selection (RRFFS) for early risk identification is attained for efficient prediction. The Sparse augmentation disease rate (SADR) finds the ideal margins the feature deficiency factor weight and features are selected using Recursive random forest feature selection (RRFFS). The selected feature are trained with cross layer Leap gated convolution neural network (CLLG-CNN) to find the disease risk factor. The proposed system produce high performance compared to the other system by identifying disease efficiently. This improve the detection rate as well precision recall rate to support from early treatment to avoid the cardiac risks.

Keywords*: Cardiac disease prediction, Data mining, deep neural network, random forest, decision tree, medical margin rate, recommendation system.*

1. Introduction

Cardiovascular disease is a leading cause of death worldwide. Data mining is a powerful tool that can be used to extract knowledge from large datasets of cardiovascular data. This knowledge can be used to improve the diagnosis and treatment of cardiovascular diseases, and to develop new preventive measures. This paper presents data mining for cardiovascular disease prediction which contains the various challenges during data analysis in observing patient data to make analysis with support data mining to classify the risk of cardiovascular disease prediction [1]. These challenges include the need for large datasets of high-quality data, the need for better data mining techniques, and the need for better understanding of the relationship between cardiovascular risk factors and cardiovascular disease.

Despite these challenges, data mining is a powerful tool that can be used to improve the diagnosis, treatment, and prevention of cardiovascular diseases. As the amount of cardiovascular data available continues to grow, data mining will become an increasingly important tool for improving the health of the population.

Cardiovascular disease (CVD) causes damage to the heart and blood vessels, and most of these losses can lead to death or stroke. Therefore early activation of CVD and self-diagnosis can save many lives. However, there is much room for improvement in efficiency and reliability, so more studies can be conducted to achieve these goals. However, millions of people worldwide suffer from chronic diseases due to various factors such as unhealthy lifestyles, work and psychological stress and external factors such as pollution, unsafe working conditions and lack of adequate medical services. Examples include CVD, which can lead to death or disability in the heart and blood vessels.

The human body has various organs, each performing its function. Also, the heart is one of the organs that pump blood throughout the body. Otherwise, the human body is likely to turn into a deadly situation. Hypertension plays a role in various cardiovascular diseases and is a significant cause of Blood Pressure (BP). Machine Learning (ML) implemented in healthcare can detect diseases early with their accuracy. Treatment for patients with chronic diseases, especially CVD, is delayed, and deaths from CVD have increased worldwide. Routine use of BP monitoring is essential for its management and diagnosis and may prevent heart disease. Photoplethysmography (PPG) can be considered a low-

[1]*Research Scholar in Computer Science Thanthai Periyar Govt. Arts & Science College (A), Tiruchirappalli Affiliated to Bharathidasan University*
E-mail: tamilfathima@jmc.edu
[2]*Assistant Professor in Computer Science, Thanthai Periyar Govt. Arts & Science College (A), Tiruchirappalli Affiliated to Bharathidasan*

cost technique to find an effective and convenient method for early diagnosis of CVD. Predicting disease incidence is a critical challenge today. Death due to heart disease has become an essential issue in today's modern life, and every minute, approximately one person suffers from heart disease [2-3].

Blood tests, electrocardiograms, computed tomography scans of the heart, and cardiac magnetic resonance imaging may traditionally be performed to diagnose heart disease. However, these traditional diagnostic methods are invasive and time-consuming and have a high impact. Heart disease is a deadly human disease rapidly increasing worldwide in developed and underdeveloped countries. In general, this disease prevents the heart from supplying enough blood to the rest of the body to function properly. Early detection of this complication is critical to preventing patient harm and saving lives [4-5].

Data mining has been successfully used in medical datasets (large) to extract hidden information and aid decision-making [6-8]. Deep learning is used to efficiently categories the cardiovascular disease. This research main contribution is to improve the CVD identification accuracy performance with less false classification rate. The contribution is to apply a enhanced deep featured neural network that is designed to analyze the cardiac risk and make efficient predictions and recommendations. The Cross Layer Leap Gated Convolution Neural Network (CLLG-CNN) using Recursive Random Forest Feature Selection (RRFFS) for early risk identification is achieved for efficient prediction. The Sparse augmentation disease rate (SADR) finds the optimal margins the feature deficiency factor weight and features are selected using Recursive random forest feature selection (RRFFS). The selected features are trained with cross layer Leap gated convolution neural network (CLLG-CNN) to find the disease risk factor. The rest of the sections follows the literature review and proposed system, result and discussion, finally the conclusion proves the result achievement.

2. Related work

This Section presents a review of the literature on data mining for cardiovascular disease prediction. Also, the different data mining techniques that have been used for cardiovascular disease prediction, and the results of these studies. The review also discusses the challenges and opportunities in using data mining for cardiovascular disease prediction. The author shows the changes in these factors and their impact on Cardiovascular Disease (CVD) in the hourly demand. This requires early disease detection using modern technology to reduce the mortality rate [6]. The author proposed that researchers can wholeheartedly develop smart healthcare systems. An automated system for predicting heart disease risk was undertaken as an outstanding achievement. The UCI Machine Learning (ML) repository dataset can be used to identify predictors of cardiovascular disease assessment [7]. The author proposed that only ten aspects can be considered in the existing studies. In addition, 14 optional components can be taken in their research assignments. Furthermore, a comparative analysis can be provided on CVD classification based on applying Random Forests (RF), Support Vector Machines (SVM), and ML techniques [8]. However, there are a number of challenges that need to be addressed before data mining can be used to its full potential in this area. These challenges include the need for large datasets of high-quality data, the need for better data mining techniques, and the need for better understanding of the relationship between cardiovascular risk factors and cardiovascular disease.

The Prospective Urban-Rural Epidemiology (PURE) model could be a future initiative to assess relevant demographic, behavioral, and CVD risk factors. With these, the data can be sorted by age, urban, and rural. Also, means and standard deviations for continuous variables can be summarized as numbers and percentages for categorical variables [9]. The authors proposed that a sophisticated ML technique, bayesian additive Regression Trees (RT), can rank socio-demographic, health behavior, prevention, and environmental factors [10]. The authors proposed that the Intima-Media Complex (IMC) composition may be more informative for cardiovascular risk assessment than the Intima-Media Thickness (IMD). The main objective was to investigate the textural features of common carotid artery IMCs and their relationship with prevalent clinical CVD. A secondary aim was to test whether the histological features of IMD and IMC differed between left and right carotid arteries [11].

Most authors proposed a time-perceived decay function based on the latent time between patient visits for age and model disease progression patterns. Then, a Bidirectional Long-Short-Term Memory (Bi_LSTM) network and a Convolutional Neural Network (CNN) can be constructed in parallel from various types of clinical data to learn temporal and temporal features [12]. The authors propose to predict Major Adverse Cardiovascular Events (MACE) using a Deep Learning (DL) model. Such a model could be developed and validated based on administrative claims of diabetic patients in the Veneto region of northeastern Italy. Notably, the 4P-MACE endpoint of death, a first event of heart failure, myocardial infarction, or stroke, was assessed using baseline patient information at the 1-year

pharmacy and hospital claims and variable predictive thresholds [13].

The author proposed that CardioXNet could use a new lightweight end-to-end architecture. Five cardiac auscultation classes can be automatically detected using normal, aortic stenosis, mitral stenosis, mitral regurgitation and mitral valve prolapse, and the raw Phono Cardio Gram (PCG) signal [14]. The datamining improving prediction accuracy using ensemble models and majority voting hybrid classifier techniques. In addition, pre-processing methods and feature selection based on genetic algorithms can be used to improve prediction performance and overall time consumption [15]. The authors proposed that four significant cardiac abnormalities can be predicted with the power of DL techniques. They are abnormal heart rate, heart attack, history of heart attack and normal person classes. Also, a dataset of heart patients' general Electro Cardio Gram (ECG) images can be used [16].

The author proposed using a ML based cardiovascular disease diagnosis (MaLCaDD) framework to predict cardiovascular disease with high accuracy effectively. However, improving the accuracy of the proposed techniques and methods is imperative due to the inherent seriousness and life-threatening risks of heart disease [17]. The Data mining method is superior to the Binomial Linear Regression (BLR) model justifying the need to apply more sophisticated methods to produce reliable CVD risk scores [18].

