International Journal of Advanced Science and Technology
Vol. 29, No. 6, (2020), pp. 4225 - 4234

Early Prediction of Coronary Heart Disease from Cleveland Dataset using Machine Learning Techniques

Ramya Perumal[1], Kaladevi AC[2]
[1]*Assistant Professor, Dept. of CSE, Sona College of Technology, Salem*
[2] *Professor, Dept. of CSE, Sona College of Technology, Salem*

Abstract

The mortality rate of the person affected by heart disease is kept on increasing day by day to a greater extent. Also, the survival rate of approximately 50% of the patient suffering from heart disease is approximately less than 10 years [9].The exponential growth rate of high-dimensional data in the medical domain needs automation for analysing the data. Hence it is necessary to have an effective computational intelligent system to detect and predict heart disease in advance. Angiography is an imaging test taken for predicting heart disease. It incurs high cost and severe side effects to the patients. Also, it requires experts to diagnose the patient data [4]. To facilitate the process, we proposed an effective computational intelligent system that integrates Principal Component Analysis (PCA) and machine learning classifier models such as k-Nearest Neighbour (k-NN), Support Vector Machine (SVM) and Logistic Regression (LR) to predict the person is affected by heart disease or not [6]. To validate the performance of model, performance measures such as accuracy, specificity, sensitivity, error rate, and Mathews Correlation Coefficient are used. These performance measures of our proposed system provide promising results in heart disease prediction. Among the three classifier models, LR outperforms in all aspects of performance measures and provides comparatively similar results when compared to the SVM. Our proposed system efficiently provides remarkable results in predicting coronary heart disease. In the future, we are focusing to gather and investigate real-world datasets and apply diverse mixture of machine learning techniques[2].

Keywords: *Machine Learning, Coronary Heart Disease, Support Vector Machine, K-Nearest Neighbour, Logistic Regression, Principal Component Analysis.*

1. Introduction

According to WHO, cardiovascular disease is the first non-communicable disease causes numerous death in the world. In all over the world, 17.9millions of people lost their lives because of heart disease [4].The survey also conveys the importance of safeguarding the life of people from heart disease. With the emerging trends of Artificial Intelligence (AI) and Machine Learning techniques (ML) in the medical domain, heart disease can be predicted efficiently with great accuracy [1][18-20]. The conventional method for detecting heart disease requires costly lab tests and severe side effects. Moreover, it is a complex and time-consuming task for medical experts to analyze the patient data and provide the result [5]. Also, it involves some medical expertise and their intuitions to confirm the disease. To overcome these limitations, an automated machine learning system acts as a diagnosing tool for performing the task. The machine itself learns from the real-life parameters and builds a classifier model to check whether the person is affected by heart disease or not.

Some of the Cardio Vascular Diseases (CVD) are Coronary Heart Disease (CHD), Arrhythmias, Stroke, Myocarditis, Cardiomyopathy, Congestive heart failure, etc.[5][15]. The proposed work is particularly focussing on Coronary Heart Disease (CHD). The reason for the cause of CHD is due to blockage of bad cholesterol in the coronary arteries. It leads to an insufficient supply of oxygen and blood to the heart results in heart failure [4]. The early symptoms of the CVD are irregular heartbeats, shortness of breath, pain or pressure in the chest, fainting, fatigue, etc. [5]. Cleveland dataset lists nearly 76 risk factors that are responsible for heart disease. Among them, 13 are considered as major risk factors contributing the heart disease. These risk factors are broadly classified into two categories, mutable and non-mutable factors. Non-mutable risk factors are age, gender, and family medical history. Mutable risk factors are dynamic because physical changes occur in the body such as BMI, blood pressure, cholesterol level, maximum heart rate, etc. [3]. Physical activity such as regular exercise and proper diet keep these risk factors like cholesterol level and blood pressure in a controlled fashion and reduce the risk of causing heart disease. Research highlights smoking, drinking and obesity are also risk factors for heart disease [1]. The proposed system is the combination of feature reduction technique and machine learning classifier models to obtain accurate heart disease prediction. It is a cost-effective and speedy computational intelligent system that provides results with good accuracy. The machine learning models learn from the inputted data and fit a model by analyzing the characteristics of data and get converge with the model by reducing the error in the data instances. The correlated features in the data and its level of degree are identified by the Pearson correlation coefficient. The degree of correlation of the value -1 represents negatively correlated and 1 represents positively correlated features [9]. Machine learning is a cognitive task that replicates the human learning system. It is an eminent field that has a great scope in the medical domain. It is defined as some class of tasks T learns the task by experience and shows its improvement in performance P [10]. It is broadly classified as three categories supervised learning, unsupervised learning, and reinforcement learning [6][19-20]. The proposed work is to predict heart disease that comes under the category of supervised learning.

