

Heart Disease Prediction using Machine Learning and Data Mining



Keshav Srivastava, Dilip Kumar Choubey

Abstract: Weighing only 300 grams, Heart is declining the mortality rate at a rapid pace from decades. The major factors that contribute to it are smoking, drinking, unbalanced diet, and many more. Even with these more technical advancements the analysis of the clinical data is a critical challenge. With the use of Machine Learning techniques, it is possible to analyse the data and interpret the cause that lead to heart diseases such as Coronary Heart Disease, Arrhythmia, and Dilated Cardiomyopathy. Many researchers are developing IoT enabled hardware to predict these diseases using various ML and DM techniques. In this study, we propose a novel method to determine the disease using Cleveland Heart Disease Dataset by combining the computational power of various ML and DM algorithms and concluded that among all the algorithms, K-Nearest Neighbors gives the highest accuracy of 87%. Along with this, a web app is developed using flask in python with which the user can enter the attributes and predict the heart disease.

Keywords: Classification, Data Mining, Decision Trees, Heart Disease, KNN, Logistic Regression, Machine Learning, Naïve Bayes, Random Forest, SVM.

I. INTRODUCTION

A lot of factors contribute to heart disease like cholesterol level, blood pressure, smoking or drinking habits, pulse rate, age, diabetes, and many more. Over the past years, many machine learning, data mining and neural network approaches are adopted to identify the disease. Various Machine Learning models like K-Nearest Neighbour, Naïve Bayes, Decision Tree, Support Vector Machines are introduced to determine the threat level these disease possess but the data from the last few decades define that many person are having these disease at an early stage and even some new born children are suffering or died from a heart disease. Machine Learning aspects can play a major role in prediction of these disease and the threat level it possesses.

Data Mining and Machine Learning methods like Decision Tree have been used predict these disease [11][16] by calculating the accuracy on a dataset.

Researchers Choubey et al. (2016, 2017, 2018, 2019, 2020) [3, 4, 17, 18, 25, 26, 29, 30] and Bala et al. (2017, 2018) [19, 35] have briefly summarized of many soft computing, data mining and machine learning methods for the classification of diabetes and thunderstorm respectively. We proposed various approaches like Decision Tree, Naïve Bayes Classifier, K-Nearest Neighbour, Random Forest, Artificial Neural Network, and Support Vector Machines on Cleveland Heart Disease dataset to train the model. These trained models are then used as packages to be uploaded on a web server. Now using Flask framework in python, a front end is created on which the various attributes of the dataset is taken as input from the user and this input will be redirected to the model packages for testing and the result will be displayed to the user.

This paper has been organized as follows: Related Work is presented in Section II, Dataset Description is discussed in section III, Proposed Methodology in section IV, Experimental Result and Discussion in section V and Conclusion and Future Work in section

II. RELATED WORK

Anitek et al. (2017) ([3]) developed a portable and real-time hardware for the classification of Heart Auscultation in which Feature Analysis and Supervised Classification is used to classify the heart rate as normal or abnormal. Statistical features like Mean and Median are also used along with SVM Classifier with linear kernel to classify the sounds with which the authors achieved an accuracy of 97.50% and 100% sensitivity.

Sneha et al. (2018) ([6]) proposed a system for the analysis of Echocardiogram in which an input video is fed to the system and then the Frame Extraction, ROI Selection, Feature Extraction and, SVM Classifier is used for training. The overall accuracy of the system is 98.30% of which Normal disorder gives 99.53%, ASD disorder gives 97.57% and DCM gives 97.40%.

Martin et al. (2017) ([7]) gives a system for the detection of Chronic Heart Failure using Filtering, Segmentation, Feature Extraction and some machine learning algorithms like Boosting, Support Vector Machines, k-Nearest Neighbours, Naïve Bayes classifier, J48 and Random Forest. LOSO cross validation evaluation is used in the experimental process and the model attained an accuracy of 96%.

Raid et al. (2017) ([8]) proposed a heart disease recommendation system on MATLAB in which Fast Fourier Transform is used to determine the properties of time series. Along with this an Ensemble model constituting Bagging, Artificial Neural Network, Least Square-SVM, and Naïve Bayes Classifier is used.

Manuscript received on April 04, 2020.

Revised Manuscript received on April 13, 2020.

Manuscript published on May 30, 2020.

* Correspondence Author

Keshav Srivastava*, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India. Email: keshav.srivastava1981@gmail.com, keshav.srivastava2018@vitstudent.ac.in

Dilip Kumar Choubey*, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India. Email: dilipchoubey_1988@yahoo.in, dilip.choubey@vit.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The accuracy provided by the system in Heart Rate is 94.80%, in DBP 94.85%, in MAP 94.20%, and in SO2 95.50% using an eight-feature set. The overall accuracy is 94.83%.

Qingxue et al. (2017) ([9]) prepared a hardware system for Cardiac Health Monitoring by using Support Vector Machine algorithm for identification of heart beat, an Unsupervised Learning approach for purification of heart beat and a Regression Model for determining the chest ECG.

Zhao et al. (2017) ([10]) developed a system for Heart Disease Classification by using two datasets, one from Shanghai Shuguang Hospital and another is UCI heart disease dataset. The model uses Support Vector Machine algorithm along with PCA, CCA and DMPCCA which are used for feature extraction and fusion. The overall study resulted that DMPCCA provided the best result.