The existing authors proposed a complete end-to-end model called Deep Risk based on focus mechanisms and deep neural networks. It automatically learns high-quality features from Electronic Health Records (EHRs), efficiently synthesizes heterogeneous time-series clinical data, and ultimately predicts a patient's CVD risk [19]. The author uses the consensus of human experts and radiologists to present basic factual information. Performance evaluation with full-field digital mammograms in the analyzed cases was performed using Free-Response Receiver Operating Characteristics (FROC) and calcium mass measurement [20].

The author proposed to use the most comprehensive set of behavioural predictors in the Qatar Biobank (QBB) CVD cohort, including anthropometric, clinical biomarkers, bio impedance, spirometry, VICORDER readouts and biomedical measurements [21]. The author proposed that the data could be filled and analyzed in the Korean National Health Insurance Service-National Health Sample Cohort (KNHSC) collection. Furthermore, the properties of ML and big data can be analyzed to predict CVD risk [22].

Most of the data mining model combining different approaches can effectively predict heart disease. Efficient data collection, pre-processing, and transformation methods can contribute to the model's success in generating accurate information for the training model [23]. The author proposed combining approaches from different models to predict heart disease effectively. Efficient data collection, pre-processing, and transformation methods can enable the model to train the model or generate accurate information successfully [24].

The author proposed to develop a method based on Heart Rate Variability (HRV) features. An ML classifier is offered for the epochal detection of Obstructive Sleep Apnea (OSA) from surface ECG signals [25]. The authors proposed that a stacking ensemble model with a straightforward prediction strategy can predict the number of daily CVD admissions using Hospital Admissions (HAs), air pollution and weather data. Also, the sequential forward floating selection method can be used for Feature Selection (FS) [26]. The authors proposed a multi-input DL model to directly utilise sleep signals and clinical features. They use data from the Pays de la Loire sleep cohort of 5,506 OSA-examined patients without a history of MACE [27].

Further describe an efficient method based on the Synthetic Minority Optimization Technique (SMOTE) to detect patient status by combining six different ML classifiers to deal with the problem of skewed distributions. SMOTE and Joint Hyper Parameter Optimization (HPO) can be proposed to find ML classifiers [28]. The author proposed that the ML approach could potentially use HRV to predict hypertensive patients from those at high risk for developing vascular events [29]. The authors proposed that the developed cardiac patient classifier be applied to a standard 10-s 12-lead ECG signal to identify hypertrophic cardiomyopathy (HCM) patients. Heart rates recorded to classify patients as having HCM can be placed in the characteristics of HCM [30].

A study was conducted using the CRoss-Industry Standard Process for Data Mining (CRISP-DM) to enhance the precision, recall, and accuracy of CVD prediction through optimized DT and RF algorithms [34]. Although the suggested model showed promising results, the author identified a high false classification performance issue. In [35], the fuzzy KNN algorithm was introduced for predicting heart disease by modeling a membership function and combining it with measurements to reduce uncertainty. However, predicting CVD remains an important but challenging task in clinical data analysis [36].

The aim of the study was to determine the key features and effective DM techniques that can enhance the precision of survival prediction in patients with CVD.

Four classification models, namely DT, RF, LR, and SMOTE, were implemented to forecast patient survival [37]. Predicting survival in patients with heart disease is a significant obstacle in clinical data analysis. Despite these challenges, data mining support to male ML and DL that can be used to improve the diagnosis, treatment, and prevention of cardiovascular diseases. As the amount of cardiovascular data available continues to grow, data mining will become an increasingly important tool for improving the health of the population.

3. Proposed system

Towards the development, an enhanced deep featured neural network is designed to analyses the cardiac risk to make efficient prediction and recommendation is carried out by step by step evaluation. But most of the prevailing techniques are failed to analyses the disease margins and feature release related to disease factor affects the precision rate, so the prediction accuracy is low and to make improper suggestion.

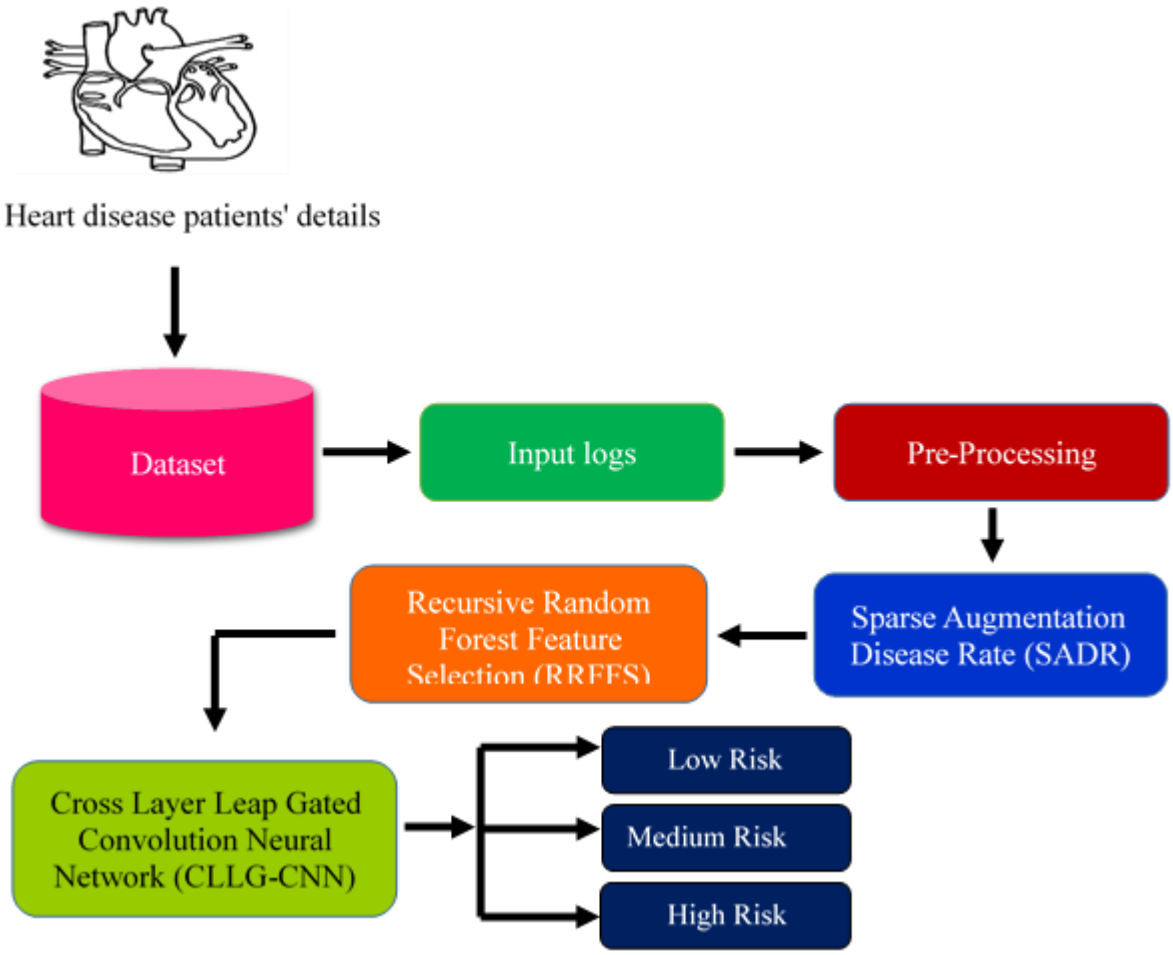

Figure 2: Proposed architecture diagram RRFFS-CLLG-CNN

To resolve tis properly a novel optimization is need to improve the prediction accuracy based on the deep learning techniques. In this paper, to propose an enhanced deep featured neural network is designed to analyses the cardiac risk to make efficient prediction and recommendation. The proposed has following phased to evaluate the disease deficiency.(1)Sparse augmentation disease rate, (2) Recursive random forest feature selection, (3) The cross layer Leap gated convolution neural network, The proposed system produce high performance compared to the other system by identifying disease efficiently. This improve the detection rate as well precision recall rate to support for early treatment to avoid the cardiac risks.

3.1 Dataset collection

The heart disease of patient dataset is applied to evaluate the proposed CNN method. The dataset includes more than and it's collected from various hospitals. Physicians therefore need to have good diagnostic programs to accurately predict heart disease. Collecting data count 2lake in UCI repository datasets link https://archive.ics.uci.edu/dataset/45/heart+disease.