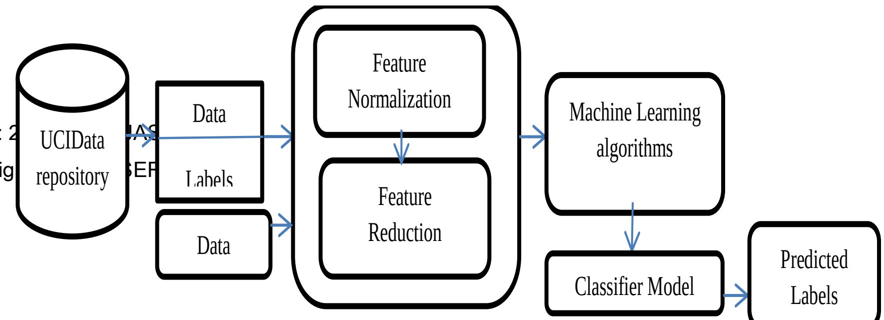
Nowadays lots of data are generated at an extremely high speed in the medical domain. Data are available in a digitalized format that requires an expert system to efficiently store and access the data. In addition to that, it also requires some computational intelligent system to analyze and diagnose those high-dimensional data. Cardiovascular disease is a life-threatening disease that requires such an effective computational intelligent system to predict the CVD. The proposed system concentrates on CHD prediction. It uses the PCA technique and machine learning classifier models to predict the likelihood of the person having the heart disease.

2. Related Works

Considerable experimentation and researches have been performed in predicting the heart disease. Though the utilization of data mining techniques in the CHD prediction able to extract the hidden pattern from the data, it requires human intervention to analyze the data and interpret the results [5]. The exponential growth rate of data in the medical domain needs automation for handling and reporting the data. The exploration of AI and machine learning (ML) techniques in the medical domain provides remarkable results to predict CHD. Many works are carried out to find the association of the risk factors causing heart disease through structural equation modelling. Research finds factors like BMI, obesity, drinking, hypertension, and smoking increase the chance of getting heart disease. It also reveals that the hypertension is strongly associated with CHD .The mortality rate due to CHD is 2.3 times greater if the person has hypertension [1].The working environment is yet another factor to be considered for heart disease. People working in iron and steel manufacturing unit have an abnormality in blood lipid due to high temperature and noise in their surroundings [2]. By considering the cost constraints many handy tool kits are available in the markets for monitoring the real-life parameters of the heart patients[3].Mostly the research work has been conducted to find optimal features that are responsible for heart disease .The feature extraction technique helps to find optimal features further improvises the prediction accuracy of heart disease [4] [9] [14] .Experimenting with different classifier models Naïve Bayes(NB), Decision Tree(DT), Support Vector Machine(SVM), k Nearest Neighbour(k-NN), Logistic Regression(LR) helps to predict heart disease with good accuracy[6][8][11].

3. Methodology

The proposed system consists of data collection, feature standardization, feature reduction, and classifier model. Figure .1 represents the block diagram of the proposed system.



ISSN: 2
Copyrig

Fig 1. Block Diagram of The Proposed System

Data Collection

The proposed system uses the Cleveland dataset consists of the data from the UCI repository. Dataset consists of 303 data instances of both healthy subjects and heart patient details. It contains 76 features out of which 13 features are important. It contains both numerical and categorical data .