Ganesan et al. (2019) ([11]) makes the use of IOT technology for prediction and diagnosis of heart diseases by taking UCI dataset and applied J48 Classifier, Logistic Regression, Multilayer Perceptron, and SVM using Java on Amazon Cloud. In this study J48 gave 91.48%, SVM gave 84.07%, LR gave 83.70%, and MLP gave 78.14% accuracy and concluded that J48 outperforms all other algorithms.

Berina et al. (2017) ([12]) proposed a model for the classification of Diabetes and Cardiovascular Disease using Artificial Neural Network's Multilayer Feed Forward Neural Network with Levenberg-Marquardt learning along with Naïve Bayesian Network such that the model gave an accuracy of 99.51% for diabetes and 97.92% for cardiovascular disease. The authors also concluded that ANN shows better mean accuracy.

Saba et al. (2019) ([13]) developed a Heart Disease Prediction system on Rapid Miner Tool using Decision Tree, Logistic Regression, Logistic Regression SVM, Naïve Bayes, and Random Forest on UCI heart disease dataset. The model showed 82.22% for Decision Tree, 82.56% for Logistic Regression, 84.85% for Logistic Regression SVM, 84.24% for Naïve Bayes, and 84.17% for Random Forest and concluded that Logistic Regression SVM is best.

Ricardo et al. (2019) ([14]) created a model for Heart Disease Detection on V.A. Medical Centre dataset by using Random Forest algorithm in which the authors divided the people as per the age group and the model showed an accuracy of 84.448%.

Hasan et al. (2019) ([15]) provided a model of Heart Disease Prediction on StatLog dataset. The model constituted Sparse Discriminant Analysis, Bagging, Random Forest, and Linear Discriminant Analysis and showed an accuracy of 96%, 76%, 80%, and 89% and concluded that Sparse Discriminant Analysis is better algorithm for prediction.

Prerna et al. (2019) ([16]) proposed a model for Heart Disease Prediction on Fingertip Video Dataset and Heart Disease Dataset from UCI Repository using Modified Artificial Plant Optimization along with several libraries like OpenCV, NumPy, Matplotlib, SciKit-Learn, Keras, TensorFlow, SciPy, and XGBoost and the accuracies attained by ANN, XGBoost, Naïve Bayes and Logistic Regression are 87%, 85.10%, 89.5%, and 89% after applying APO which concludes that MAPO is better.

Rui et al. (2019) ([20]) designed a system which is helpful for prediction of Cardiovascular Diseases. The authors divided the disease into levels: Normal, Mild and Moderate and using Naïve Bayes classifier the system predicted that

57% of the population are in Normal level, 27% are in Mild level and 16% are in Moderate level.

Samuel et al. (2019) ([21]) uses the CARES Registry Dataset to predict Heart Disease. The model used k-Nearest Neighbour, EFCN, Decision Tree, Random Forest, and LightGBM of which the sensitivity obtained by LightGBM is 67%, Random Forest is 68%, Decision Tree is 70%, k-Nearest Neighbour is 70%, and ECNF is 82.50%.

Nahian et al. (2019) ([22]) uses PTB Diagnostic ECG Dataset, MIT-BIH Arrhythmia Dataset, St. Petersburg Institute of Cardiological Technique's 12-Lead Dataset and applied Convolution Neural Network Algorithm which attained an accuracy of 98.24% in PTB Diagnostic ECG Dataset, 97.70% in MIT-BIH Arrhythmia Dataset, 99.71% in St. Petersburg Institute of Cardiological Technique's 12-Lead Dataset.

Sabrina et al. (2019) ([23]) took Italian and American Datasets for the prediction of Cardiovascular Disease on which the authors applied Support Vector Machine with RBF Kernel Algorithm and optimized the output using GridSearch. The model attained an accuracy of 95.25% in Italian Dataset and 92.15% in American Dataset.

Ashok et al. (2018) ([24]) proposed a performance evaluation method for prediction of Heart Disease on StatLog Heart Disease Dataset using Classification Tree, k-Nearest Neighbour, Logistic Regression, Naïve Bayes Classifier, SVM, and Artificial Neural Network whose accuracies are 77%, 80%, 80%, 83%, 82%, and 84% thus concluding that ANN is better.

Gunasekaran et al. (2018) ([27]) designed a recommendation system for Heart Disease on KEGG Metabolic Reaction Network Dataset using PCA, LDA, MKL, LS, GDA with SVM and ANFIS. The algorithms are applied in combinations and the model attained 98% Sensitivity, 99% Specificity and 0.01% Less Mean Square.

Subhashini et al. (2019) ([31]) designed a prediction system for Chronic Heart Disease by combining Artificial Neural Network, Naïve Bayes Classifier and Support Vector Machine as an ensemble model. Bagging is then applied on this ensemble model using Fast Fourier Transform method thus making the combined model attain 93% overall accuracy.

Sangeeta et al. (2018) ([32]) used Cleveland Heart Disease Dataset for the prediction of Cardiovascular Disease Prediction. The Rough Set Classifier is used to classify the data i.e. the maximum accuracy attained by Rough Set Classifier 1 is 91.65% and Rough Set Classifier 2 is 89.75% along with this the model was able to handle missing values.