There is an urgent need to develop reliable diagnoses to reduce time to diagnosis and to support the still uncomplicated and increasingly complex diagnostic decision-making processes. It's about improving accuracy. The entire dataset is already classified benign or malignant based on the physician suggestion.

3.2 Dataset preprocessing

In this step, the heart disease dataset records undergo preprocessing to get them ready for feature analysis. The first step is to check whether each record contains all the necessary data values. Next, we perform the filling, removal, and cleaning of checksum and index records to deal with noisy data. The proposed performs all necessary checks to ensure each record is within the defined margins. The proposed method also verifies the existence of all registry values associated with property attributes. Finally, we prepare all records in the dataset to reduce its dimensionality for diagnosing heart disease.

Algorithm

Input: Heart disease dataset (H_D)

Output: Prepared dataset ($\mathfrak{P}_\mathfrak{b}$)

Begin

Step 1: Initialize theH_D

Read collective dataset H_{Di}, $i = 1,2,3 \dots n$

For all record as $H_R \leftarrow H_{D1}, H_{D2} \dots H_{Dn}$

If check each feature ($H_{Di} values$! = $field\ null$)

Check the weight $\mathbb{W} \rightarrow H_{Di}$

Check empty, duplicated records to remove

Update record set

End If

End for

Step 2: Check each feature intensity margins

Calculate the maximum and minimum range $Mx_R$and Mm_R

Retain to index margin values to update

Step 3: For each feature read weights

Compute redundant feature weight list $R_{Fw} = \int_{i=1}^{n} \sum Mm_R$

Step 4: Verify the actual range values

If each feature $Mx_R < \mathfrak{H}_m$
//$\mathfrak{H}_m$ heart disease margin

Obtain the processed records $\mathfrak{P}_\mathfrak{b}$

End if

Stop

The above algorithm steps provide a heart disease-prepared dataset to reduce dimensionality. The proposed system efficiently identifies the inconsistency values presented in the collective dataset.

3.3 Sparse Augmentation Disease Rate (SADR)

This stage finds the ideal margins of disease deficiency factor weight. The intensification variation feature relations are differed from each other feature limits in this time series data limits. So the convocation normal features limits between ideal and actual margins are marginalized through sparse augmentation technique. The minimum convocation of feature difference is evaluated to fins the disease rate weight according the medical margin which support to risk factor identification. This can be calculated as,

.

$$F\ (x) \min_{s \in conv\{X \cup Y\}} ||s - r||^2 \quad (1)$$

Where F(x) is the features from CVD logs A possible approaches for verify whether or not any point of Y belongs to Conv $\{ X \cup Y \}$ to solve the normalization problems. To reduces the computing burden it proceed has the $Y \in C \triangleq conv\ \{X \cup Y\}$

$$Y = \begin{Bmatrix} \sum_{a=1}^{n} \lambda_a\ s_a : s_a \epsilon X \cup Y, \\ \lambda_a \geq 0, x = 1, \dots n, \\ \sum_{a=1}^{n} \lambda_a = 1 \end{Bmatrix} \quad (2)$$

Let, σ_x be the support function of x. For any $X \cup Y$ have$\sigma_x(Y) \geq \sigma_1^*$. λ_a be the Maximum range of the dataset values.

$$\sigma_1^{*\triangleq}\{\min \sigma_x\ (Y), Y \in x\} \quad (3)$$

In this support function observe the nearest values of $\sigma_1^{*\triangleq}$ can obtained by solving form the dataset.

In this stage the attribute features are identified based on ideal margins from the relevance rate which is from finest feature limits. This match the intensity of cardiac rate to select the relevance features.

Input: F_p → created attribute pattern of n feature

Output: f_b → finest features (Ps-Cds)

Start

Import: F_p, S

S → empty set to store and sort out the best features using mutual information.

C
→ mutual information for each feature with a class label

$$S \leftarrow \arg\max (F_p, C)$$

While do for all to process up to intensity margin validation

$I_f = \sum_{i=0}^{n} S$

$f_b = I_f$

End while

Attain Relevance rate Rlr (Fb)→ F(x)

Stop

Related information between each feature to predict the optimal feature f_b from the pattern $F_p$is restrained and examined the article weightf_b. Atlas, the supreme attribute heaviness designs are organized out and made the f_b done restatement

3.4 Recursive Random Forest Feature Selection (RRFFS)

The marginalized feature are trained with In this section, DM based Random Forest (RF) bootstrapping approach is employed to produce training features and samples, and the number of training assemblies is randomly determined. The recursive factor are carried out by decision tree to choose the random factor to normalize the feature weights. Moreover, the versatility and performance variability feature are selected based on nature of the training sets after implementing accurate DT. By the optimization of recursive decision tree and random forest finds the feature limits variation depends on margin related to CVD feature limits without any contradiction. For preference of understanding the individual DT, the final output can be predicted based on most predictions. Additionally, an RF can take predictions from each tree. More trees in the forest, better accuracy, and can be used to avoid over complications. RF consists of multiple DTs, each accomplished with independent predictors and classifiers.

In this category, calculate the total number of individuals in the risk group using a split statistic, (Equation 4).

$$\sum_{R\in M}\sum_{J=1}^{SR}\left[-\hat{\mu}_{Rj}^{M} + y_{Rj} * \log(\hat{\mu}_{RS}^{M})\right] + \sum_{R\in N}\sum_{S=1}^{SR}[-\hat{\mu}_{RS}^{N} + y_{RS} * \log(\hat{\mu}_{RS}^{N})] + \sum_{R\in P}\sum_{S=1}^{Si}[-\hat{\mu}_{RS}^{U} + y_{RS} * \log(\hat{\mu}_{RS}^{U})] \quad (4)$$

In this segment, estimate the expected number of events at various nodes, (Equation 5).

$$\hat{\mu}_{RS}^{z} = \hat{\lambda}_{RS}^{z}\, n_{RS} \quad (5)$$

Compute the Bayesian estimate of each node's occurrence rate if the nodes share the same occurrence rate at a given time, as stated in Equation 6.

$$\hat{\lambda}_{R}^{z} = \frac{\alpha+\sum_{R\in z} y_{RS}}{\beta+\sum_{R\in z} n_{RS}} \quad (6)$$

Compute using Out-of-the-Box (OOB) overall estimation by averaging the process information units. (Equation 7).

$$\hat{\lambda}_{f}^{ooB}(S|x_R) = \frac{\sum_{v=1}^{V} R_{r,v}\hat{\lambda}_{v}(S|a_R)}{\sum_{v=1}^{V} R_{r,v}} \quad (7)$$

Calculate the average value of all the trees in the forest using the new observations (Equation 8).

$$\hat{\lambda}_{f}(S|a_R) = \frac{\sum_{v=1}^{V}\hat{\lambda}_{V}(S|a_R)}{V} \quad (8)$$

In this category, computes the marginal function of a training set drawn randomly from a distribution of random vectors, (Equation 9).

$$wo(a,b) = av_Q R\big(G_Q(a)=b\big) - \max_{S\neq b} av_Q S\big(G_Q(a)=S\big) \quad (9)$$

In this section, the high confidence of the average vote count class measuring the margin is estimated as shown in equation 10.

$$UF^{*} = U_{a,b}(mg(a,b)<0) \quad (10)$$

In this section, calculate the number of trees with bounds on generalization error (equation 11).

$$U_{a,b}\big(U_{\Theta}\big(G(a,\Theta)\big)=b\big) - \max_{S\neq b} U_{\Theta}(U_{\Theta}(G(a,\Theta)=S) < 0) \quad (11)$$

Calculate the marginal function of the random forest and the strength of the classifier ensemble as shown in Equation 12.

$$wN(a,b) = U_{\Theta}(G(a,\Theta)=b) - \max_{S\neq b}(G(a,\Theta)=S) \quad (12)$$

In this category, equation 13 and 14 as shown in the average correlation is calculated.

$$\bar{\rho} = F_{\Theta,\Theta'}\big(\rho(\Theta,\Theta')zx(\Theta)zx(\Theta')\big)/F_{\Theta,\Theta'}\big(zx(\Theta)zx(\Theta')\big) \quad (13)$$

$$F_{\Theta}var(\Theta) \le F_{\Theta}\left(F_{a,b}Nwo(\Theta,a,b)\right)^{2} - z^{2} \le 1 - z^{2} \quad (14)$$

This section evaluated the upper bound of the generalization error (Equation 15).

$$UF^{*} \le \bar{\rho}(1-z^{2})/z^{2} \quad (15)$$

In this segment, estimate the strength number of the classifier ensemble as shown in Equation 16.

$$UF^{*} \le \sum_{S} var\left(\rho_{\Theta}(G(a,\Theta)=b) - \rho_{\Theta}(G(a,\Theta)=S)\right)z_{S}^{2} \quad (16)$$

The maximum number of trees in RF gives high accuracy and high-performance estimation (Equation 17).