3.1 Pre-processing Techniques

3.1.1 Feature Standardization

The Cleveland dataset contains numerical features such as age, blood pressure, maximum heart rate, etc. with a different unit of measurements. Feature standardization is bringing the numeric features to the same unit of measurement with zero mean and unit standard deviation [9]. It also consists of categorical features

Table 1. Heart Disease Features of Cleveland Dataset

Sno	Features	Feature Description	Feature Type	Feature Data Range
1	Age	Age in years	Numeric	[29-77]
2	Sex	Sex	Binary	0=female 1=male
3	Cp	Chest Pain Type	Nominal	1=typical angina 2=atypical angina 3=non-anginal pain 4=asymptomatic
4	Trestbpss	Resting blood pressure in mm Hg on admission to the hospital	Numeric	[94,200]
5	Chol	Serum Cholesterol in mg/dl	Numeric	[126,564]
6	FBS	Fasting blood sugar>120 mg/dl	Binary	1=true 0=false
7	Restecg	Resting electrocardiographic results	Nominal	0=normal 1=ST-T wave abnormality 2=left ventricular hypertrophy
8	Thalach	Maximum heart rate	Numeric	[71,202]

ISSN: 2005-4238 IJAST

		achieved		
9	Exang	Exercise induced angina	Binary	0=no 1=yes
10	Oldpeak	ST depression induced by exercise relative to rest	Numeric	[0,6.2]
11	Slope	The slope of the peak exercise ST segment	Nominal	1=up sloping 2=flat 3=down sloping
12	Ca	Number of major vessels coloured by fluoroscopy	Nominal	0-3
13	Thal	Defect type	Nominal	3=normal 6=fixed defect 7=reversible defect
14	Num	Diagnosis of heart disease	Binary	0=less than 50%diameter narrowing (Normal) 1=greater than 50%diameter narrowing (Patient)

such as sex, chest pain type, blood sugar, etc. are transformed into a one-hot encoding mechanism [14]. Feature normalization is the necessary pre-processing step in the heart disease prediction system [4]. Our dataset has no missing values and wrong interpretation of data. If it exists, that has to be taken into an account as a first pre-processing step to fill the missing entry, verifying the data format and correcting the wrong interpretation of the data.

3.1.2 Feature Reduction

The proposed system uses PCA as a feature reduction technique. It eliminates redundant and irrelevant features by preserving significant features. PCA transforms the features into a new set of features known as principal components that are orthogonal to each other. It ranked principal components according to its variance. The features and their relationships in the dataset are identified by the Pearson correlation coefficient. Features are categorized as positively correlated, negatively correlated and uncorrelated. As the number of features increases, it is hard to visualize it. Hence PCA finds the uncorrelated or independent features from the dataset. It identifies the principal components that are the eigenvectors of a correlation matrix, and they are orthogonal to each other [9]. Using PCA, 7 independent principal components are identified that maximizes the results of the classifiers. The proposed system enhances its performance with a minimal set of features.

$$\mathbf{Av = \lambda v}$$

Where A is a correlation matrix,

v is the eigenvectors

λ is the eigenvalues associated with v

3.3. Coronary Heart Disease classifier model

The classifier model uses a machine learning technique to predict the likelihood of a person affected by heart disease. The proposed heart disease prediction system consists of two phases. They are the training phase and the testing phase. In the training phase, sampled data with its labels are given as an input to the system. The proposed system uses the learning algorithm to derive a fitting function by concerning the characteristics of the data. At the testing phase, the unseen data with no labels are provided to the system. The system predicts the correct label of the unseen data by experience. More the sampled data you provided during the training phase, the higher the accuracy in prediction at the testing phase you arrived[10].In the proposed system, three classifier models such as kNN, SVM and LR are used[7].

3.3.1. K-Nearest Neighbour Classifier

It is a supervised ML model used both for regression and classification. It is simple and easy to implement. It predicts the label of unseen data based on distance and proximity of the neighbourhood classes of data. The popular measure for calculating the distance is Euclidean distance. The trial and error method is used to find k value. k is the number of neighbours to be considered. The data instance is categorized into a specific class based on the majority votes of its neighbour data instance's classes. In our proposed system, k is assigned with a value 12 using trial and error method. We incrementally assign the value for k starting from 1, up to a certain limit it gives similar results. When k value exceeds 12, the performance of the proposed system degrades. Hence we fix the k values as 12.