Costa et al. (2019) ([33]) proposed a system for the Detection of Cardiovascular Heart Disease on Cleveland Heart Disease Dataset using Artificial Neural Network on Amazon EC2 Cloud Server. Amazon EC2 Cloud server provides a platform for faster computation which normal PC fails to provide. The model attains an overall accuracy of 90.74%.

Chandra et al. (2018) ([34]) designed a system for Heart Disease Diagnosis using Cleveland Heart Disease Dataset on MATLAB. Genetic algorithm is used for feature selection and Support Vector Machine is used for classification; in this way the model achieved an overall accuracy of 88.34%.

III. DATASET DESCRIPTION

We have used the Cleveland Heart Disease Dataset [2] from the UCI Machine Learning repository. The dataset consists of 303 records of which 6 records have missing values. The

Table 1 Dataset Description

S. No.	Attribute	Description	Type
1	Age	Age of patient (in years)	Numerical
2	Sex	Gender of patient (divided into two categories: 1 for male and 0 for female)	Nominal
3	Cp	Type of chest pain (divided into four categories: 1 for typical angina, 2 for atypical angina, 3 for non-anginal pain, and 4 for asymptomatic)	Nominal
4	Trestbps	Resting blood pressure (in mm Hg)	Numeric
5	Chol	Cholesterol level (in mg/dl)	Numeric
6	FBS	Blood sugar level at fasting mode (empty stomach) (divided into two categories: 1 for true and 0 for false)	Nominal
7	RestECG	Electrocardiographic (ECG) results at rest	Numeric
8	Thalach	The maximum heart rate attained by patient	Numeric
9	Exang	Exercise induced angina (divided into two categories: 1 for yes and 0 for no)	Nominal
10	Oldpeak	ST depression induced angina	Numeric

Fig. 1 determines the instances having heart disease or not. Here the samples in green column having value 0 don't have any heart disease whereas the samples on red column having value 1 have heart disease. The actual dataset have 5 classes in target attribute of which the values 1, 2, 3, 4 represents different heart disease which is combined as one i.e. the red class consists of the 4 classes. The dataset [2] has been divided training and testing i.e. 70% is used for training and 30% for testing i.e. 212 samples are used for testing and 91 samples are used for training.

dataset has numerical and categorical values. In pre-processing phase, the missing values have been removed and from the 297 patients 137 have heart disease and the remaining 160 don't have heart disease. The dataset contains 14 among which the last one is target class attribute.

		relative to rest state	
11	Slope	Peak exercise ST segment slope (divided into three categories: 1 for upsloping, 2 for flat and 3 for downsloping)	Nominal
12	Ca	Number (0-3) of major vessels colored by fluoroscopy	Numeric
13	Thal	Heart status (divided into three categories: 3 for normal, 6 for fixed defect, 7 for reversible defect)	Nominal
14	Num	Heart disease diagnose of narrowing diameter (divided into two categories: 0 for <50% and 1 for >50%)	Nominal

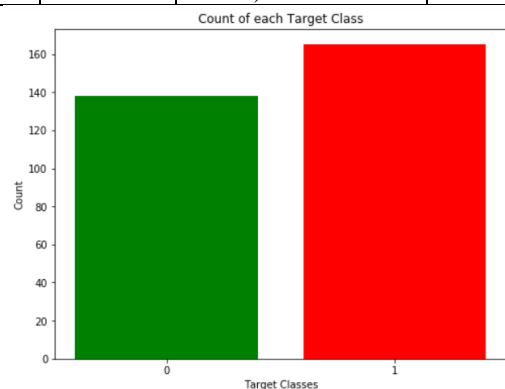


Fig. 1 Count of each Target Class

IV. PROPOSED METHODOLOGY

The attributes and samples from the dataset are pre-processed to rectify the null values after that various Machine Learning algorithms are applied on the dataset to make the predictions as shown in Fig. 2. The detailed description of the algorithms used in the model are given below:

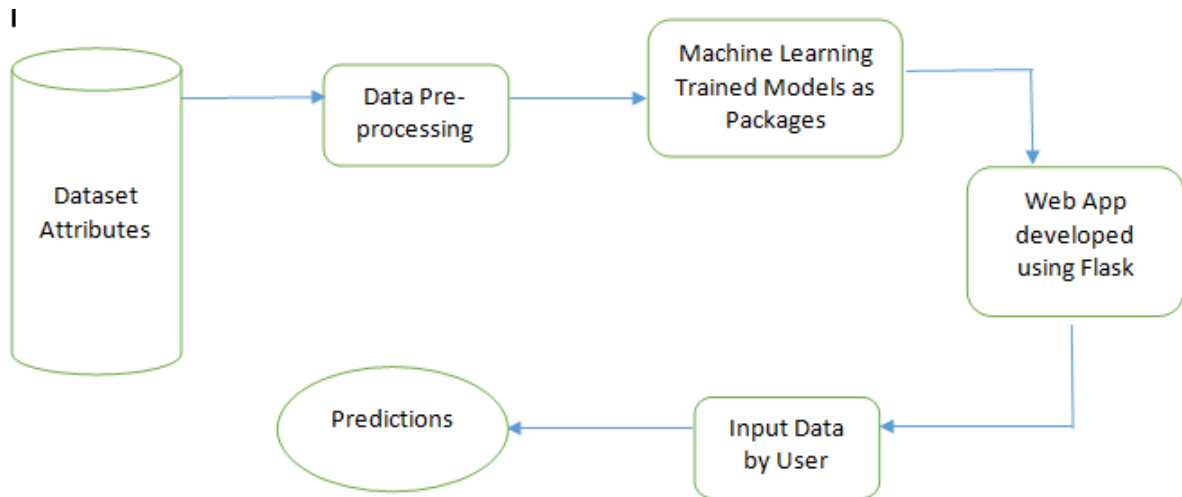


Fig. 2 Proposed Architecture

A. Decision Trees

Decision Tree works on the high entropy of input samples. It is a Divide and Conquer (DAC) technique in which the trees are constructed in a top-down approach. To remove the non-significant samples tree pruning is performed on the training samples. The entropy can be calculated by using:

$$DT_Entropy = -\sum_{z=1}^n p_{ij} \log_2 p_{ij}$$

B. K-Nearest Neighbour

It uses Euclidian Distance function $dis(x_i, x_j)$ to extract the information out of the samples by using the majority vote of the k-nearest neighbours.

$$Dis(x_i, x_j) = \sqrt{(x_{i,1} - x_{j,1})^2 + \dots + (x_{i,m} - x_{j,m})^2}$$

C. Support Vector Machine

$f(x) = w^T x + b$ is the equation of hyperplane which SVM locates by using the samples. This hyperplane is also called

the optimal solution. To find the hyperplane some mathematical computation is done:

$$\text{Min}_{w,b,E_i} \frac{1}{2} w^2 + C \sum_{i=1}^n E_i$$

Where, $y_i (w^T x_i + b) \geq 1 - E_i$, $E_i \geq 0$,

'w' is Dimensional Coefficient and 'b' is offset

D. Random Forest

This is a bagging technique which classifies the samples by building several decision trees to find the optimal result. For a given dataset D , $X = \{x_1, x_2, x_3, \dots, x_n\}$ with target values $Y = \{y_1, y_2, y_3, \dots, y_n\}$ where bagging is performed from $i = 1$ to N . to find the unseen samples we take the average of the predictions on x , i.e.

$$y = \frac{1}{N} \sum_{i=1}^N f_b(x)$$

the error is derived by calculating the standard deviation of the trees

$$S.D. = \left(\frac{\sum_{i=1}^N (f_b(x) - f)^2}{N-1} \right)^{1/2}$$

Algorithm 1 Decision Tree Classifier

Input: m samples from dataset with target classes

for all attributes **do**
 for each record **do**
 Perform Decision Tree Classifier algorithm
 end for
 classify the attribute space
end for

determine the total leaf nodes $n_1, n_2, n_3, \dots, n_m$
 divide the samples into $m_1, m_2, m_3, \dots, m_m$ according to the leaf nodes

Output: Partitioned samples $m_1, m_2, m_3, \dots, m_m$

Algorithm 2 K-Nearest Neighbour

Input: m samples from dataset with target classes

1. Calculate Euclidean Distance $dis(x, x_i)$ and arrange it in ascending order.
2. Take the first value from Step 1 and name it **k**.
3. Find the distance corresponding to the k-points.
4. If $k_j > k_i$, where $j \neq i$ then put x in class j .

Algorithm 3 Support Vector Machine

Input: m samples from dataset with n target classes $\alpha \leftarrow 0$ or $\alpha \leftarrow$ partially trained SVM

1. $K \leftarrow$ Consider some value
2. Repeat
3. **for all** $\{m_i, n_i\}, \{m_j, n_j\}$ **do**
4. Optimize α_i and α_j , where $\alpha > 0$



```

5.         if  $\alpha$  changes
6.         break
    
```

Algorithm 4 Random Forest

```

Input: m samples from dataset with n target classes
H: hypothesis
I: Iterations
1. for all m
2.     Do  $a_m \leftarrow 0$ 
3.     for all m
4.         for all I
5.              $x_1 \leftarrow$  random sample from m
6.              $p \leftarrow H(x_1)$ 
7.              $x_2 \leftarrow$  random sample from m
8.              $x_1[m] \leftarrow x_2[m]$ 
9.             if  $H(x_1) \neq p$  then
10.                 $a_m \leftarrow a_m + 1$ 
    
```

Table 2 Confusion Matrix of used Algorithms

DT	k-NN					SVM					RF								
39	5	3	1	1	46	2	1	0	0	47	0	2	0	0	44	3	2	0	0
10	5	2	6	0	17	4	1	1	0	17	2	0	4	0	15	2	2	4	0
5	3	2	3	1	5	7	1	1	0	8	0	0	6	0	6	0	5	3	0
3	2	2	2	1	3	4	2	1	0	3	1	0	6	0	4	2	1	2	1
0	0	2	2	0	1	1	1	1	0	1	1	0	2	0	0	1	0	3	0

Here Table 2 consists of the confusion matrix obtained by various algorithm. Here the five columns represent the different types of heart disease labelled as 0, 1, 2, 3, 4 where 0 represents no heart disease and the values 1, 2, 3, 4 represents that the person has heart disease.