$\hat{b} = \frac{1}{t}\sum_{T=1}^{t} f_T\,(a)$

(17)

Let's assume, R & S- individual interval, z-node, M-left node, N, n-right node,- U-parent node, $\hat{\mu}_{R,S}^{z}$-expected number, $\hat{\lambda}_{R}^{z}$-estimate event rate, $\hat{b}$- overall event rate, OOB-out of bag, f-estimate tree, R-time period, $R_{r,V}$-OOb trees, α& β-standard deviation mean value order .and overall event rate, V-Bayes estimate, a and b-random vector, wo-margin function, $G_Q(a)$-training set, av-average number, F-Error, UF-generalization error, UF^{*}-number of sequence error, z-strength, t-average of the output, $f_T(a)$-denoted by decision tree, ρ-rho, Σ-sigma, λ-lambda, Θ-Theta. In this type, averaging the output of multiple decision trees can improve the prediction accuracy of RF classifiers. Thus increasing the number of RF trees results in much better accuracy.

To marginalize the feature limits f(x)→ x1, x2, x3…are Especially to estimate the medical margin limits f(x)>Mw(Medical margin) at new margin N(w) is represented as,

$$N(w) = {}^{n}/_{\sum_{n_{o,n\in f}}^{n} n_0 + n_1 + \cdots + n_x}$$

(18)

From the above equation, the mean count of features is 13, reduced to 8 after feature selection (sex, cb, fps, xang, old peak, slope, ca and thal). $n(w)$ Individual features (W is the weight to calculate the weight of each feature, n denotes the feature. The mutual representation of feature limits are represented as,

$$fw_{(val=x)} = \frac{X}{\mathrm{M} \cup \mathrm{F}}$$

(19)

$$w(tx) = w(n) * w(val)$$

(20)

Feature weights $w(tx)$ and Feature value weights $w(val)$ give the total feature weights($W(n)$).

To create subset index from Sm all heart disease feature limit and default to (Ma)

$$F(Ma) = (\sum fi(Ma)in\,)/Na$$

(21)

Choose the cluster head margin to split the Max- Min groups

$$D\,(F)\,D(F) = |Fs(D(F)) - FM((F))|$$

(22)

Return subset cluster frequency group Fp

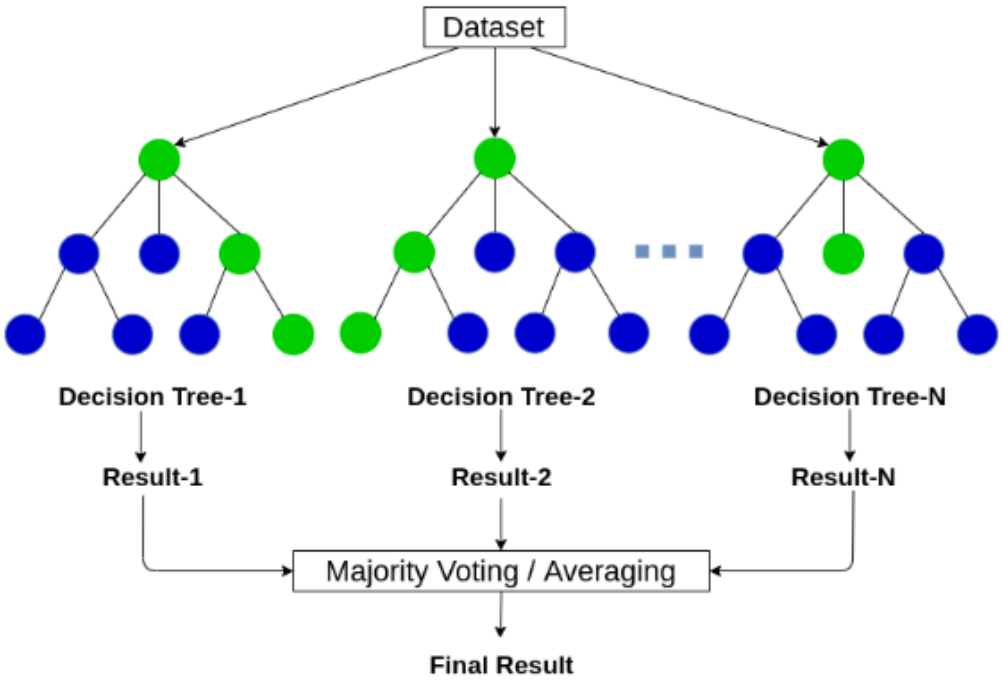

Figure 1. Random Forest

In this section, a DT can be constructed between each model to select an event from the dataset and determine the prediction result as shown in Figure 1. Then, the result with the highest number for each result is referred to as the final prediction

3.5 Cross Layer Leap Gated Convolution Neural Network (CLLG-CNN)

The selected feature are trained with CNNs perform mappings in the middle of spatially / temporally detached arrays in arbitrary dimensions. These

constraints make CNNs perform like interconnected filter systems, and other filter systems can be used to allow valuable assessments. Consequently, the CNN neural weights behave like a Finite Impulse Response (FIR) or bandpass filter plates in the device system. Thus, a trained CNN can be tailored to the application of a specific functional graph to act as a trainable filter system. Finally, CNNs allow the dispensation of large spatially distributed categorizations without the essential for large amounts of free constraints, which increases the possibility of minimal avoidance and generality. Succeeding the embedding layer, a 1D convolutional layer with a kernel size of 2 × 2 and a Rectifier Linear Unit (ReLU) activation function is presented. A maximum pooling layer with a 2 × 2 pooling size is used to map the salient features of the 1D convolution output. Finally, a Flatten layer is added to convert the output back to a 1D array for the ML model. A CNN is "duplicated" based on translations and allows neurons to share weights and biases.

Calculate the magnitude of the network weights of the region around each translation point, as shown in Equation 23.

$$b_{x,Q}^{\mu} = \sum_{L} \sum_{R} Z_{x,L}^{\mu} \, \xi_{L,Q+R}^{\mu} + S_x \qquad (23)$$

In this category, compute the hidden neural outputs that form the array of features introduced by the transfer function (Equation 24).

$$T_{x,Q}^{\mu} = e\left(b_{x,Q}^{\mu}\right) \qquad (24)$$

Equation 25 shows that the neurons in the transformation sequence of the output layer compute the net input.

$$b_{y,c}^{\mu} = \sum_{x} \sum_{D} V_{y,x}^{D} \, T_{x,c+D}^{\mu} + D_y \qquad (25)$$

Computes the filter length of the output layer, as shown in Equation 26.

$$G_{y,c}^{\mu} = e\left(b_{x,Q}^{\mu}\right) = g\left(\sum_{x} \sum_{D} V_{y,x}^{D} \, T_{x,c+D}^{\mu} + D_y\right) \qquad (26)$$

Compute update weights of normal neural networks for spatial translation and feature maps (Equation 27).

$$F = \frac{1}{2} \sum_{\mu,y,c} \left[\zeta_{y,c}^{\mu} - G_{y,c}^{\mu}\right]^2 \qquad (27)$$

Compute the error weights of the sequential concatenation filter on the feature array (Equation 28).

$$\frac{\partial F}{\partial V_{x,y}^{D}} = -\sum_{\mu,c} \left[\zeta_{y,c}^{\mu} - g\left(b_{y,c}^{\mu}\right)\right] e'\left(b_{y,c}^{\mu}\right) T_{x,c+D}^{\mu} \qquad (28)$$

Calculate the familiar gradient-down weight optimization rule shown in equations 29 and 30.