3.3.2. Logistic Regression classifier

LR classifier is used both for regression and classification. It is also used when the target variable is of categorical type. It depends on the linear combination of multivariate features and perform logistic function (sigmoid) to classify data instances. The hypothesis function is given by,

$h\Theta(x) = \text{sigmoid}(Z)$

$h\Theta(x) = P(Y=1|X;\text{theta})$

Probability that Y=1 given X which is parameterized by theta

$P(Y=1|X;\text{theta}) + P(Y=0|X;\text{theta}) = 1$

$P(Y=0|;\text{theta}) = 1 - P(Y=1|X;\text{theta})$

The logistic function uses Z values ranges between -theta and theta. Its value is 0.5 at the origin. Given the input variable X, if the Z value goes to negative infinity then the output of the target variable is Y=0. If the Z value goes to infinity, then the output of the target variable is Y=1.

3.3.3.Support Vector Machine

SVM is the supervised model also used for both classification and linear regression. It is used to fit a hyperplane in an N-dimensional data that distinctly classifies the data instances. Where N is the number of features and hyperplane is usually the linear decision boundary to separate the classes. There are n number of hyperplanes can be drawn to separate the classes. Finding the optimal hyperplane that maximizes the margin is necessary to separate the two classes. The maximum margin is the distance between data instances of the two classes. The kernel function in SVM is of two types linear and radial basis function which is a non-linear function. Our proposed uses linear SVM to classify the heart patient from healthy subjects.

Consider the training data is of the form {(x1,y1),(x2,y2),….(xn,yn)}
x1, x2…xn are the data instances
y1, y2 …yn are the labels of the given data instances
In testing data, no labels are provided only the data instances are given
Based on the linear hyperplane, the fitting function is given by
yi=w.xi + b
If yi >=1, the data points belong to class 1
If yi <1, the data points belong to class 2

4. Experimental Setup

All computations are performed on Intel (R) Core (TM) i5-8250U CPU @1.80GHz with 64bit windows 10 as the operating system. All the experimentsare performed using the Python software package. The proposed system uses the Cleveland dataset obtained from the UCI repository. It consists of 165 instances of heart patients and 138 instances of healthy subjects [9]. The data are partitioned into 80:20.80% of data are used as the training dataset and the remaining 20% of data are used for the test phase. During the testing phase, it validates the performance of the classifier model [5]. The following are the performance measures conducted to evaluate the potential of our proposed system.

4.1.Performance measure

The performance of the classifier model is evaluated by using the confusion matrix. It is of 2*2 confusion matrix. It consists of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values [12].

True Positive is a measure that defines the number of instances has CHD and it is also correctly predicted by the classifier model

True Negative is a measure that defines the number of instances has no CHD disease and it is also correctly predicted by the classifier model

False Positive is a measure that defines the number of instances has CHD but it is predicted as normal by the classifier model

False Negative is a measure that defines the number of instances has no CHD but it is predicted as patients by the classifier model

Sensitivity, specificity and accuracy are described in terms of TP, TN, FN, and FP.

Accuracy is defined as the number of correct predictions made by classifier out of the total number of predictions.

Accuracy = (TN + TP)/(TN+TP+FN+FP)

Error rate is defined as the number of incorrect predictions made by the classifier out of the total number of predictions

Sensitivity is defined as the number of positive predictions made by the classifier out of the total number of positive predictions. It is also known as True Positive Rate or Recall.

Sensitivity = TP/(TP + FN)

Specificity is defined as the number of negative predictions made by the classifier out of the total number of actually negative predictions. It is also known as True Negative Rate.

Specificity = TN/(TN + FP)

Mathews Correlation Coefficient (MCC) is the measure used to predict the classifier score ranges between [-1, 1]. It returns the values -1, +1, 0 for a completely wrong, correct and random class prediction respectively [6][8][9].

4.2.Result

The following table2 describes the characteristics of data. It is observed from the sample data, that the person on an average of age 55 years with an average of blood pressure 132 mmHg, serum cholesterol 246 mg/dl, depression induced by exercise related to rest as 1 and maximum heart rate achieved as 150 tends to be affected by CHD. The percentage of male count in the dataset is 68% out of which 45% of them are affected by CHD. The percentage of female count in the dataset is 32% out of which 75% of them are affected by CHD. It is also found that most females have high cholesterol and males are suffering from high BP and maximum heart rate from the dataset.