V. EXPERIMENTAL RESULT AND DISCUSSION

In this study, Jupyter Notebook is used to apply the algorithms on the Cleveland Heart Disease Dataset. Here, the diagnostic performance is evaluated in terms of accuracy, precision, sensitivity, specificity, f1-score, ROC. The factors contributing to these are discussed below: -

A. True Positive

In this scenario, the algorithm predicted true and the actual result is also true.

B. True Negative

In this scenario, the algorithm predicted false and the actual result is also false.

C. False Positive

In this scenario, the algorithm predicted true and the actual result is also false.

D. False Negative

In this scenario, the algorithm predicted false and the actual result is also true.

However, accuracy, sensitivity/recall, specificity, precision and f1-score can be calculated using these four outcomes, i.e.

$$\text{Accuracy} = (\text{True Negative} + \text{True Positive}) / (\text{True Negative} + \text{False Negative} + \text{False Positive})$$

$$\text{Sensitivity} = \text{True Positive} / (\text{True Positive} + \text{False Negative})$$

$$\text{Specificity} = \text{True Negative} / (\text{True Negative} + \text{False Positive})$$

$$\text{Precision} = \text{True Positive} / (\text{True Positive} + \text{False Positive})$$

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Table 3 Different Measures of the algorithms

Algorithm	Accuracy	Sensitivity/Recall	Specificity	Precision	F1- Score	ROC
DT	71	73	69	84	70	71.1
KNN	87	73	86	83	77	88.5
SVM	84	75	82	80	77	90.4
RF	83	79	86	68	81	90.8

Table 3 illustrates the different measures like accuracy, sensitivity/recall, specificity, precision, f1-score, and ROC area under the curve of the algorithms.

In Table 4, the classification accuracy of proposed algorithms and various other classifiers ([1]) are mentioned.



Table 4 Classification accuracy of proposed algorithms and other classifiers

Source	Algorithm	Accuracy %
Our Study	k-Nearest Neighbours	87
Our Study	Decision Tree	79
Our Study	Support Vector Machine	83
Our Study	Random Forest	84
WD/KG	Neural Network	85.1
Ster, Dobniker	LDA	84.5
Rafal, Ster, Dobniker	Naïve Bayes	82.5
Bennet and Blue	Decision Tree and Cross Validation	82.5
Karol Grudzinski	k-Nearest Neighbours	82.1
Bennet and Blue	Support Vector Machine and Cross Validation	81.5
Zarndt	C4.5	53.8
Rafal, Ster, Dobniker	Multi-Layer Perceptron	81.3
Ster, Dobniker	k-Nearest Neighbours	81.5

The various accuracy graph of Decision Tree, K-Nearest Neighbour, Support vector Machines, and Random Forest are shown in Fig. 3, 4, 5, and 6 respectively. The various ROC curve of Decision Tree, K-Nearest Neighbour, Support vector Machines, and Random Forest are shown in Fig. 7, 8, 9, and 10.

