Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7685512

ROBUST REGRESSED CONGRUENCE CONVOLUTIVE DEEP LEARNING CLASSIFIER FOR HEART DISEASE PREDICTION WITH THE CAUSE OF DIABETES

[1]S.M.Meenaatchi [2]Dr.K.Rajeswari
PhD Research Scholar[1], Associate Professor [2]
PG and Research Department of Computer Science,
Tirppur Kumaran College for Women, Tirppur, Tamil Nadu, India [1,2]
Email id :rajeswarik1224@gmail.com

Abstract

Prediction of heart disease is a major issue to begin and personalize the treatments and maximize the benefit of those treatments. Heart disease is a major noticeable illness that increases heart disease by influencing several risk factors faced by the people such as age, high blood pressure, cholesterol, blood sugar, etc. The risk of diabetes increases constantly with the increasing nature of glucose levels and it causes heart disease. The conventional prediction model is a great deal for predicting heart disease with the cause of diabetes at an earlier stage. But achieving a higher degree of accuracy rates using conventional algorithms and feature selection methods is a difficult task. In order to improve the accuracy of heart disease digenesis, a novel deep learning technique called Hot Deck Imputed Robust Congruence Convolutive DEep Neural Learning (DICDEN) technique is introduced. The proposed DIRCOL technique consists of three different processes namely data-preprocessing, feature selection, and classification. The number of patient data related to diabetes is collected from the dataset. Then the data preprocessing is carried out using Manhattan hot deck imputed technique to transform raw data into the structured format by handling the missing value and data duplication. After the data preprocessing, the Rand indexive robust linear regression is employed in the proposed DICDEN technique for significant feature selection and removing the other irrelevant features. Followed by, classification of data samples is done with the selected features by using a deep learning technique called Tucker Congruence Radial Damped Convolutive Deep Learning Classifier. The proposed classifier includes numerous layers such as one input layer, multiple hidden layers, and one output layer. First, the number of selected features with the training patient data is given to the input layer. Then the input is transferred into the hidden layer where the feature mapping is performed in the convolution layer using the Tucker congruence correlation coefficient. The radial activation function is applied to provide the final disease classification results at the output layer. In order to minimize the error, the damped least square method is applied. Finally, accurate classification results with a minimum error are obtained at the output layer. Based on the classification results, heart disease is correctly predicted. Experimental evaluation is carried out with different quantitative metrics such as disease prediction accuracy, precision, recall, F-measure, and disease prediction time. The analyzed results reveal the performance of our proposed DICDEN technique when compared with the existing deep learning methods.

Keywords: Heart disease predicting, Manhattan hot deck imputed technique based data preprocessing, Rand indexive robust linear regression based feature selection, Tucker Congruence Radial Damped Convolutive Deep Learning Classifier

1. INTRODUCTION

A huge amount of medical data is available in the healthcare industry and this data is used by the physician to make precise decisions about patients' health conditions. Heart disease is one of the major chronic problems affecting human health. Therefore, an efficient intelligent system is required to extract medical data from the medical organization for accurate disease diagnosis. Data mining techniques have a massive effect on extracting information from a data set and predicting the disease of humans. A lot of new technologies have been developed for predicting Heart disease.

A Combined Reinforcement Multitask Progressive Time-Series Network (CRMPTN) was developed in [1] to detect coronary heart disease through echocardiography reports, and blood biochemical indicators about the patients. However, accurate heart disease prediction was not performed with minimum time. A deep belief network with a cuckoo search bio-inspired algorithm (DBF+CSA) was introduced in [2] for finding the accurate prediction of cardiac disease. But it failed to combine the huge datasets with deep learning classification methods for cardiac disease prediction. A filter-based feature selection technique was introduced in [3] to select the most relevant features for identifying heart disease. However, the prediction accuracy was not improved since it failed to apply a combination of machine learning and deep learning models to obtain the best feasible model for heart disease diagnosis. Recursion-enhanced random forest with an improved linear model (RFRF-ILM) was developed in [4] to identify heart disease based on key features. However, the deep learning model was not applied for accurate heart disease prediction. The performance of risk prediction scores for coronary heart disease and type-2 diabetes was developed in [5]. But the performance of higher risk prediction accuracy was not improved. An analysis of the random forest model was designed in [6] for detecting cardiovascular diseases. But the preprocessing and significant feature selections were not performed to enhance the performance of cardiovascular diseases prediction.

Extreme Gradient Boosting classifier was introduced in [7] to improve the prediction accuracy of heart disease. But the deep learning classifier model was not applied for enhancing the prediction performance. An accurate fuzzy rule-based classification method was developed in [8] for predicting heart disease. However, it failed to optimize the accuracy of classification for improving the prediction results. A two-deep neural network classifier was developed in [9] to build more accurate prediction models for identifying the risk of coronary heart disease. But it failed to focus on handling missing values by generating new values. A hybrid classifier using the ensembled model was developed in [10] to improve prediction accuracy by using preprocessing techniques and feature selection and minimize the overall time consumption. However, the deep learning model was not efficient to improve prediction accuracy.

1.1 Major contributions of the paper

In this section, a novel DICDEN technique is introduced with the following novel contributions,

- To predict the effectiveness of diabetes-based heart disease, the DICDEN technique is introduced based on three different processes namely preprocessing, feature selection, and classification.
- To minimize the heart disease prediction time, the Manhattan hot deck imputed technique is applied for handling the missing value and data duplication. Then the Rand indexive robust linear regression is also applied for selecting the more significant features for predicting diabetes-based heart disease. Robust linear regression is a machine learning technique used for finding the more related features with help of the rand similarity index.
- A Tucker Congruence Radial Damped Convolutive Deep Learning Classifier is applied in DICDEN for predicting heart disease by analyzing the testing and training samples by using the Tucker congruence correlation coefficient. The radial activation function is also applied for providing the classification result. The damped least square method is also used to minimize the error of the classification result.
- Finally, a comprehensive experimental assessment is carried out with a variety of performance metrics to illustrate the improvement of the DICDEN technique over conventional deep learning methods.

1.2 Outline of paper

The rest of the paper is arranged into different sections. Section 2 reviews the related works of heart disease prediction. Section 3 provides a brief explanation of the proposed DICDEN technique with a neat architecture diagram and different processes. Section 4 describes the experimental settings with the dataset description. In section 5, the performance results of the proposed DICDEN technique and conventional deep learning methods are discussed with different metrics. At last, Section 6 concludes the paper.