$$\Delta V_{y,x}^{D} = \eta \sum_{\mu,c} \delta_{y,c}^{\mu} \, T_{x,c+D}^{\mu} \qquad (29)$$

$$\delta_{y,c}^{\mu} = \left[\zeta_{y,c}^{\mu} - e\left(b_{y,c}^{\mu}\right)\right] e'\left(b_{y,c}^{\mu}\right) \qquad (30)$$

In this category, compute incremental rules for input weight changes hidden links and index changes (Equation 31 and 32).

$$\Delta z_{x,L}^{D} = \eta \sum_{\mu,Q} \delta_{y,Q}^{\mu} \, \xi_{L,Q,R}^{\mu} \qquad (31)$$

$$\delta_{x,Q}^{\mu} = \sum_{D} e'\left(b_{x,Q}^{\mu}\right) \sum_{y} \delta_{y,Q-D}^{\mu} \, Z_{x,y}^{D} \qquad (32)$$

Convert the training set to the desired input pattern and compute the embedding layer as illustrated in Equation 33 and 34.

$$FW = embedding_{layer(T_D, H_D, Y)} \qquad (33)$$

$$FH_D = FW(E_D) \qquad (34)$$

Embedding layers process input data and produce outputs for further processing of the model, (Equation 35).

$$1D - conv_s = N(E, L_D, IE) \leftarrow FH_D \qquad (35)$$

Calculate using the Rectified Linear Unit (ReLU) activation function to set all negative values to zero (Equation 36).

$$E(I) = max(0, F)_D \qquad (36)$$

Compute the pooling window size used by the maximum pooling layer to map important features, (Equation 37).

$$C_f = E_{map} \left\lfloor \frac{I - P_s}{s} \right\rfloor + 1 \qquad (37)$$

Evaluate the local part of the input in the output given by the recursive expression. (Equation 38 and 39).

$$P_{m+1} = P_m + E_m - 1 \qquad (38)$$

$$P_{m+1} = 2(P_m + E_m) - 3 \qquad (39)$$

Let's assume, x, y, L-index array, G, T,ζ –array label, δ-delta, ζ-zeta, ∂- partial differential, ξ-Xi, c, Q, t-spatial index, D, R- weight index, z-weight, x, Q=hidden units, μ-input pattern, G_c-input contributing to the function, S_x-constant bias, c-neuron translation, D-length of the filter, x- output array, y- feature array, Δ-delta, $V_{y,x}^{D}$-hidden connections, S_x and S_y-bias term, m- layer, E_m −filter size, E_D-feature selection, FH_D-embedding layer, $1D - conv_s$-output of the 1D convolutional layer, IE-activation function, L_D-kernel size, E_{map}-feature map. In this sense, embedding layers process the input data and generate outputs for further processing through the CNN model.

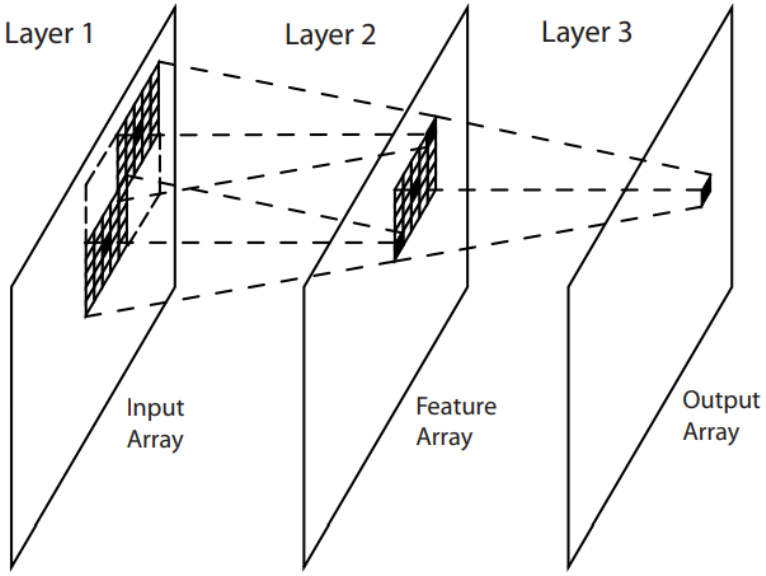

Figure 2. Single Convolutional neuron

In this sense, a convolutional neuron can be used to learn about the structure of a CNN in two layers. As shown in Figure 2, two differently shifted 5x5 filters are mapped to the shaded pixels in the first feature row.

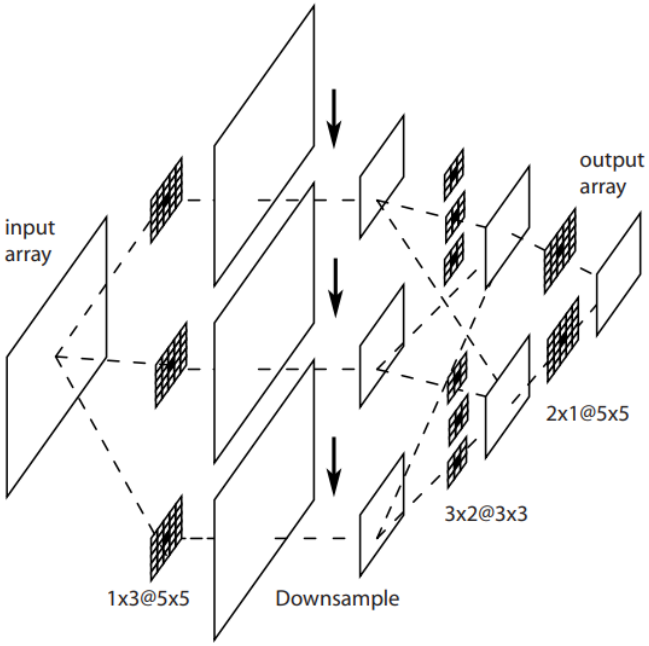

Figure 3. Architecture of a fully interconnected layer

In this section, Figure 3 presents a fully interconnected CNN with a coefficient of 2 in the second layer and several neurons in the structural and convolutional layers. A discriminant function is an activation function that modifies the convolution technique by fixing the consistent beginning loads to the pooling as bias weight. The pooling layer conseques ness are processedas relu activation function,

$$R_t = relu(W_{uv}x_c + W_v R_{c-1} + b_R) \qquad (40)$$

$$x_t = W_{uv} + b_R \qquad (41)$$

The cross layer designed a number of hidden points R is threshold support,

$$R_t = \left(spe\left(x_{t-1}w_{uv}^{(2)}\right) + spe\left(x_{t-1}w_{uv}^{(1)}\right)\right)/2 \qquad (42)$$

Convolutional input features into a recursive input layer using the active support points of the initiation purpose.

$$\frac{\partial x_t^i}{\partial x_{s-1}^n} = \frac{\left(\sum jx_{t-1}^j U_{ji}\right)}{mod\left(\sum jx_{t-1}^j U_{ji}\right)^2}\ \frac{s_{xi}}{2} + \frac{\left(\sum jx_{t-1}^j r_{ji}\ \right)}{mod\left(\sum jx_{t-1}^j r\ \right)^2} \ldots\ \frac{r_{ki}}{2} \qquad (43)$$

This optimized leafs convulsed pods based on threshold margins, maturity and immature intents. This logically activates the threshold average points that segment the classes by their weights for continuous feature support.

Step 1: Choose selected feature F(x).

For each classifier f in x

Using features based intended weights

Marginal ideal values of feature weights using CNN result W_x

End for

Forecasting the Risk level of class

Step 2: CNN → Train a classifier perfect using the relu activation function with values for test data and residual exercise facts.

Step 3: The results of all classification models are predicted using a RCNC approach to improve forecast correctness.

Above every step, forecasts are complete and the exercise data are the predicted outputs of the dataset used for HD classification

Start

Step 1: initialize the feature class Fc(Id)

Step 2: Process to predict the heart disease patient label from dataset Fc(Id)

For each i=0 in id, do

ed→ compute the clause class;

End for

Step 3: compute patient margin ← Fc(Id) at all classes to create support margin.

Step 4: Regularize Ascended attributesX^2

$$X^2 = \frac{x-\text{minimum}(x)}{\text{maximum}(x)-\text{minimum}\,(x)}$$

Step 4: elect the risk classes depends on heart class margin Cd

For each i = 0 in Fc(Id), do then

Select the type definition == heart risk to categorize the class

End for

Stop

The above process classifies the trained cardiology classes based on the logical features of the evaluated training set. A neural classifier predicts recommendations based on disease prevalence for premature treatment based on a course by the representation of disease-affected rate.