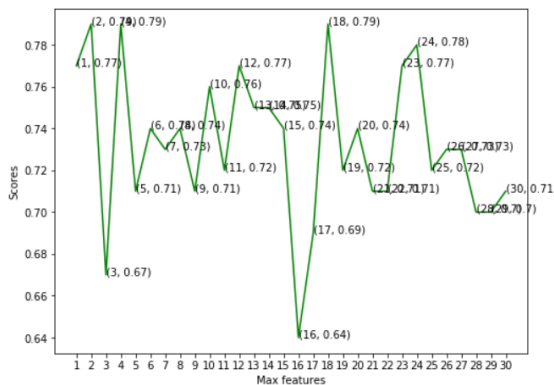


Fig. 3 Decision Tree Classifier Score for different number of maximum features

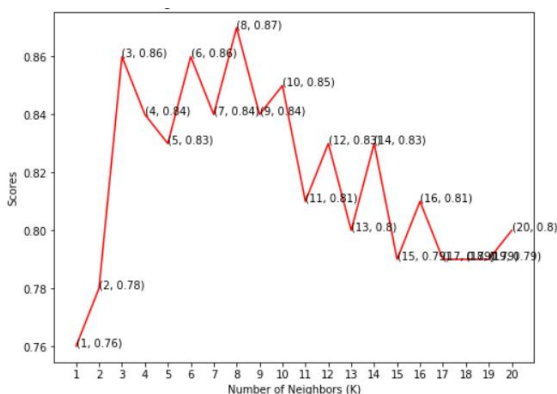


Fig. 4 K-Nearest Neighbour Score for different K Values

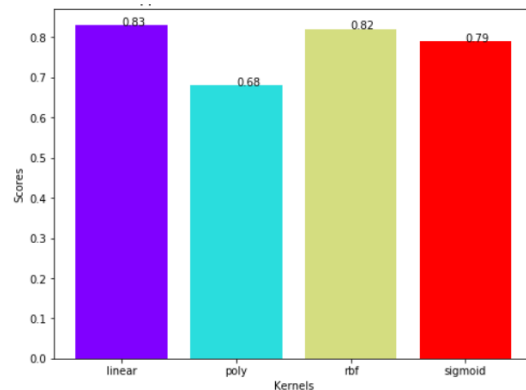


Fig. 5 Support Vector Classifier Scores for different Kernels

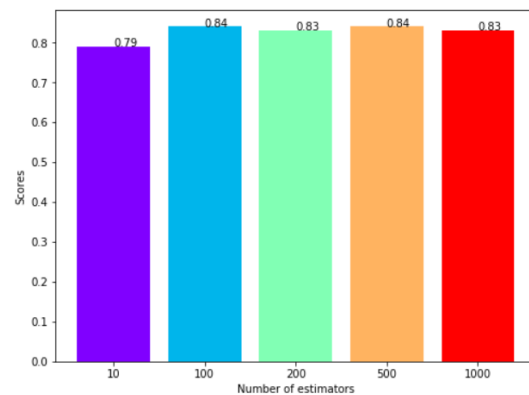


Fig. 6 Random Forest Score for different number of estimators

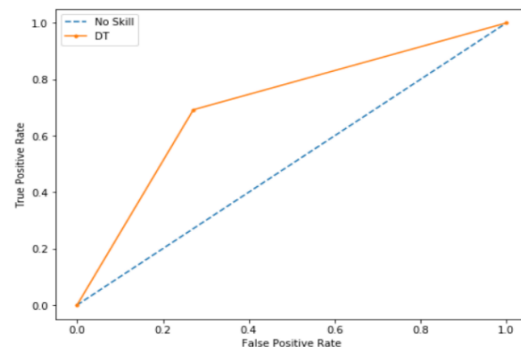


Fig. 7 ROC Curve of RF

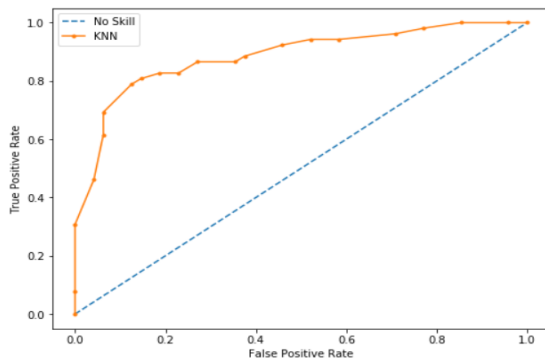


Fig. 8 ROC Curve of KNN

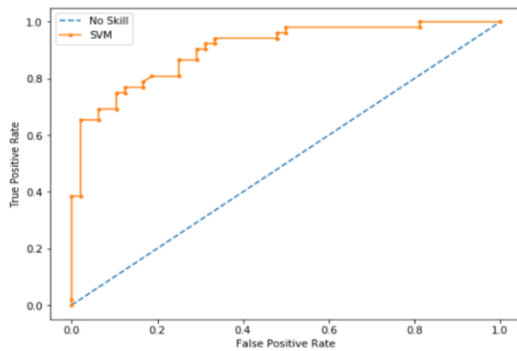


Fig. 9 ROC Curve of SVM

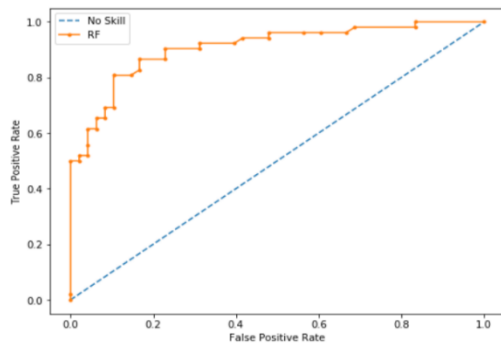


Fig. 10 ROC Curve of RF

VI. CONCLUSION AND FUTURE WORK

Determining any heart disease on some raw data is really difficult for even a doctor which is why many healthcare sectors are opting for machine learning techniques to determine it. In our experimental study we took the Cleveland Heart Disease dataset obtained from UCI repository and applied pre-processing to drop the data with missing values and applied some algorithm like Decision tree, K-Nearest Neighbour, Support Vector Machines and, Random Forest of which Decision Tree gives an accuracy of 79% as shown in Fig. 3, K-Nearest Neighbour gives an accuracy of 87% as shown in Fig. 4, Support vector Machines gives an accuracy of 83% as shown in Fig. 5, and Random Forest Gives an accuracy of 84% as shown in Fig. 6. The ROC curve for Decision Tree, K-Nearest Neighbour, Support vector Machines, and Random Forest gives an AUC of 71.6%, 88.5%, 90.4%, 90.8% as shown in Fig. 7, 8, 9, and 10 respectively.

Now these models with these accuracies are converted into packages using pickle library in python. Using flask, a web app is created where these packages are used to do prediction on the data entered by the user. For future researchers, the accuracies can be improved by using some

data mining techniques which can extract the hidden information from the samples.