2. RELATED WORKS

A novel hybrid technique was introduced in [11] to predict the accuracy level of prediction. But it failed to apply the proposed model to different machine learning algorithms for different metrics comparison. Supervised machine learning algorithms were introduced in [12] for heart disease prediction to perform comparison analysis.A Swarm-Artificial Neural Network (Swarm-ANN) was developed in [13] to predict heart disease. However, the feature fusion and selection methods were not used for increasing the accuracy of heart disease prediction. Unsupervised deep learning–based method was introduced in [14] to improve the prediction of cardiovascular patients. However, the complexity of cardiovascular disease prediction was not minimized. Machine learning algorithms with lasso feature selection methods were developed in [15] for predicting cardiovascular disease. However, it failed to apply deep learning algorithms for improving cardiovascular disease prediction.

An explainable transformer-based deep learning method was introduced in [16] for identifying the occurrence of heart failure. The model and analysis were not applied more deeply for the discovery of heart disease in other complex conditions. An ensemble-based voting method was introduced in [17] for predicting heart disease. But the method was not efficient for analyzing diabetes-based heart disease prediction. IoT-centered Deep Learning Modified Neural Network

(DLMNN) was introduced in [18] for heart disease diagnosis. But the heart disease was not identified more accurately for rendering the treatment to the patient immediately by the doctor. An ensemble-based method was developed in [19] includes machine learning (ML) and deep learning (DL) to identify cardiovascular disease. A novel recurrent neural network framework was developed in [20] for predicting heart failure. But the accurate prediction of heart failure patients at high risk was not performed.

3. METHODOLOGY

Medical organizations and hospitals are generating a huge volume of data on daily basis. Due to the massive amount of medical data in the healthcare industry, a clinical system analyzes and identifies the patient's health conductions and minimizes medical errors. In healthcare data, predicting heart disease is an extremely complex process. It is done only if the doctor with well-experienced and has good awareness concerning the disease. The clinical system analyzes a set of attributes for heart disease prediction of the patients but the irrational factors expose to time delay and made uncertainty in the heart disease prediction. Based on this motivation, a novel technique called DICDEN is introduced in this paper for improving heart disease prediction. The objective of the proposed DICDEN is to improve the accuracy of heart disease prediction.

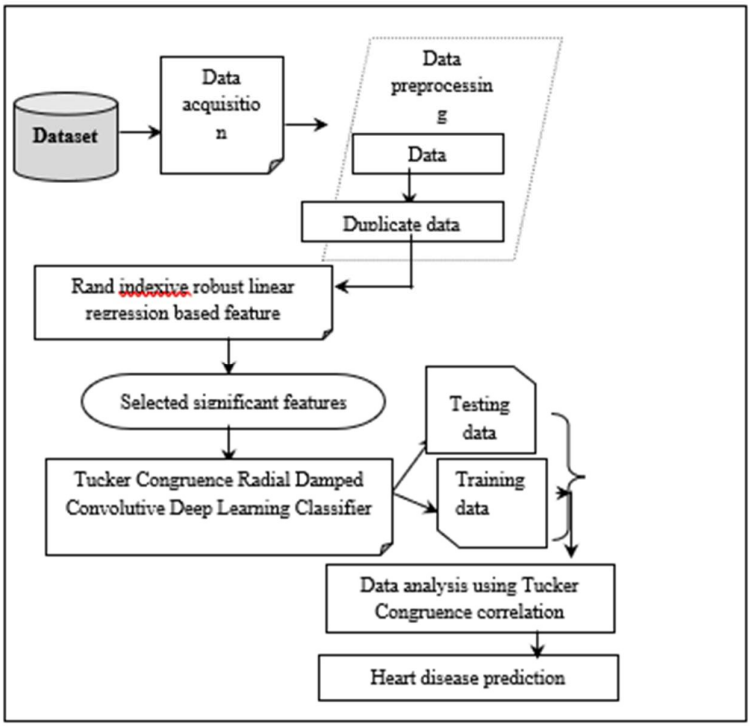

Figure 1 Architecture of the proposed DICDEN technique

Figure 1 given above depicts the architecture of the proposed DICDEN technique consisting that includes three major processes for improving heart disease prediction accuracy. Initially, the Heart Disease Dataset is considered as input for disease prediction. The dataset consists of 'n' number of patient data $D_i = D_1, D_2, .., D_n$' and 'm' number of attributes '$A_j = A_1, A_2, .., A_m$' such as age, gender or sex, chain pain type, resting blood pressure, serum cholesterol in mg/dl, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, old peak = ST depression induced by exercise relative

to rest, the slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, thal, and target.

After the data acquisition process, pre-processing is carried out to structure the dataset. Data preprocessing is necessary for any machine learning or deep learning approach since the performance of machine learning methodology depends on the dataset that is prepared and structured. The proposed DICDEN technique uses the Manhattan hot deck imputed data preprocessing method to handle the missing data and remove the outlier or duplicate data to show the model's efficiency and obtain adequate and better accuracy for disease prediction. In the pre-processing stage, there are two processes are included such as handling missing values and data duplication removal

After the data pre-reprocessing, Rand indexive robust linear regression-based Feature selection is performed in the DICDEN technique to select the relevant attributes from the heart disease dataset for timely disease prediction with minimum complexity. Robust linear regression is a machine learning technique used for finding well-matched attributes with help of the statistical method called the Rand similarity index. With the extracted significant features from the training patient data, Tucker Congruence Radial Damped Convolutive Deep Neural Learning Classifier is applied for predicting heart disease. First, the data is split into testing and training samples. The Convolutive Deep Neural Learning Classifier consists of different layers for learning the given input. The convolution layer uses the Tucker Congruence correlation coefficient to perform the feature map by analyzing the patient training data and testing data. Then the max-pooling layer minimizes the dimensionality of the data by means of the Wilcoxon rank-sum test. It is a statistical measure used for finding stochastically correlated results greater than the other by setting the threshold value. Then the radial activation function is used in a fully connected layer for analyzing the correlation results with the mean of the particular class and provides the output classification results as normal or heart disease at the output layer. This different process of the proposed DICDEN technique is explained in the following subsections.