4. Result Evaluation and outcomes

Results are evaluate in confusion matrix under various parameters collected in the cardiac database of the UCI repository and its performance was evaluated. This method measures the efficacy against disease prognosis under the feature rate is evaluated. The dataset contains cardiac feature up to 30 features observed from patients which is Figure 2: The collective CVD Dataset features. The results are compared to evaluated using python framework and Jupiter notebook under the feature description. Thus, the method measures performance by various parameters. The expected result observation as shown in figure2 evaluated based on confusion matrix.

id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
1	63	Male	Cleveland	0	125	212	0	1	168	0	1	2	2	3	0
2	67	Male	Cleveland	0	140	203	1	0	155	1	3.1	0	0	3	2
3	67	Male	Cleveland	0	145	174	0	1	125	1	2.6	0	0	3	1
4	37	Male	Cleveland	0	148	203	0	1	161	0	0	2	1	3	0
5	41	Female	Cleveland	0	138	294	1	1	106	0	1.9	1	3	2	0
6	56	Male	Cleveland	0	100	248	0	0	122	0	1	1	0	2	0
7	62	Female	Cleveland	0	114	318	0	2	140	0	4.4	0	3	1	3
8	57	Female	Cleveland	0	160	289	0	0	145	1	0.8	1	1	3	0
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

Figure 3: CVD features dataset

Before being used for classification, the datasets are cleaned and filtered to remove missing or redundant values. The dataset is split into training and testing datasets using 75% and 25% samples. Out of 1,025 patient records, 820 were used for training and used the remaining 205 samples for testing. Training data is used to train the model using the DN-MPNN algorithm, and validation data is used to verify the performance of the trained model shows in figure 3.

Accuracy is the correct intrusion detection based on the total number of intrusion attacks.

$$\text{CVD-disease Classification Accuracy (DCA)} = \frac{C_d}{T(I_g)} \quad (44)$$

Here we assume, C_d denotes correctly detect the CVD pattern and $T(I_g)$ is a total number of usage for recommend generation.

Sensitivity is the probability that the proposed method identified IDS correctly.

Sensitivity $=\frac{TP^2V}{TP^2V+FNV}$ (45)

TP^2V Refers to True positive predictive values, FNV false negative predictive values

The specificity is the probability that the proposed technique rightly estimates a negative.

Specificity$=\frac{FP^2V}{FP^2V+TNV}$ (46)

FP^2V False positive predictive values, TNV refers true negative predictive values False ratio (Fr) $=\frac{T_f+F_t}{R_t}$ Here T_f is a classified as false, F_t is a categorize as true and R_t is a total number of request.

Time complexity measures how long it takes to detect or classify an incoming packet. The O(n) classified at in maximum time taken for processing at in leverage point of data execution. Figure, defines a high impact rate level highlights outputs that detect edges of feature representation that can affect interdependencies of service level integrity rate.

The results and performance of the proposed DN-MPNN for implementation results are tested using the trained features on the UCI repository dataset. Performance evaluations are performed to test the compliance of accuracy and reproducibility of the measurements obtained at this stage. The text case measure is calculated from the implemented error rate true/false state. Performance values are evaluated as precision, recall, and accuracy for a set of positive and negative values trained on tests.

Table 1: Simulation parameters

Parameters	Values
Dataset name	CVD dataset
Simulation tool	Anaconda
Simulation language	Python
Total Records	1000
Training	700
Testing	300

The proposed classifier produces feature categorizing and additional techniques; as shown in Figure 3, the rate of feature classification on the CLLG-CNN is higher than other approaches.

Table 2: Analysis of precision performance

Number of records	EDCNNs in %	LASSO in %	LRCN in %	LIDPF-DNMPMM in %	CLLG-CNN in %
50	46	52	57	70	78
100	52	60	67	75	81
150	61	72	75	82	85
200	70	75	77	84	89
250	75	79	85	88	92
300	80	84	90	92	94

Table 2 describes the precision performance analysis also is known as positive value is the percentage of events that are related. For the problem of unbalanced classification with two classes, the accuracy is divided by the number of true positive and false positives.

$$Precision\ (P)\ =\ TP\ /\ (TP\ +\ FP)\ *100$$

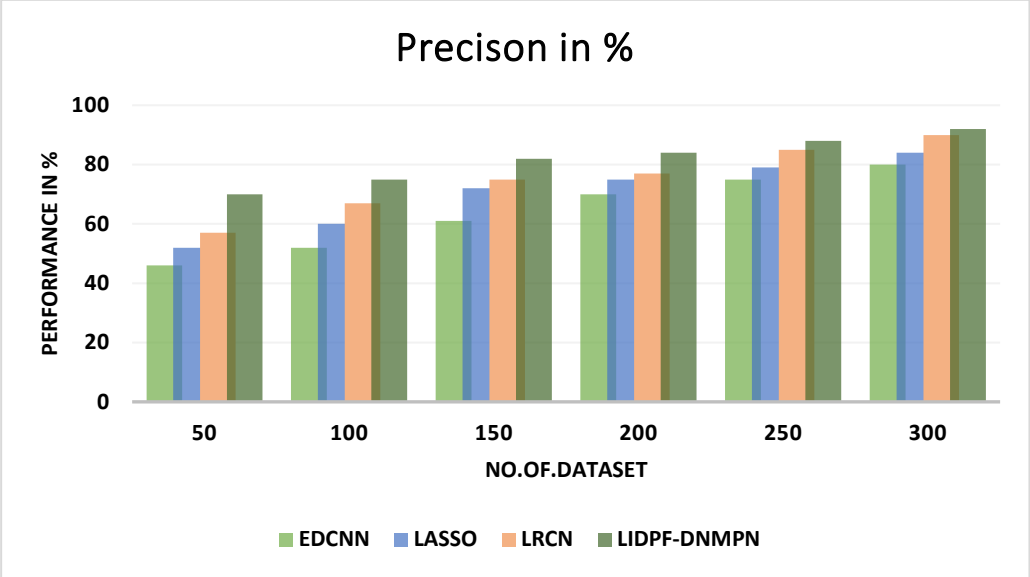

Figure 4: Analysis of precision performance

Figure 4 describes precision values of the True positive accuracy is comparing the different methods, the proposed implementation produces higher performance other Algorithms. In the existing methods Enhanced Deep Learning-Assisted Convolutional Neural Networks (EDCNNs) is 80% and Recurrent Neural Network (LASSO) is 84%, Long short Term Memory (LRCN) is 90% but the proposed method Deep Spectral Recursive Convolutional Neural Classifier (LIDPF-DNMPMM) is 92% is high precision better than previous methods using 300 Records.

Table 3: Analysis of Recall performance

Number of records	EDCNNs in %	LASSO in %	LRCN in %	LIDPF-DNMPNN in %	CLLG-CNN
50	45	50	56	69	76
100	51	60	68	77	81
150	60	70	78	80	85
200	75	77	80	82	87
250	78	80	84	88	92
300	82	85	89	90	94

Table 3 describes the recall performance number of true positives is divisible by the total number of elements that belong to the positive class.

$$Recall\ (R)\ =\ TP\ /\ (TP\ +\ FN) * 100$$

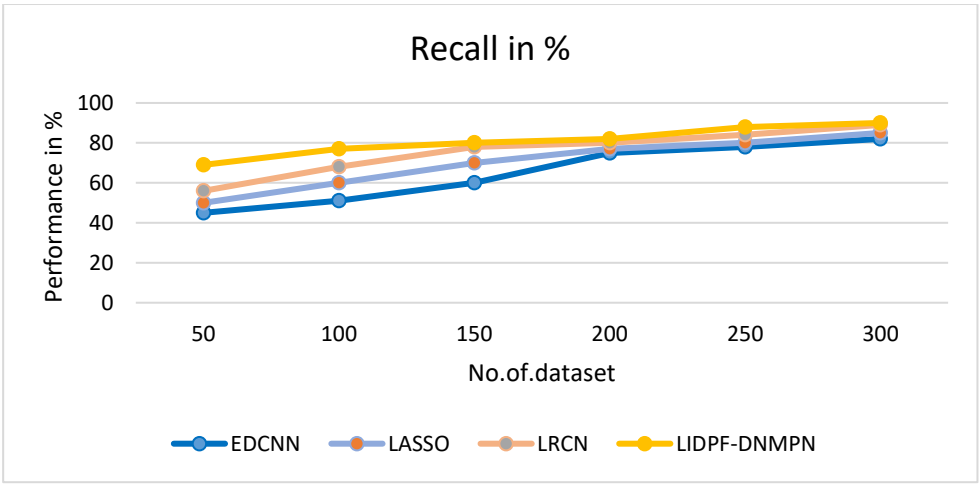

Figure 5: Analysis of recall performance

Figure 5 describes recall performance values for the True positive recalled accuracy is comparing the different methods, the proposed implementation produces higher performance other Algorithms. In the existing methods Enhanced Deep Learning-Assisted Convolutional Neural Networks (EDCNNs) is 82% and Recurrent Neural Network (LASSO) is 85%, Long short Term Memory (LRCN) is 89% but the proposed method Deep Spectral Recursive Convolutional Neural Classifier (LIDPF-DNMPMM) is 90% is high precision better than previous methods.