REFERENCES

- Humar Kahramanli, Novruz Allahverdi, "Design of a hybrid system for the diabetes and heart diseases", Expert Systems with Applications, Elsevier, pp: 82–89, 2008.
- Dataset: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- Anitek Bhattacharya, Mohan Mishra, Anushikha Singh & Malay Kishore Dutta, "Machine Learning Based Portable Device for Detection of Cardiac Abnormality" International Conference on Emerging Trends in Computing and Communication Technologies (ICETCCT), IEEE, pp: 1-4, 2017.
- Dilip Kumar Choubey, Sudhakar Tripathi, Prabhat Kumar, Vaibhav Shukla, Vinay Kumar Dhandhanian, "Classification of Diabetes by Kernel based SVM with PSO", Recent Advances in Computer Science and Communications, Bentham Science, Vol. 12, No. 1, pp. 1-14, 2019.
- Dilip Kumar Choubey, Prabhat Kumar, Sudhakar Tripathi, Santosh Kumar, "Performance Evaluation of Classification Methods with PCA and PSO for Diabetes", Network Modeling Analysis in Health Informatics and Bioinformatics, Springer, Vol. 9, Issue 1, Article 5, pp. 1-30, 2019.
- Sneha Borkar, Prof. M. N. Annadate, "Supervised Machine Learning Algorithm for Detection of Cardiac Disorders" Fourth International Conference on Computing Communication Control and Automation (ICCUCEA), IEEE, 2018.
- Martin Gjoreski, Monika Simjanoska, Anton Gradisek, Ana Peterlin, Matjaz Gams, Gregor Poglajen, "Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine Learning Classifiers" International Conference on Intelligent Environments (IE), IEEE, pp: 14-19, 2017.
- Raid Lafta, Ji Zhang, Xiaohui Tao, Yan Li, Xiaodong Zhu, Yonglong Luo and Fulong Chen, "Coupling a Fast Fourier Transformation with a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment" IEEE, pp:10674-10685, 2017.
- Qingxue Zhang, Dian Zhou, Xuan Zeng, "Hear the Heart: Daily Cardiac Health Monitoring Using Ear-ECG and Machine Learning" 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), IEEE, pp: 448-451, 2017.
- TING-TING ZHAO, YU-BO YUAN, YING JIE WANG, JU GAO, PING HE, "HEART DISEASE CLASSIFICATION BASED ON FEATURE FUSION" International Conference on Machine Learning and Cybernetics (ICMLC), IEEE, vol. 1, pp: 111-117, 2017.
- M.Ganesan, Dr.N.Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models" International Conference on System, Computation, Automation and Networking (ICSCAN), IEEE, pp: 1-5, 2019.
- Berina Ali, Lejla Gurbetal, Almir Badnjevi, "Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases" 6th Mediterranean Conference on Embedded Computing (MECO), IEEE, pp: 1-4, 2017.
- Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, Khurram Bashir, "Improving Heart Disease Prediction Using Feature Selection Approaches" 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), IEEE, pp: 619-623, 2019.
- Ricardo Buettner, Marc Schunter, "Efficient machine learning based detection of heart disease" Proc. IEEE Healthcom, 2019.
- K. M. Zubair Hasan, Shourab Datta, Md Zahid Hasan, Nusrat Zahan, "Automated Prediction of Heart Disease Patients using Sparse Discriminant Analysis" International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, pp: 1-6, 2019.
- Perna Sharma, Krishna Choudhary, Kshitij Gupta, Rahul Chawla, Deepak Gupta, Arun Sharma, "Artificial Plant Optimization Algorithm to detect Heart Rate and Presence of Heart Disease using Machine Learning" Artificial Intelligence in Medicine, Elsevier, pp:101752, 2019.
- Dilip Kumar Choubey, Sanchita Paul, "Classification techniques for diagnosis of diabetes: a review", International Journal of Biomedical Engineering and Technology (IJBET), Inderscience, Vol. 21, No. 1, pp. 15-39, 2016.