3.1 Data preprocessing

The fundamental process of the proposed DICDEN technique is the data preprocessing that is used to transform raw data into a structural format. The raw data typically has conflicting data formats, and also be incomplete. Therefore, the data preprocessing step resolves the above-said issues and creates a dataset more absolute and efficient to perform disease prediction. The data preprocessing in the proposed DICDEN technique is performed by filling in the missing value and duplicate data removal.

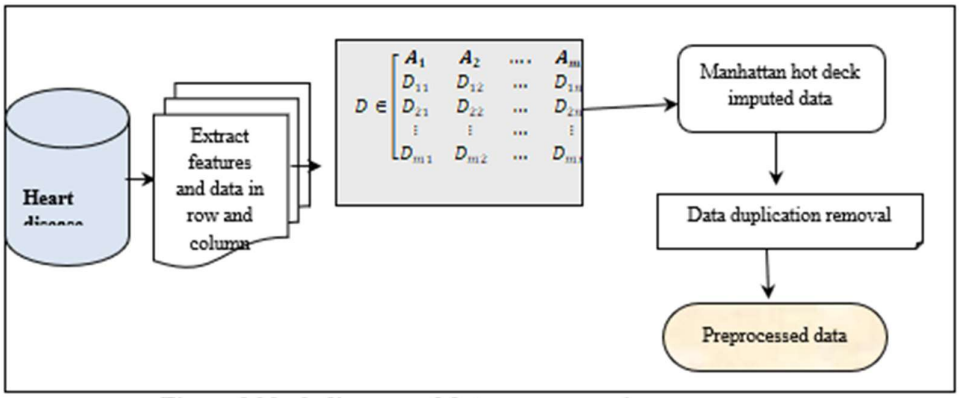

Figure 2 block diagram of data pre-processing

Figure 2 demonstrates dataset pre-processing to obtain the structured output. First, a number of features and data are arranged in the form of a sample input matrix with respect to rows and columns. Second, data imputation is performed for handling the missing data in the corresponding column and row by applying a Manhattan hot deck imputed data preprocessing. Followed by, the data duplication is identified and removed. Finally, the structured dataset is obtained for the next processing.

Let us consider the 14 attributes ''$A_j = A_1, A_2, .., A_m$ are arranged in the row of the matrix and the corresponding patient data samples $D_i = D_1, D_2, .., D_n$' are positioned in the column of the matrix. Then, the input sample matrix 'X' is formulated as given below.

$$X = [A_1\ A_2\ \ldots A_m\ D_{11}\ D_{12}\ \ldots\ D_{1n}\ D_{21}\ D_{22}\ \ldots\ D_{2n}\ \vdots\vdots\ \ldots\ \vdots\ D_{m1}\ D_{m2}\ \ldots\ D_{mn}] \quad (1)$$

From the above formulation (1), the matrix ' ' is constructed with the number of attributes and the corresponding data. The data imputation technique is applied for handling the missing value in the dataset. In statistics, data imputation is the process of substituting the missing data with alternative values. When substituting a data point into the respective column, it is called unit imputation. By applying the Manhattan distance measure, the data value is filled with randomly selected and validated.

$$d = \sum \quad |D_{ij} - D_{ik}| \quad (2)$$

Where, denotes a Manhattan distance, denotes a data value stored in row, column, denotes a data value stored in row, column. A condition which satisfies to fill the new randomly number into the respective column as given below,

$$Z = min\ d \quad (3)$$

Where, denotes an output of preprocessing, denotes a minimum distance ' '. Similarly, data duplication is performed depending on the large deviation between the values of the data. It means that the two data values of the respective column have maximum deviation.

$$B = max\ d \quad (4)$$

Where B denotes an output of duplication removal, max denotes a maximum distance function 'd'. As a result, the data imputation and missing values are handled, and obtain the preprocessed data. The algorithmic process of the data preprocessing is given below,

// Algorithm 1: Data Pre-processing
Input: Heart disease dataset, attributes $A = A_1, A_2, .., A_m$', $D_i = D_1, D_2, .., D_n$
Output: Obtain the pre-processed data

Begin
1: Collect the number of attributes $A_1, A_2, .., A_m$ and data $D_i = D_1, D_2, .., D_n$ from the dataset
2: **For** each data in the dataset
3: Construct the matrix 'X'
4: Measure the Manhattan distance 'd'
5. Filling the missing value by finding the minimum distance (2)
6. Remove duplicate data using (4)
7. **End for**
8. Return (preprocessed data)
End

Algorithm 1 given above provides the different steps of data pre-processing. The input data samples and attributes are collected from the heart disease dataset. First, the data imputation process is said to be performed for filling the missing value with minimum distance. Duplicate data are identified and removed from the dataset. As a result, the data pre-processed results are obtained to minimize the complexity of heart disease prediction.

3.2 Rand indexive robust linear regression-based Feature selection

After the data preprocessing, the proposed DICDEN technique performs the feature selection to minimize the time complexity of disease prediction. The heart disease dataset comprises the attributes for processing the input data. It is necessary to select important attributes for data classification. The main aim of feature selection is to minimize the dimensionality of the dataset by finding the more informative feature. The proposed technique uses the Rand indexive robust linear regression for selecting the more relevant features for heart disease prediction. It is a machine learning technique used to measure the relationship between the variables (i.e. features or attributes) with help of the Rand similarity index.