Table 4: Analysis of Accuracy

Number of records	EDCNNs in %	LASSO in %	LRCN in %	LIDPF-DNMPNN in %	CLLG-CNN
50	50	55	60	65	72
100	57	62	68	73	80
150	66	75	79	84	87
200	75	80	89	91	93
250	78	82	90	92	94
300	82	84	92	95	96

Table 4 describes the generates different levels of users to perform HD features detection accuracy. The proposed system produces a higher impact on prediction performance than other different methods.

$$Accuracy\ (A) = TP\ /\ (TP + TN) * 100$$

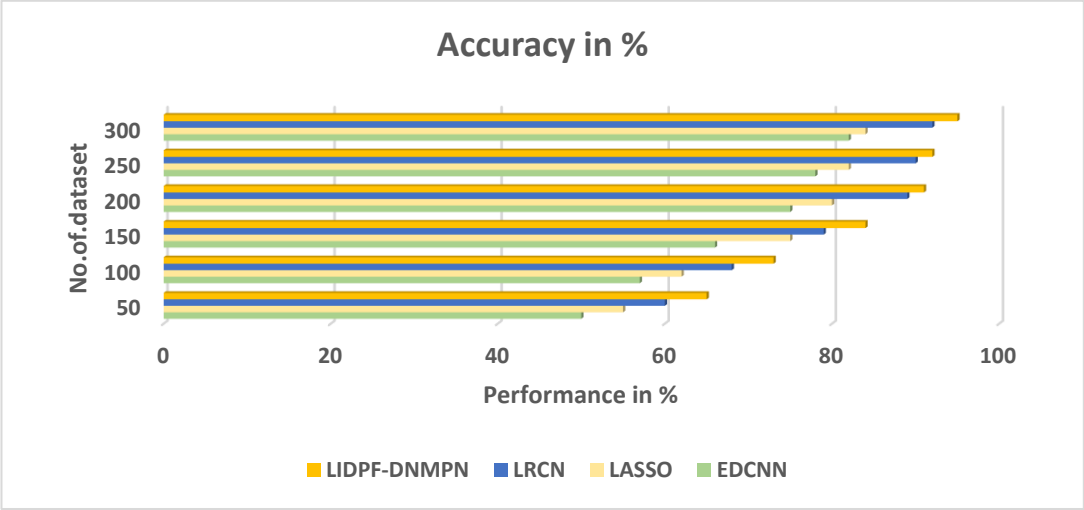

Figure 6: Analysis of accuracy

Figure 6 describes accuracy values for the comparing the different methods, the proposed implementation produces higher performance comparing to other Algorithms. In the existing methods Enhanced Deep Learning-Assisted Convolutional Neural Networks (EDCNNs) is 82% and Recurrent Neural Network (LASSO) is 84%, Long short Term Memory (LRCN) is 92% but the proposed method Deep Spectral Recursive Convolutional Neural Classifier (LIDPF-DNMPMM) is 95% is high precision better than previous methods.

Table 5: Analysis of False score

Number of records	EDCNNs in %	LASSO in %	LRCN in %	LIDPF-DNMPNN in %	CLLG-CNN
50	36	34.2	31.1	28.4	22.6
100	41.1	39.2	36.2	30.1	23.9
150	45.2	40.2	41.3	32.4	24.4
200	50.2	49.2	47.4	34.6	25.2
250	55.2	50.1	49.2	45.2	28.7
300	59.2	47.2	45.2	42.8	32.4

Table 5 describes false rate performance comparing the different levels of HD features for reducing the errors. The proposed methods reducing the errors for the records both train and test dataset.

$$F - score\ (F)\ =\ (2\ *\ Precision\ (P)\ *\ Recall\ (R))\ /\ (Precision\ (P)\ +\ Recall\ (R))$$

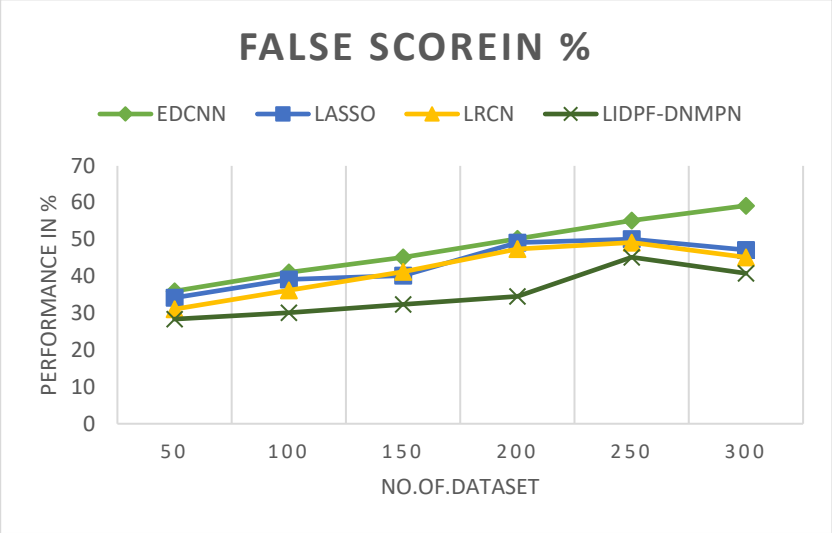

Figure 7: Analysis of False rate

As shown in Table 5. The percentages of misclassifications introduced by various methods have been measured, as shown in Table 5. The misclassification rate of the proposed CLLG-CNN method is lower than that of other methods, which is 4%.Recurrent Neural Network (LASSO) is 59.2%, long short Term Memory (LRCN) is 45.2% but the proposed method Deep Spectral Recursive Convolutional Neural Classifier (LIDPF-DNMPMM) is 40.8% .The disease prediction rate defines the high level performance levels in various levels of feature description to predict best level as shown in figure 3. The impact of this HSFS-DCNN method best predicted in feature evaluation and higher classification compared to the other methods up to 96 %. This proposed system proves the best performance compared to the prevailing methods

5. Conclusion

To conclude, cardiac disease prediction based on extensive data analysis cross layer Leap gated convolution neural network using Recursive random forest feature selection for early risk identification has produced the best prediction accuracy. DM techniques can efficiently extract useful information related to CVD. Therefore, the proposed system analyses the SADR with RF model to select the cardiac features depending on marginal accuracy. It reduces the big data dimensionality problems, which makes features trained on recurrent neural networks for best training features. The resultant factors prove that the best classification accuracy is achieved concerning cardiac disease influence rate, up to 97 %. It supports a better way to make early risk disease predictions for premature treatment to reduce cardiac attack

Reference

[1] Pal M, Parija S, *Panda* G, Dhama K, Mohapatra RK. Risk prediction of cardiovascular disease using machine learning classifiers. Open Med (Wars). 2022 Jun 17;17(1):1100-1113. doi: 10.1515/med-2022-0508. PMID: 35799599; PMCID: PMC9206502.

[2] Dalin Zhang" Detection of Cardiovascular Disease Based on PPG Signals Using Machine Learning with Cloud Computing" Volume 2022 | Article ID 1672677 | https://doi.org/10.1155/2022/1672677.