18. Dilip Kumar Choubey, Sanchita Paul, "GA_MLP NN: A Hybrid Intelligent System for Diabetes Disease Diagnosis", International Journal of Intelligent Systems and Applications (IJISA), MECS, Vol. 8, No. 1, pp. 49-59, 2016.
19. Kanchan Bala, Dilip Kumar Choubey, Sanchita Paul, "Soft Computing and Data Mining Techniques for Thunderstorms and Lightning Prediction: A Survey", International Conference of Electronics, Communication and Aerospace Technology (ICECA 2017), IEEE Xplore Digital Library, Vol. 1, pp. 42-46, 2017.
20. Rui Chen, Aijia Lu, Jingjing Wang, Xiaohai Ma, Lei Zhao, Wanjia Wu, Zhicheng Du, Hongwen Fei, Qiongwen Lin, Zhuliang Yuf, Hui Liu "Using machine learning to predict one-year cardiovascular events in patients with severe dilated cardiomyopathy" European Journal of Radiology, Elsevier, 2019.
21. Samuel Harford, Houshang Darabi, Marina Del Rios, Somshubra Majumdar, Fazle Karim, Terry Vanden Hoek, Kim Erwin, Dennis P. Watson, "A machine learning based model for Out of Hospital cardiac arrest outcome classification and sensitivity analysis" Resuscitation 138, Elsevier, pp: 134-140, 2019.
22. Nahian Ibn Hasan, Arnab Bhattacharjee, "Deep Learning Approach to Cardiovascular Disease Classification Employing Modified ECG Signal from Empirical Mode Decomposition" Biomedical Signal Processing and Control 52, Elsevier, pp: 128-140, 2019.
23. [23] Sabrina Mezzatesta, Claudia Torino, Pasquale De Meo, Giacomo Fiumara, Antonio Vilasi, "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis" Computer Methods and Programs in Biomedicine 177, Elsevier, pp: 9-15, 2019.
24. Ashok Kumar Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease" Neural Computing and Applications 29, Springer, pp: 685-693, 2018.
25. Dilip Kumar Choubey, Sanchita Paul, Smita Sandilya, Vinay Kumar Dhandhanian, "Implementation and Analysis of Classification algorithms for Diabetes", Current Medical Imaging Reviews, Bentham Science (In Press 2018) Dilip Kumar Choubey, Manish Kumar, Vaibhav Shukla, Sudhakar Tripathi, Vinay Kumar Dhandhanian, "Comparative Analysis of Classification Methods with PCA and LDA for Diabetes", Current Diabetes Reviews, Bentham Science (In Press 2020).
26. Dilip Kumar Choubey, Sanchita Paul, Smita Sandilya, Vinay Kumar Dhandhanian, "Implementation and Analysis of Classification algorithms for Diabetes", Current Medical Imaging, Bentham Science, Vol. 16, Issue 4, pp. 340-354, 2020.
27. Gunasekam Manogaran, R. Varatharajan, M. K. Priyan, "Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System" Multimedia tools and applications 77, Springer, pp: 4379-4399, 2018.
28. Subhashini Narayan, E. Sathiyamoorthy, "A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases" Neural Computing and Applications 31, Springer, pp: 93-102, 2019.
29. Dilip Kumar Choubey, Sanchita Paul, Vinay Kumar Dhandhanian, "Rule Based Diagnosis System for Diabetes", Biomedical Research, Allied Academies, Vol. 28, Issues 12, pp. 5196-5209, 2017.
30. Dilip Kumar Choubey, Sanchita Paul, "GA_RBF NN: A Classification System for Diabetes", International Journal of Biomedical Engineering and Technology (IJBET), Inderscience, Vol. 23, No. 1, pp. 71-93, 2017.
31. Subhashini Narayan, E. Sathiyamoorthy, "A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases" Neural Computing and Applications 31, Springer, pp: 93-102, 2019.
32. Sangeeta Bhanja Chaudhuri and Mirzanur Rahman, "Design of a Medical Expert System (MES) Based on Rough Set Theory for Detection of Cardiovascular Diseases" Progress in Advanced Computing and Intelligent Engineering, Springer, pp. 325-332, 2018.
33. W. L. Costa, L. S. Figueiredo, and E. T. A. Alves, "Application of an Artificial Neural Network for Heart Disease Diagnosis" Brazilian Congress on Biomedical Engineering, Springer, pp. 753-758, 2019.
34. Chandra Babu Gokulnath, S. P. Shantharajah, "An optimized feature selection based on genetic approach and support vector machine for heart disease" Cluster Computing, Springer, pp: 1-11, 2018.
35. Kanchan Bala, Dilip Kumar Choubey, Sanchita Paul, Mili Ghosh Nee Lala, "Classification Techniques for Thunderstorms and Lightning Prediction- A Survey", Soft Computing-Based Nonlinear Control Systems Design, IGI Global, pp. 1-17, 2018.
36. Dilip Kumar Choubey, Sanchita Paul, "GA_SVM-A Classification System for Diagnosis of Diabetes", Handbook of Research on Nature

Inspired Soft Computing and Algorithms, IGI Global, pp. 359-397, 2017.

AUTHORS PROFILE



Keshav Srivastava, pursuing M. Tech. (2018-2020) in Computer Science and Engineering from Vellore Institute of Technology in Vellore, completed B. Tech (2014-2018). in Computer Science and Engineering from IMS Engineering College in Ghaziabad. His area of interest are Machine Learning, Database and Web Development. Ph. No. 7291002329, email Id: keshav.srivastava1981@gmail.com, keshav.srivastava2018 @vitsstudent.ac.in.



Dr. Dilip Kumar Choubey, is currently working as an Assistant Professor (Senior) in School of Computer Science and Engineering, Vellore Institute of Technology Vellore, India since 13th May 2019. He worked as an Assistant Professor (On Contract) in NIT Patna, India during 30th March 2017 to 09 May 2019. He worked as an Asst. Prof. in Lakshmi Narain College of Technology (L.N.C.T), Bhopal, India and Oriental College of Technology (O.C.T), Bhopal, India. He has more than 7 years of teaching and research experience. He, received his Ph.D. in Engineering from Birla Institute of Technology (B.I.T), Mesra, Ranchi, India in 2018. He received his M.Tech degree in Computer Science and Engineering from Oriental College of Technology (O.C.T), Bhopal, India in 2012 and has B.E. degree in Information Technology from Bansal Institute of Science and Technology (B.I.S.T), Bhopal, India in 2010. He is a author of 1 book and has more than 37 reputed research publications in international journals, book chapters and conference proceedings. He has also one project as a Co-PI funded by TEQUIP Collaborative Scheme. Being a Coordinator, He has conducted some of the FDP/Workshop. He has been also the member of the organizing and technical program committees of many conferences and Workshop. He has reviewed many research papers in reputed journal/Conferences/book chapters. He has been attended so many workshop's/Seminar's/Conferences. He is the lifetime members of several professional bodies. His research interests include Machine Learning, Soft Computing, Bioinformatics, Data Mining, Pattern Recognition, and Database Management System, etc. Ph. No. 7033789676, email Id: dilipchoubey_1988@yahoo.in, dilip.choubey@vit.ac.in.