Table 2. Characteristics of Data In the Cleveland Dataset

ISSN: 2005-4238 IJAST

	Age	sex	cp	Trest Bps	chol	fbs	Rest ecg	Thal ach	exang	oldpeak	Slope	ca	Thal	target
count	303	303	303	303	303	303	303	303	303	303	303	303	303	303
mean	54.36	0.68	0.96	131.62	246.26	0.14	0.53	149.64	0.32	1.04	1.39	0.72	2.3	0.54
std	9.08	0.46	1.03	17.53	51.83	0.35	0.52	22.90	0.46	1.16	0.61	1.0	0.61	0.50
min	29	0	0	94	126	0	0	71	0	0	0	0	0	0
max	77	1	3	200	564	1	2	202	1	6.2	2	4	3	1
Std Error	0.52	NA	NA	1.00	2.97	NA	NA	1.31	NA	0.06	NA	NA	NA	NA
Median	55	NA	NA	130	240	NA	NA	153	NA	0.8	NA	NA	NA	NA
Mode	58	NA	NA	120	204	NA	NA	162	NA	0	NA	NA	NA	NA
Variance	82.48	NA	NA	307.58	2686.42	NA	NA	524.64	NA	1.34	NA	NA	NA	NA
Kurtosis	-0.54	NA	NA	0.93	4.51	NA	NA	-0.061	NA	1.58	NA	NA	NA	NA
Skew	-0.20	NA	NA	0.71	1.14	NA	NA	-0.54	NA	1.27	NA	NA	NA	NA
Range	48	NA	NA	106	438	NA	NA	131	NA	6.2	NA	NA	NA	NA

Table 3 represents the correlation matrix of all the features in the data and its relationship using the Pearson correlation coefficient. The correlation matrix represents the strength of the relationship that exists between features in the data. The value +1 represents features are perfect positively correlated,-1 represents the features are perfect negatively correlated and 0 represents features are uncorrelated.

Table 3. Pearson Correlation Coefficient Considering All Features of Cleveland Dataset

Age	sex	Cp	Trest bps	chol	fbs	rest ecg	Thal Ach	exang	oldpeak	Slope	ca	thal
-	-0.098	-0.069	0.279	0.214	0.121	-0.116	-0.399	0.097	0.210	-0.169	0.276	0.068
-0.098	-	-0.049	-0.057	-0.198	0.045	-0.058	-0.044	0.142	0.096	-0.031	0.118	0.210
-0.069	-0.049	-	0.048	-0.077	0.094	0.044	0.296	-0.394	-0.149	0.120	-0.181	-0.162
0.279	-0.057	0.048	-	0.123	0.178	-0.114	-0.047	0.068	0.193	-0.121	0.101	0.062
0.214	-0.198	-0.077	0.123	-	0.013	-0.151	-0.010	0.067	0.054	-0.004	0.071	0.099
0.121	0.045	0.094	0.178	0.013	-	-0.084	-0.009	0.026	0.006	-0.060	0.138	-0.032
-0.116	-0.058	0.044	-0.114	-0.151	-0.084	-	0.044	-0.071	-0.059	0.093	-0.072	-0.012
-0.399	-0.044	0.296	-0.047	-0.010	-0.009	0.044	-	-0.379	-0.344	0.387	-0.213	-0.096
0.097	0.142	-0.394	0.068	0.067	0.026	-0.071	-0.379	-	0.288	-0.258	0.116	0.207
0.210	0.096	-0.149	0.193	0.054	0.006	-0.059	-0.344	0.288	-	-0.578	0.223	0.210
-0.169	-0.031	0.120	-0.121	-0.004	-0.060	0.093	0.387	-0.258	-0.578	-	-0.080	-0.105
0.276	0.118	-0.181	0.101	0.071	0.138	-0.072	-0.213	0.116	0.223	-0.080	-	0.152
0.068	0.210	-0.162	0.062	0.099	-0.032	-0.012	-0.096	0.207	0.210	-0.105	0.152	-

Table 4 consists of the reduced feature set after implementing PCA. Large numbers of features are hard to visualize. PCA can capture the relationship of features in lower-dimensional representation. The red colour indicates the principal components(pc) are negatively correlated to the features and blue colour indicates the pc is negatively correlated to the features.