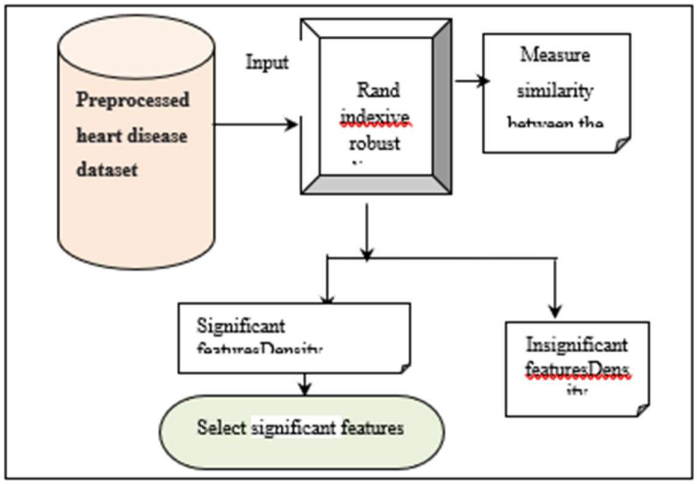

Figure 3 Block diagram of regression-based feature selection

Figure 3 given above illustrates the block diagram of feature selection. First, the preprocessed dataset is taken as input and given to the regression technique. The regression technique measures the relationship between the attributes with the help of the Rand similarity coefficient. Rand similarity coefficient is a statistical method used to measure the similarity and diversity of sample sets. This is also used to determine the well-matched attributes as well

as diabetes-related attributes for heart disease prediction. The similarity analysis between the features is given below.

$$\varphi = \left[\frac{[A_i \cap A_j]}{Number\ of\ attributes}\right] \quad (5)$$

Where, φ indicates a similarity coefficient, $A_i \cap A_j$ represents a mutual dependence between the attributes. The coefficient (φ) returns the similarity output ranges between 0 and 1. From the similarity index value, the well-matched significant attributes are selected for disease prediction as follows,

$$\varphi - \{1; \quad relevant\ features\ 0; \quad irrelevant\ features \quad (6)$$

Where ' φ 'indicates a similarity coefficient returns '1' indicates two attributes are highly matched, and the coefficient returns '0' indicates two attributes are not matched. The highly matched features are selected as significant and other features are removed. These relevant features are used for heart disease prediction to minimize the time complexity. The algorithmic process of Rand Indexive Robust linear Regression is given below

// Algorithm 2: Rand Indexive Robust linear Regression based feature selection

Input: Preprocessed Dataset, number of attributes i.e. features $A = \{A_1, A_2, A_3, \ldots, A_m\}$.
Output: Select significant features

Begin

1. **Collect the** number of features $A = \{A_1, A_2, A_3, \ldots, A_m\}$ from the preprocessed dataset
2. **for each feature**
3. Apply robust linear regression to analyze the similarity coefficient 'φ'
4. **if** $(\varphi - 1)$ **then**
5. Highly matched attributes
6. **else**
7. attributes are not matched
8. **End if**
9. **Select** highly matched features
10. Remove the other features from the dataset

end for
end

Algorithm 2 given above illustrates the different processes of significant feature selection for heart disease prediction. At first, the number of features is collected from the heart disease dataset. After that, the regression is applied for finding the relationship between the features with the help of the rand similarity coefficient. The similarity coefficient returns the similarity values in the ranges from 0 to1. Based on the similarity index results, more significant features are identified for classification. Otherwise, the attributes are removed from the dataset.

3.3 Tucker Congruence Radial Damped Convolutive Deep Learning Classifier

Finally, the heart disease prediction is performed using Tucker Congruencive Radial Damped Convolutive Deep Learning Classifier with the selected relevant features. A Convolutive Deep learning classifier is s a type of fully connected feed-forward deep neural network that uses a mathematical operation called convolution in at least one of its layers. The main advantage of

the Convolutive Deep learning classifier is it uses an efficient dense network that performs the prediction efficiently much more than regular neural networks. Therefore, the proposed technique uses the Convolutive Deep learning classifier for accurate heart disease prediction. The structure of the Convolutive Deep learning classifier is shown in figure 4.

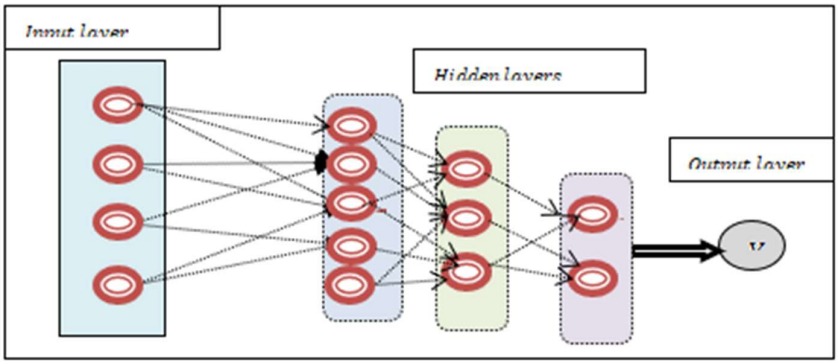

Figure 4 Schematic illustration of Tucker Congruencive Radial Damped Convolutive Deep Learning Classifier

Figure 4 illustrates a schematic diagram of the Tucker Congruencive Radial Damped Convolutive Deep Learning Classifier consists of three types of layers such as input, hidden (i.e. middle), and output layers. The input and output layers are always single layers, whereas the middle layer includes numerous sublayers for analyzing the given input data samples. Each layer typically includes small individual units called artificial neurons or nodes. This neuron has the ability to process the given weighted inputs and forward the output to other nodes with help of an activation function. An input to an artificial neuron is a number of training patient data from the input layer or outputs from a previous layer's neurons. The connection between these neurons is called a synapse.

From the above figure, the input layer consists of input training patient data. The proposed deep learning classifier includes three hidden layers and multiple sub-layers for learning the given input data. The three hidden layers are the convolutional layer, the maxpooling layer, and the fully connected layers. Finally, the result of the output layer provides the final disease prediction results. The convolutional layer is the first hidden layer of a deep learning classifier. Followed by, pooling layers, the fully-connected layers are presented.

Let us consider the number of training patient data , are given to the input layer. The input is transferred into the hidden layers.

- Convolutional layers

The convolution layer is the building block of the deep learning classifier. Convolutional layers convolve the input with the set of weights and pass its result to the next layer. Each convolutional neuron processes the input patient data only for its receptive field. This layer performs a mathematical operation called a "convolution". Convolution is a linear mathematical operation that involves the multiplication of a set of weights with the input patient data. The activity of the artificial neuron in the convolution layer is shown in figure 3.