[3] Apurv Garg et al "Heart disease prediction using machine learning techniques" 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012046 DOI 10.1088/1757-899X/1022/1/012046.

[4] Gihyun Lee "Effective Heart Disease Detection Based on Quantitative Computerized Traditional Chinese Medicine Using Representation Based Classifiers" Volume 2017 | Article ID 7483639 | https://doi.org/10.1155/2017/7483639.

[5] Muhammad, Y., Tahir, M., Hayat, M. et al. Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci Rep 10, 19747 (2020). https://doi.org/10.1038/s41598-020-76635-9.

[6] N. K. S. Banu and S. Swamy, "Prediction of heart disease at early stage using data mining and big data analytics: A survey," 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), Mysuru, India, 2016, pp. 256-261, doi: 10.1109/ICEECCOT.2016.7955226.

[7] Kirmani M. M. Cardiovascular Disease Prediction Using Data Mining Techniques: A Review. Orient. J. Comp. Sci. and Technol, Vol. 10, No. (2), pp. 520-528, 2017.

[8] R Fadnavis, K Dhore, D Gupta, J Waghmare and D Kosankar, “Heart disease prediction using data mining”, International Conference on Research Frontiers in Sciences (ICRFS), 2021, pp.1-7.

[9] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques", *ICT Exp.*, 2021, [online] Available: https://doi.org/10.1016/j.icte.2021.08.021.

[10] K. Dissanayake and M. G. Md Johar, "Comparative study on heart disease prediction using feature selection techniques on classification algorithms", *Appl. Comput. Intell. Soft Comput.*, vol. 2021, 2021, [online] Available: https://doi.org/10.1155/2021/5581806.

[11] P. E. Rubini, C. A. Subasini, A. V. Katharine, V. Kumaresan, S. G. Kumar and T. M. Nithya, "A cardiovascular disease prediction using machine learning algorithms", *Ann. Romanian Soc. Cell Biol.*, vol. 25, no. 2, pp. 904-912, 2021.

[12] K. F. Alhabib, M. A. Batais, T. H. Almigbal, M. Q. Alshamiri, H. Altaradi, S. Rangarajan, et al., "Demographic behavioral and cardiovascular disease risk factors in the Saudi population: Results from the prospective urban rural epidemiology study (PURE-Saudi)", *BMC Public Health*, vol. 20, no. 1, pp. 1-14, Dec. 2020.

[13] L. Hu, B. Liu and Y. Li, "Ranking sociodemographic health behavior prevention and environmental factors in predicting neighborhood cardiovascular health: A Bayesian machine learning approach", *Preventive Med.*, vol. 141, Dec. 2020.

[14] C. P. Loizou, E. Kyriacou, M. B. Griffin, A. N. Nicolaides and C. S. Pattichis, "Association of Intima-Media Texture With Prevalence of Clinical Cardiovascular Disease," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 68, no. 9, pp. 3017-3026, Sept. 2021, doi: 10.1109/TUFFC.2021.3081137.

[15] Y. An, K. Tang and J. Wang, "Time-Aware Multi-Type Data Fusion Representation Learning Framework for Risk Prediction of Cardiovascular Diseases," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 6, pp. 3725-3734, 1 Nov.-Dec. 2022, doi: 10.1109/TCBB.2021.3118418.

[16] E. Longato, G. P. Fadini, G. Sparacino, A. Avogaro, L. Tramontan and B. Di Camillo, "A Deep Learning Approach to Predict Diabetes' Cardiovascular Complications From Administrative Claims," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3608-3617, Sept. 2021, doi: 10.1109/JBHI.2021.3065756.

[17] 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.

[18] S. E. A. Ashri, M. M. El-Gayar and E. M. El-Daydamony, "HDPF: Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm," in IEEE Access, vol. 9, pp. 146797-146809, 2021, doi: 10.1109/ACCESS.2021.3122789.

[19] M. B. Abubaker and B. Babayiğit, "Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods," in IEEE Transactions on Artificial Intelligence, vol. 4, no. 2, pp. 373-382, April 2023, doi: 10.1109/TAI.2022.3159505.

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

[21] 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.

[22] Y. An, N. Huang, X. Chen, F. Wu and J. Wang, "High-Risk Prediction of Cardiovascular Diseases via Attention-Based Deep Neural Networks," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 3, pp. 1093-1105, 1 May-June 2021, doi: 10.1109/TCBB.2019.2935059.

[23] J. Wang et al., "Detecting Cardiovascular Disease from Mammograms With Deep Learning," in IEEE Transactions on Medical Imaging, vol. 36, no. 5, pp. 1172-1181, May 2017, doi: 10.1109/TMI.2017.2655486.

[24] H. R. H. Al-Absi, M. A. Refaee, A. U. Rehman, M. T. Islam, S. B. Belhaouari and T. Alam, "Risk Factors and Comorbidities Associated to Cardiovascular Disease in Qatar: A Machine Learning Based Case-Control Study," in IEEE

Access, vol. 9, pp. 29929-29941, 2021, doi: 10.1109/ACCESS.2021.3059469.

[25] G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in IEEE Access, vol. 8, pp. 157643-157653, 2020, doi: 10.1109/ACCESS.2020.3015757.

[26] P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in IEEE Access, vol. 9, pp. 19304-19326, 2021, doi: 10.1109/ACCESS.2021.3053759.

[27] D. Xu, J. Q. Sheng, P. J. -H. Hu, T. -S. Huang and C. -C. Hsu, "A Deep Learning–Based Unsupervised Method to Impute Missing Values in Patient Records for Improved Management of Cardiovascular Patients," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 2260-2272, June 2021, doi: 10.1109/JBHI.2020.3033323.

[28] D. Padovano, A. Martinez-Rodrigo, J. M. Pastor, J. J. Rieta and R. Alcaraz, "On the Generalization of Sleep Apnea Detection Methods Based on Heart Rate Variability and Machine Learning," in IEEE Access, vol. 10, pp. 92710-92725, 2022, doi: 10.1109/ACCESS.2022.3201911.

[29] Z. Hu, H. Qiu, Z. Su, M. Shen and Z. Chen, "A Stacking Ensemble Model to Predict Daily Number of Hospital Admissions for Cardiovascular Diseases," in IEEE Access, vol. 8, pp. 138719-138729, 2020, doi: 10.1109/ACCESS.2020.3012143.

[30] M. Blanchard, M. Feuilloy, A. Sabil, C. Gervès-Pinquié, F. Gagnadoux and J. -M. Girault, "A Deep Survival Learning Approach for Cardiovascular Risk Estimation in Patients With Sleep Apnea," in IEEE Access, vol. 10, pp. 133468-133478, 2022, doi: 10.1109/ACCESS.2022.3231743.

[31] Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," in IEEE Access, vol. 10, pp. 79974-79985, 2022, doi: 10.1109/ACCESS.2022.3191669.

[32] M. Alkhodari, D. K. Islayem, F. A. Alskafi and A. H. Khandoker, "Predicting Hypertensive Patients With Higher Risk of Developing Vascular Events Using Heart Rate Variability and Machine Learning," in IEEE Access, vol. 8, pp. 192727-192739, 2020, doi: 10.1109/ACCESS.2020.3033004.

[33] Q. A. Rahman, L. G. Tereshchenko, M. Kongkatong, T. Abraham, M. R. Abraham and H. Shatkay, "Utilizing ECG-Based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification," in IEEE Transactions on NanoBioscience, vol. 14, no. 5, pp. 505-512, July 2015, doi: 10.1109/TNB.2015.2426213.

[34] Bárbara Martins, Diana Ferreira, Cristiana Neto, António Abelha, José Machado "Data Mining for Cardiovascular Disease Prediction" Journal of Medical SystemsVolume 45Issue 1Jan 2021https://doi.org/10.1007/s10916-020-01682-8.

[35] Krishnaiah, V., Narsimha, G., Chandra, N.S. (2015). Heart Disease Prediction System Using Data Mining Technique by Fuzzy K-NN Approach. In: Satapathy, S., Govardhan, A., Raju, K., Mandal, J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer, Cham. https://doi.org/10.1007/978-3-319-13728-5_42.

[36] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[37] Ishaq et al., "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," in IEEE Access, vol. 9, pp. 39707-39716, 2021, doi: 10.1109/ACCESS.2021.3064084.