Table 4.Reduced Feature Set after Implementing PCA On Cleveland Dataset

Features	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Sex	-0.050	0.552	-0.464	0.322	-0.224	-0.504	0.263
Trestbps	-0.471	-0.263	-0.201	-0.375	-0.606	0.164	0.366

Chol	-0.266	-0.560	0.285	0.247	0.110	-0.664	0.144
Fbs	-0.243	-0.153	-0.642	-0.378	0.562	-0.134	-0.172
Restecg	0.333	0.216	0.284	-0.738	-0.041	-0.450	0.117
thalach	0.464	-0.377	-0.336	0.030	-0.470	-0.192	-0.520
oldpeak	-0.566	0.317	0.241	-0.080	-0.178	-0.145	-0.680

Table 5 consists of principal components and their variance. Upon implementing the PCA, it can pertain to the features with high variance with a new set of principal components that are ranked in descending order. The first principal component (PC) is the linear combination of original features with preserving high variance in data. Subsequently, the principal components explain the remaining variances in the data under the condition the principal component are orthogonal to the previous one. According to the table.5, the first four principal components capture the 70% proportion of data with unit variance.

Table 5. Principal Components and Its Variances

PC	Variance	Proportion	Cumulative proportion
PC1	1.514	0.216	0.216
PC2	1.278	0.183	0.399
PC3	1.113	0.159	0.558
PC4	0.967	0.138	0.696
PC5	0.805	0.115	0.811
PC6	0.729	0.104	0.915
PC7	0.595	0.085	1.000

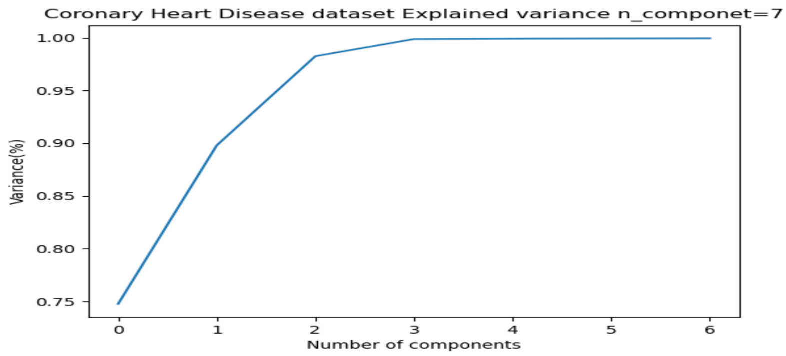
In the proposed system, three classifier models are used [7]. Table 6 represents the performance measures of the three proposed classifier models. Since data distribution is linear in the dataset, the multivariate linear combination of variables in LR classifier and SVM classifier gives comparatively similar results in all aspects of performance measures. The performance of KNN classifier is low when compared to SVM and LR classifiers. MCC represents the test accuracy of the prediction and approximately nearer to the correct prediction[11].

Table.6 Performance Measures of the Proposed System

Model	**SVM**	**LR**	**kNN=12**
Sensitivity	82%	85%	62%
Specificity	87%	88%	78%
Accuracy	85%	87%	69%
MCC	0.7	0.73	0.4
Error Rate	15%	13%	31%

Fig.1.represents the plot show selecting the number of principal components as 7, PCA preserves 100% of the total variance of the data. It indicates the best 7 principal components that are the linear combination of the original features and are orthogonal each other.Fig.2 represents the plot that

captures performance measures such as accuracy, error rate, sensitivity, specificity and MCC of the proposed classifier models of the kNN, SVM and LR[11].