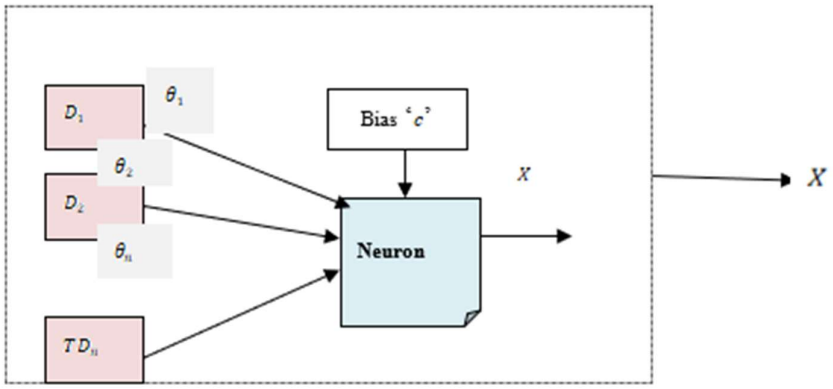

Figure 5. Flow process of artificial neuron

The input unit receives the selected input features i.e. training patient data and is given to the convolution layer. Therefore, the neurons take the inputs, $D_1, D_2, D_3, \ldots . D_n$, convolved with the set of weights, $\theta_1, \theta_2, \ldots, \theta_n$ and adds the bias term, B, then computes the linear function. Therefore, the output of the neuron in the convolution layer is given below,

$$X = \sum (\theta_j * D_i) + B \quad (7)$$

Where, X indicates a Convolutional layer output, θ_j denotes a set of weights assigned to an input patient data 'D_i'. Here, '$*$' denotes a mathematical operation called a "convolution". In this Convolutional layer, the feature map process is performed by analyzing the testing and training data using the Tucker congruence correlation coefficient. It is a statistical function used to measure the association between two variables based on such as testing and training patient data. The testing and training data analysis is performed as follows,

$$CC = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} \left| D_T(i) \cap D_{Tr}(j) \right|}{\sqrt{\sum D_T(i)^2} \sqrt{\sum D_T(j)^2}} \quad (8)$$

Where 'CC' indicates a Tucker congruence correlation coefficient, $\cap$ indicates a mutual dependence between the testing data D_T and training data $'D_{Tr}$, $\sum D_T(i)^2$ represents a squared score of $D_T(i)$, $D_{Tr}(j)^2$ indicates a squared score of $D_{Tr}(j)$. The congruence correlation coefficient provides the output ranges from 0 to 1. The mathematical results are transferred to the next pooling layer.

- **Max pooling layers**

It is the second layer of the convolutional deep learning classifier and it is used for reducing the dimensions of the datasets thereby minimizing the computational time of heart disease prediction. Max Pooling is the method, where the highly correlated results are taken as input from the feature map. Max Pooling provides the dimensions reduced output data. The Wilcoxon rank-sum test is a statistical measure used to Max Pooling operations for finding stochastically correlation results that are greater than the other by setting the threshold value.

$$W = \sum Q(CC, \vartheta) \quad (9)$$

$$Q = \{CC > \vartheta ; \quad Selected \quad CC < \vartheta; \quad Removed \quad (10)$$

Where W denotes an output of the Wilcoxon rank-sum test, Q denotes an output of the pooling layer, ϑ denotes a threshold, ϑ indicates a Tucker congruence correlation coefficient. As a result, a higher correlation than the threshold is selected for the next classification process. Otherwise, the results are removed resulting in minimizing the similarity of the data. The output of the pooling layer is sent to the fully connected layer for data classification.

- **Fully connected layer**

The last layers which determine the output of heart disease prediction are the fully connected layers. The output from the pooling layer is sent to the next hidden layer called the fully connected layer where each neuron receives input from the neurons of the previous hidden layer. A dense Layer is used to classify data based on output from convolutional layers. The classification is performed using the Radial activation function. An activation function performs a deterministic task that defines the output of that particular node for a given set of inputs. The data classification is performed by the Radial activation functions.

$$R = exp\ exp\left(\frac{|CC - m_c|^2}{2\,v^2}\right) \quad (11)$$

Where R denotes a Radial activation function, 'CC' indicates the correlation coefficient results, m_c denotes a mean of the particular class, 'v' indicates a deviation. Based on activation function results, data is correctly classified into particular classes namely the presence of heart disease and the absence of heart disease. After the data classification, the Huber loss also called squared error loss is measured for comparing the actual and predicted output.

$$L_H = \frac{1}{2}\,[\,R_A - R\,]^2 \quad (12)$$

Where the Huber loss 'L_H' is measured as a squared difference between the actual classification results 'R_A' and output predicted by the activation function 'R'. The iterated damped least-squares method helps to solve non-linear least-squares problems by finding the minimum Huber loss by updating the weight.

$$\theta_j{}' = \theta_j - \tau \left[\frac{\partial L_H}{\partial \theta_j}\right] \quad (13)$$

Where the $\theta_j{}'$ updated weight, θ_j denotes a current weight, τ denotes a learning rate (> 0) learning rate is a tuning hyperparameter that moves toward a minimum of a loss function, $\frac{\partial L_H}{\partial \theta_j}$ indicates a derivative of the loss with respect to weight θ_j. This process is iterated till the damped least-squares method finds the minimum loss. Finally, the classification results are obtained at the output layer. Based on the classification results, the heart disease prediction is correctly predicted with minimum error. The Tucker Congruencive Radial Damped Convolutive Deep Learning Classifier algorithm is described as given below,

//Algorithm 3: Tucker Congruencive Radial Damped Convolutive Deep Learning Classifier

Input: Selected relevant features $A = \{A_1, A_2, A_3, \ldots A_n\}$ and training patient data $D_1, D_2, D_3, \ldots D_m$

Output: Increase the heart disease prediction accuracy

Begin

1. **Number of selected features** $\{A_1, A_2, A_3, \ldots A_n\}$ with training data $D_1, D_2, D_3, \ldots D_m$ taken **into the input layer**
2. **For each** training data D **[convolutional layer]**
3. Convolve the weight 'θ_j' with the input ' D_i ' and add bias 'B'
4. Obtain the artificial neuron activity '$X(t)$'
5. **end for**
6. **For** each training data with testing data
7. Measure the Tucker Congruencive correlation coefficient 'CC'
8. **End for**
9. Measure Wilcoxon rank-sum test 'W'
10. *if* ($CC > v$) **then –[max pool layer]**
11. Select the data for classification
12. **else**
13. Remove the data
14. **end if**
15. **For selected correlation results 'CC' –[fully connected layer]**

16. Apply radial activation function 'R'
17. **If** ($R = +1$) **then**
18. heart disease is correctly predicted
19. **else**
20. Absence of heart disease
21. **End if**
22. **For each prediction results**
23. Calculate the Huber loss 'L_H'
24. Update the weight '$\theta_j^{\prime\prime}$'
25. Find minimum Huber loss by updating the weight
26. Obtain final disease prediction **at the output unit**

End

Algorithm 3 given above describes the step-by-step process of heart disease prediction with higher accuracy. The selected relevant features with the training data are given to the input layer of the deep learning classifier. Then the input is transferred into a convolution layer. The inputs are convolved with the set of weights and added to the bias. In the convolutional layer, the Tucker Congruencive coefficient is measured to find the correlation between the training data and testing patient data. Then the correlation results are transferred into the max pooling layer. In that layer, the higher correlation between training data and testing data is selected for the classification process by using the statistical measure. Then the output of the max pooling layer is given to the fully connected layer where the radial activation function is applied for classifying the data into a particular class. Followed by, the Huber loss is calculated for each predicted output. After that, the weight gets updated and finds the minimum Huber loss. Finally, the accurate heart disease prediction results are displayed at the output layer of deep learning classifier.

4.EXPERIMENTAL SETUP

In this section, experimental evaluation of the proposed DICDEN technique and existing CRMPTN [1], DBN+CSA [2] are implemented in Java with heart disease dataset taken from https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset. The main aim is to find the presence and absence of heart disease in the patient. The dataset consists of 14 attributes and 1025 instances. After collecting the data, pre-processing is carried out to fill in the missing value and remove the duplicate data. Then the significant features are selected for accurate disease prediction and other features are removed. Finally, the classification is performed with the selected features and identifies the presence of heart disease. Table 1 provides the attribute information.

Table 1 Attribute Description

S.no	Attributes	Description
1	Age	Patient age in days
2	Sex	The gender of the patient
3	cp	chest pain type Value 0: typical angina Value 1: atypical angina Value 2: non-anginal pain Value 3: asymptomatic
4	trestbps	Resting blood pressure (in mm Hg on admission to the hospital)
5	Chol	Serum Cholesterol
6	Fbs	Fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

7	Restecg	Resting electrocardiographic results
8	Thalach	Maximum heart rate achieved
9	exang	Exercise induced angina (1 = yes; 0 = no)
10	Oldpeak	ST depression induced by exercise relative to rest
11	Slope	the slope of the peak exercise ST segment
12	Ca	number of major vessels (0-3) colored by flourosopy
13	Thal	Thallium Stress Test, 1 = normal; 2 = fixed defect; 3 = reversable defect
14	Target	1 presence of disease no disease

5.PERFORMANCE RESULTS ANALYSIS

In this section, experimental results of the proposed DICDEN and existing CRMPTN [1], DBN+CSA [2] are discussed with the different performance metrics such as accuracy, precision, recall, F-measure, and time complexity. The performances of the proposed DICDEN with different metrics are discussed with the aid of a table and graphical representation.

5.1. Impact of Disease prediction Accuracy

It is measured as the number of patient data accurately diagnosed as heart disease presence or absence. Therefore, the overall disease prediction accuracy is mathematically calculated as given below,

$$DPA = \left[\frac{Tpos+Tneg}{Tpos+Tneg+Fpos+Fneg}\right] * 100 \qquad (14)$$

Where, DPA indicates disease prediction accuracy, $Tpos$ indicates a true positive rate, $Tneg$ indicates a true negative, $Fpos$represents a false positive, $Fneg$ denotes a false negative. Therefore the accuracy is measured in terms of percentage (%).

True positive: Disease patients correctly identified as sick
False positive: Normal patients incorrectly identified as sick
True negative: Normal patients correctly identified as normal
False negative: Disease patients incorrectly identified as normal

Table 2 comparative analysis of disease prediction accuracy

Number of patient data	Disease prediction accuracy (%)		
	DICDEN	CRMPTN	DBN+CSA
100	97	93	91
200	96.5	92.5	91
300	96.33	92.33	90.66
400	96.25	92	90.5
500	95.8	91.8	90.4
600	95.5	91.66	90.16
700	95.42	91.71	90.14
800	95.37	91.37	90
900	95.33	91.33	89.66
1000	95.2	91.3	89.5

Table 2 demonstrate comparative analysis of heart disease prediction accuracy with respect to the number of patient data taken from 100 to 1000 from the dataset.. The average of ten results indicates that the performance of the DICDEN technique is considerably improved by 4% when compared to [1] and 6% when compared to [2] respectively.

5.2 Impact of precision

It is measured based on the performance results of true positives as well as false positives. The formula for calculating the precision is given below,

$$PN = \left(\frac{Tpos}{Tpos + Fpos}\right) * 100 \quad (15)$$

Where, denotes a Precision, symbolizes the true positive, represents the false positive. The Precision is measured in percentage (%).

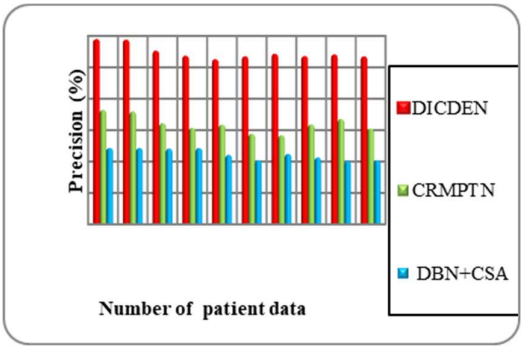

Figure 7 Graphical illustration of precision

Figure 7 depict overall performance assessment of precision of heart disease prediction with respect to a number of patient data. Average is taken for en comparison results and it indicates that the performance of precision is found to be increased by 2% when compared to [1] and 3% when compared to [2] respectively.

5.3 Impact of Recall

It is evaluated based on a number of true positives as well as false negatives during the heart disease prediction. It is mathematically formulated as given below,

$$RL = \left(\frac{Tpos}{Tpos + Fneg}\right) * 100 \quad (16)$$

Where 'RL 'indicates a recall, $Tpos$ represents the true positive, $Fneg$ denotes the false negative. The recall is measured in percentage (%).