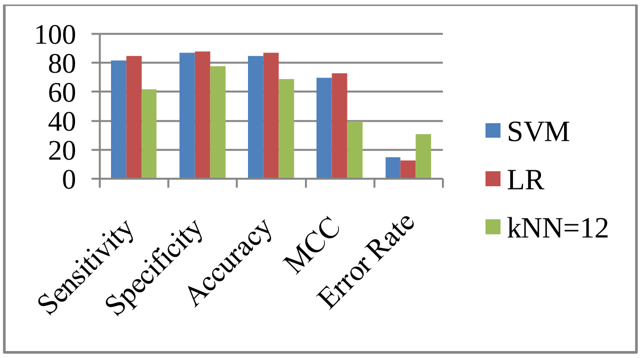


Fig.2.The Performance Measure of the Proposed System

5. Conclusion and Future Enhancement

Cardiovascular disease becomes one of the primary concerns to be taken care for safeguarding the people. Particularly South Asian people are greatly affected by CHD. Despite the conventional risk factors, prenatal, early life exposure and long working hours also play an important role in the excess of CHD in South Asian people[16][17]. Hence with proper diet control, lifestyle and exercises help to control CVD. Our contribution towards CVD is to implement a cost-effective system that predicts the disease with precise results. We restrict our experimentation with three classifiers namely LR, SVM, KNN, and its performance are assessed by the evaluation metrics such as accuracy, error rate, sensitivity and specificity, and MCC. Among the three classifiers, LR and SVM outperform the kNN classifier. The sensitivity, specificity, accuracy, MCC and error rate of LR classifier are 85%,88%,87%,0.73,13% respectively .Also, the SVM provides similar results in terms of sensitivity, specificity, accuracy, MCC and error rate are 82%,87%,85%,0.7 and 15% respectively. The k-NN classifier provides lower performance in all aspects of evaluation metrics.

The proposed system can upgrade its performance by using different feature extraction techniques, feature optimization techniques and diverse mixture of machine learning techniques [4] [7]. The application of Deep Neural Network (DNN) in the field of natural language processing, image processing, and sentiment analysis and video recognition has shown tremendous improvement in performance. Hence our future work is to implement deep neural networks in predicting the CHD which will further enhance the results [2]. Also, we are concentrating on working with real-world datasets.

References

[1] Lewellyn L. R. Rodrigues, Dasharathraj K Shetty, Nitesh Naik, Chethana Balakrishna Maddodi, Anuradha Rao, Ajith Kumar Shetty, Rama Bhat &Zeeshan Hameed," Machine learning in coronary heart disease prediction: Structural equation

modeling approach" Rodrigues et al., Cogent Engineering ISSN: (Print) 2331-1916 (Online) pp.1-15https://doi.org/10.1080/23311916.2020.1723198 February 2020

[2] Shiyue Cui1, Chao Li1, Zhe Chen1, Jiao Jiao Wang1, and Juxiang Yuan Research on Risk Prediction of Dyslipidemia in Steel Workers Based on Recurrent NeuralNetwork and LSTM Neural Network" S. Cui *et al.*, IEEE Access vol. 8 *pp.* 34153- 34160 *Digital Object Identifier 10.1109/ACCESS.2020.2974887* February 2020

[3] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (Ph.D.)"Prediction of heart disease using machine learning" IEEE Xplore ISBN: 978-1-5386-0965-1 1275- 1278

[4]M. Thiyagaraj and G. Suseendran "Enhanced Prediction of Heart Disease Using Particle Swarm Optimization and Rough Sets with Transductive Support Vector Machines Classifier Springer Naturehttps://doi.org/10.1007/978-981-13-9364-8_11 pp.141-151 2020

[5] Yogita Solanki,Sanjiv Sharma," A Survey on Risk Assessments of Heart Attack" Using Data Mining Approaches I.J. Information Engineering and Electronic Business, 2019, DOI: 10.5815/ijieeb.2019.04.05 vol. 4, pp.43-51 July 2019

[6] E. Punarselvam, Mohamed Yacin Sikkandar, ·Mohsen Bakouri, N. B. Prakash and T.Jayasankar and S.Sudhakar," Different Loading Condition and Angle Measurement of Human Lumbar Spine MRI Image Using ANSYS", Springer Journal of Ambient Intelligence and Humanized Computing (2020), ISSN: 1868-5137 , (Print) 1868-5145 (Online)
https://doi.org/10.1007/s12652-020-01939-7

[7] Liaqat Ali 1,2, Atiqur Rahman3, Aurangzeb Khan3, Mingyi Zhou1,Ashir Javeed4, And Javed Ali Khan5,6" An Automated Diagnostic System for Heart Disease Prediction Based on _2 Statistical Model and Optimally Configured Deep Neural Network ", Digital Object Identifier 10.1109/ACCESS.2019.2904800, IEEE Access vol.7 pp. 34938- 34945February 2019