Table 4 Comparison of recall

Number of patient data	Recall (%)		
	DICDEN	CRMPTN	DBN+CSA
100	98.94	96.70	95.50
200	98.41	96.13	95.50
300	98.58	96.29	95.13
400	98.67	96.13	94.97
500	98.30	95.81	95.06
600	97.88	95.94	94.93
700	97.73	96.04	94.71
800	97.75	95.32	94.67
900	97.65	95.12	94.38
1000	97.57	95.37	94.16

Table 4 given above indicate the overall performance results of recall using three different methods namely the DICDEN and existing CRMPTN [1], and DBN+CSA [2]. The overall comparison results indicate that the performance of recall using the DICDEN technique is

significantly enhanced by 2% when compared to CRMPTN [1] and 3% when compared to DBN+CSA [2] respectively.

5.4 Impact of F-measure

It is measured as the average of performance results of precisions as well as recall. It is calculated as given below,

$$F - me = \left[2 * \frac{PN * RL}{PN + RL}\right] * 100 \quad \textbf{(17)}$$

Where $F - me$ indicates an F-measure computed based on precision PN and recall 'RL'. F-measure is measured in terms of percentage (%).

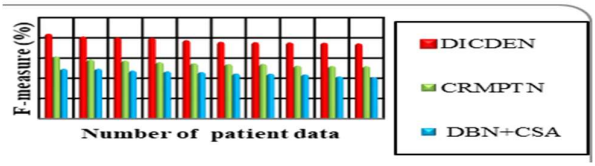

Figure 9 Graphical illustration of F-measure

Figure 9 demonstrate the experimental results of the F-measure along with the number of patient data. The average of ten comparison values indicates that the overall performance of F-measure is increased by 2% when compared to [1], and 3% when compared to [2] respectively.

5.5 Impact of disease prediction time

It is defined as the amount of time consumed by the algorithm to accurately predict heart disease presence or absence. Therefore, the overall time consumption is measured as given below,

$$DPT = n * T\ (PoD) \quad (18)$$

Where DPT indicates a disease prediction time, n denotes the number of patient data, and T (PoD denotes a time for predicting one data sample (PoD). The overall prediction time is measured in terms of milliseconds (ms).

Table 6 Comparative analysis of disease Prediction time

Number of patient data	Disease Prediction time (ms)		
	DICDEN	CRMPTN	DBN+CSA
100	18	20	25
200	24	28	30
300	33	37.5	38.7
400	36	40	44
500	40	45	50
600	42	48	54
700	45.5	49	56
800	49.6	52	57.6
900	52.2	53.1	59.4
1000	55	60	63

Table 6 indicates performance analysis of heart disease prediction time versus the number of patient data. The overall performance results indicate that the heart disease prediction time is considerably minimized by 9% and 18% when compared to existing [1] [2] respectively.

5.6 Comparative analysis of chest pain type based on gender analysis

In this section, the performance analysis of chest pain type is based on gender analysis with respect to fasting blood sugar using the proposed DICDEN. The different results are obtained as shown in the Table 7.

Table 7 Comparative analysis of chest pain type based on gender analysis

	Male Fasting blood sugar <120 (mg/dl)	**Fasting blood sugar >120 (mg/dl)**	**chest pain type**
Male (number of count)	23	5	0- typical angina
	6	3	1- atypical angina
	15	8	2- non-anginal
	3	4	3- asymptomatic
Female(number of count)	12	3	0- typical angina
	5	1	1- atypical angina
	8	2	2- non-anginal
	1	1	3- asymptomatic
Total	**73**	**27**	

Figure 11 Graphical illustration of Chest Pain Based on Gender Analysis (Fasting blood sugar (<120 (mg/dl))

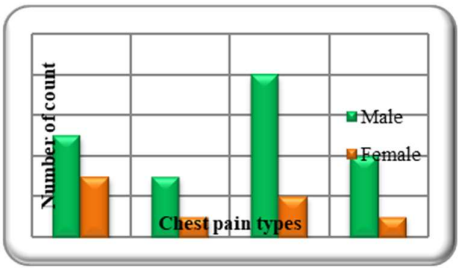

Figure 12 Graphical illustration of Chest Pain based on Gender Analysis (Fasting blood sugar (>120 (mg/dl))

Figure 11 and 12 illustrate performance results of chest pain based on gender analysis with the fasting blood sugar (<120 (mg/dl) and >120 (mg/dl)) versus chain pain types. Let us consider 100 patient data for conducting the experiment. The proposed technique accurately finds the number of female and male patients who affect different chest pain. There are four different chest pains such as typical angina, atypical angina, non-anginal and asymptomatic. Totally, 73 patients have chest pain with Fasting blood sugar (<120 (mg/dl)) and 27 patients have chest pain with Fasting blood sugar (<120 (mg/dl)).

5.7 Comparative analysis of parameters

The dual formation of parameters like Age, Gender, pressure, cholesterol, heart rate, blood sugar, and blood sugar are shown in Table 8.

Table 8 Comparative analysis of different parameters

Age	Gender	Pressure	cholesterol	heart rate	blood sugar	Chest pain types	Target
58	Male	114	318	140	<120 mg/dl	0- typical angina	1(heart disease presence)
52	Male	128	205	184	>120 mg/dl	1- atypical angina	1(heart disease presence)
65	Female	140	417	157	>120 mg/dl	2- non-anginal	1(heart disease presence)
58	Female	150	283	162	>120 mg/dl	3- asymptomatic	1(heart disease presence)

Table 8 indicates the performance results of different parameters. The experiments demonstrated that the proposed DICDEN improves the performance of the classifier for early identification of heart disease with different chest pain types.

6. CONCLUSION

Early detection of heart disease prediction in Diabetic Patients is very significant for reducing the causes of death around the world. A novel DICDEN technique is developed to increase the classifier's performance for heart disease prediction with three different stages. In the proposed DICDEN technique, the Manhattan hot deck imputed technique is applied for data preprocessing to obtain the structured format. Followed by, Rand indexive robust linear regression is employed to perform significant feature selection for improving the classification with minimum time. Finally, Tucker Congruence Radial Damped Convolutive Deep Learning Classifier is applied for predicting the data as disease presence or absence. As a result, the classifier provides precise classification results with minimum error. Based on the classification results, heart disease is correctly predicted. A comprehensive experimental evaluation is carried out using the heart disease dataset with different parameters such as Accuracy, precision, recall, F-measure, and disease prediction time. The quantitative performance result indicates that the proposed DICDEN achieves higher accuracy in heart disease prediction for diabetes patients.