[8] Liaqat Ali, Awais Niamat, Javed Ali Khan, Noorbakhsh Amiri Golilarz, Xiong Xingzhong, Adeeb Noor, Redhwan Nour, and Syed Ahmad Chan Bukhari" An Optimized Stacked Support Vector Machines Based Expert System For The Effective Prediction Of Heart Failure" L. Ali et al.: IEEE Access Digital Object Identifier 10.1109/ACCESS.2019.2909969, vol.7, pp. 54007- 54013 March 2019

[9] Ankur Gupta, Rahul Kumar 1, Harkirat Singh Arora, And Balasubramanian Raman 1, (Member, IEEE)," MIFH: A Machine Intelligence Framework For Heart Disease Diagnosis", IEEE Access, Digital Object Identifier 10.1109/ACCESS.2019.2962755vol.8, pp. 14659-14674, January 2020

[10]Tom M.Mitchell,"Machine Learning" publication McGraw-Hill Science ISBN 0070428077

[11] Zahia Aouabed, Moloud Abdar(&), Nadia Tahiri, Jaël Champagne Gareau, and Vladimir Makarenkov," A Novel Effective Ensemble Model for Early Detection of Coronary Artery Disease Springer Nature https://doi.org/10.1007/978-3-030-36778-7_53pp.480-489 2020

[12] Abderrahmane Ed-daoudy***,** Khalil Maalmi," Real-time machine learning for early detection of heart disease using big data approach IEEE ISSN: 978-1-5386-7850-3 vol.19**,** 2019

[13] Mohd Ashraf1, M. A. Rizvi2 and Himanshu Sharma3,"Improved Heart Disease Prediction Using Deep Neural Network", Asian Journal of Computer Science and Technology, ISSN: 2249-0701 Vol.8 No.2, 2019, pp. 49-54 April 2019

[14] Bo Jin 1, (Senior Member, IEEE), Chao Che2, Zhen Liu3, Shulong Zhang4,, "Predicting The Risk Of Heart Failure With EHR Sequential Data Modeling", Special Section On Recent Computational Methods In Knowledge Engineering and Intelligence Computation, Digital Object Identifier 10.1109/ACCESS.2017.2789324 ,vol.6,pp. 9256-9260,March 2018

[15] P.Shanmugapriya, V.Mohan, T.Jayasankar, Y.Venkataramani, "Deep Neural Network based Speaker Verification System using Features from Glottal Activity Regions", Appl. Math. Inf. Sci. vol.12, no.6, Nov 2018, pp.1147–1155.
DOI: http://dx.doi.org/10.18576/amis/120609

[16]Jingyi Wang, Therese Tillin, Alun D. Hughes, Nish Chaturvedi, "Associations between family history and coronary artery calcium and coronary heart disease in British Europeans and South Asians Elsevier https://doi.org/10.1016/j.ijcard.2019.07.101 August 2019

[17]Prof Kivimäki et al." Long working hours and risk of coronary heart disease and stroke: asystematic review and meta-analysis of published and unpublished data for 603838 individuals" Elsevier Open Access article October 2015.

[18] Stephe.S, T.Jayasankar, "Motor Imagery Recognition of EEG Signal using Cuckoo-Search Masking Empirical Mode Decomposition", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.8, Issue-11, September 2019, pp.2717-2720.
DOI: 10.35940/ijitee.K2175.0981119

[19]Sangeetha .J, Jayasankar.T , "Emotion Speech Recognition based on Adaptive Fractional Deep belief Network and Reinforcement Learning, Springer- Advances in Intelligent Systems and Computing - International Conference on Cognitive Informatics & Soft Computing (CISC-2017),pp-167-174.
https://doi.org/ 10.1007/978-981-13-0617-4_16 .

[20] J.Sangeetha,T.Jayasankar," A Novel Whispered Speaker Identification System Based on Extreme Learning Machine", International Journal of Speech Technology, Springer,(2018) 21 (1), pp.157–165.
DOI: https://doi.org/10.1007/s10772-017-9488-z