REFERENCES

1. Wenqi Li, Ming Zuo, Hongjin Zhao, Qi Xu, Dehua Chen , “Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks”, MethodsVolume 198, February 2022, Pages 96-106.https://doi.org/10.1016/j.ymeth.2021.12.009

2. Nandakumar P, Subhashini Narayan, “Cardiac disease detection using cuckoo search enabled deep belief network”, Intelligent Systems with Applications, Volume 16 , 2022, Pages 1-14. https://doi.org/10.1016/j.iswa.2022.200131

3. Muhammad Salman Pathan, Avishek Nag, Muhammad Mohisn Pathan, Soumyabrata Dev, “Analyzing the impact of feature selection on the accuracy of heart disease prediction”, Healthcare Analytics, Elsevier, Volume 2, 2022, Pages 1-9. https://doi.org/10.1016/j.health.2022.100060

4. Chunyan Guo, Jiabing Zhang , Yang Liu, Yaying Xie, Zhiqiang Han, and Jianshe Yu, "Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform", : IEEE Access, Volume 8, 2020, Pages 59247 – 59256. DOI: 10.1109/ACCESS.2020.2981159

5. Katarzyna Dziopa, Folkert W. Asselbergs, Jasmine Gratton, Nishi Chaturvedi & Amand F. Schmidt, "Cardiovascular risk prediction in type 2 diabetes: a comparison of 22 risk scores in primary care settings", Diabetologia, Springer, Volume 65, 2022, Pages 644–656. https://doi.org/10.1007/s00125-021-05640-y

6. Xi Su1, Yongyong Xu, Zhijun Tan, Xia Wang, Peng Yang, Yani Su, Yangyang Jiang, Sijia Qin, Lei Shang, "Prediction for cardiovascular diseases based on laboratory data: An analysis of random forest model", Journal of Clinical Laboratory Analysis, Wiley, Volume 34, Issue 9, 2020,

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

8. Khalid Bahani, Mohammed Moujabbir, Mohammed Ramdani, "An accurate fuzzy rule-based classification systems for heart disease diagnosis", Scientific African, Elsevier, Volume 14, 2021, Pages 1-14. https://doi.org/10.1016/j.sciaf.2021.e01019

9. Tsatsral Amarbayasgalan, Van-Huy Pham, Nipon Theera-Umpon, Yongjun Piao, And Keun Ho Ryu , "An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well Ordered Training Datasets", IEEE Access , Volume 9, Pages 135210 – 135223. DOI: 10.1109/ACCESS.2021.3116974

10. Sarria E. A. Ashri, M. M. El-Gayar, Eman M. El-Daydamony, "HDPF: Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm", IEEE Access, Volume 9, Pages 146797 – 146809. DOI: 10.1109/ACCESS.2021.3122789

11. Pavithra V and Jayalakshmi V, "Hybrid feature selection technique for prediction of cardiovascular diseases", Materials Today: Proceedings, 2021, Pages1-5.https://doi.org/10.1016/j.matpr.2021.03.225

12. Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian M. W. Quinn, Mohammad Ali Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison", Computers in Biology and Medicine, Elsevier, Volume 136, 2021,Pages1-10. https://doi.org/10.1016/j.compbiomed.2021.104672

13. Sudarshan Nandy, Mainak Adhikari, Venki Balasubramanian, Varun G. Menon, Xingwang Li, Muhammad Zakarya, "An intelligent heart disease prediction system based on swarm artificial neural network", Neural Computing and Applications , Springer, 2021, Pages 1-15. https://doi.org/10.1007/s00521-021-06124-1

14. Da Xu, Jessica Qiuhua Sheng, Paul Jen-Hwa Hu, Ting-Shuo Huang, and Chih-Chin Hsu, "A Deep Learning–Based Unsupervised Method to Impute Missing Values in Patient Records for Improved Management of Cardiovascular Patients", IEEE Journal of Biomedical and Health Informatics , Volume 25, Issue 6, 2021, Pages 2260 – 2272. DOI: 10.1109/JBHI.2020.3033323

15. Pronab Ghosh, Sami Azam , Mirjam Jonkman, Asif Karim, F. M. Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu, and Friso De Boer,

"Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques", IEEE Access, Volume 9, 2021, Pages 19304 – 19326. DOI: 10.1109/ACCESS.2021.3053759

16. Shishir Rao , Yikuan Li , Rema Ramakrishnan , Abdelaali Hassaine , Dexter Canoy, John Cleland , Thomas Lukasiewicz, Gholamreza Salimi-Khorshidi, and Kazem Rahimi, "An Explainable Transformer-Based Deep Learning Model for the Prediction of Incident Heart Failure", IEEE Journal of Biomedical and Health Informatics , Volume 26, Issue 7, 2022, Pages 3362 – 3372. DOI: 10.1109/JBHI.2022.3148820

17. Saba Bashir, Abdulwahab Ali Almazroi, Sufyan Ashfaq, Abdulaleem Ali Almazroi, And Farhan Hassan Khan, "A Knowledge-Based Clinical Decision Support System Utilizing an Intelligent Ensemble Voting Scheme for Improved Cardiovascular Disease Prediction", IEEE Access, Volume 9, Pages 130805 – 130822. DOI: 10.1109/ACCESS.2021.3110604

18. Simanta Shekhar Sarmah, "An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System Using Deep Learning Modified Neural Network", IEEE Access, Volume 8, 2020. Pages 135784-135797. DOI: 10.1109/ACCESS.2020.3007561

19. Abdullah Alqahtani , Shtwai Alsubai , Mohemmed Sha , Lucia Vilcekova, and Talha Javed, "Cardiovascular Disease Detection using Ensemble Learning", Computational Intelligence and Neuroscience, Hindawi, Volume 2022, August 2022, Pages 1-9. https://doi.org/10.1155/2022/5267498

20. Xing Han Lu, Aihua Liu, Shih-Chieh Fuh, Yi Lian, Liming Guo, Yi Yang, Ariane Marelli, Yue Li, "Recurrent disease progression networks for modelling risk trajectory of heart failure", PLoS ONE, Volume 16, Issue 1, https://doi.org/10.1371/journal.pone.0245177