European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 4, 2020

IMPROVING THE ACCURACY IN THE PREDICTION OF HEART DISEASE USING MACHINE LEARNING ALGORITHM

KUMARAN.K, N.ANANTHI, G.SARANYA, P.M.LAVANYA, A.SRIDEVI, S.RESHMI SHREE
Assistant Professor, Department of Information Technology Easwari Engineering College, Chennai, India kumaran.me.cse@gmail.com
Professor, Department of Information Technology, Easwari Engineering College, Chennai, India. hod.it@eec.srmrmp.edu.in
Assistant Professor, Department of Computer science and Engineering, SRM institute of science and technology, Chennai, India. saranyag@srmist.edu.in
Assistant Professor, Department of Information Technology, Easwari Engineering college, Tamil nadu, India. lavanya.m@eec.srmrmp.edu.in
UG scholar, Department of Information Technology, Easwari Engineering College, Chennai, India. srideviayyam@gmail.com
UG scholar, Department of Information Technology, Easwari Engineering College, Chennai, India. reshmishree21@gmail.com

Abstract*—Heart disease is one of the most huge sickness disease in the world. Expectation of cardiovascular sickness is a major test in the clinical information investigation. Machine learning (ML) is utilized in settling on choices and forecasts from the enormous amount of information created by the human services industry. Different investigations offer coronary illness with ML methods. This paper is target discovering highlights by applying ML strategies for improving the accuracy level in heart disease. Various techniques have been presented to predict the heart disease. We produce the presentation level of about 88.7% through the forecast model for heart disease with the hybrid random forest techniques.

1. INTRODUCTION

It is hard to distinguish the heart disease due to few factors for example, diabetes, hypertension, elevated cholesterol, anomalous heartbeat rate and numerous different components. Data mining and neural networks have been employed to give out the seriousness of heart disease people. The range disease is dependent in different calculations like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA), and Naïve Bayes (NB). Heart disease is very complex. so,it must be handled carefully. Data mining with order assumes a significant job in the expectation of coronary illness and information examination .Decision trees have been utilized in forecasting the precision to heart disease. Different techniques have been utilized for information deliberation by utilizing known strategies for information digging for forecast of coronary illness. In this work, many readings have been completed to deliver an expectation model utilizing particular strategies as well as by relating at least two procedures. These new systems are normally known as crossover techniques . We present neural networks utilizing pulse time. This technique utilizes different clinical records for expectation, for example, Left group branch square (LBBB), Right pack branch square (RBBB), Normal Sinus Rhythm (NSR), Premature Ventricular Contraction (PVC)), and Second degree square (BII) to out the specific state of the patient corresponding to coronary illness. The dataset with a Radial Basis Function Network(RBFN) is utilized for arrangement, where the 70% of the information is utilized for preparing and the staying 30% is utilized for characterization . We additionally present Computer Aided Decision Support System (CADSS) in the held of medication and research. In past work, the wellbeing industry utilized information mining to demonstrated less time for the forecast of sickness with progressively exact outcomes . We propose the analysis of coronary illness utilizing the General anaesthesia(GA). This strategy utilizes powerful affiliation rules induced with the GA for competition determination, hybrid and the change which brings about the new proposed work. For exploratory approval,

we utilize the Cleveland dataset which is gathered from a UCI AI store. The algorithm called Particle Swarm Optimization (PSO) is presented and a few standards are created for coronary illness. The standards are applied arbitrarily with encoding procedures which bring about progress of the general precision . Heart disease can be anticipated dependent on side effects to be specific heartbeat rate, sex, age, and numerous others.

2. LITERATURE SURVEY

Adam portrays his experience on finding affiliation manages in clinical information to anticipate coronary illness. The main source of death in Heart infection is about 32% , a rate is high as in Canada (35%) and USA. Affiliation rule mining is utilized to recognize the variables that can add to coronary illness, an information base is considered alongside the calculation – Apriori. It is believed that females have more. The outcome demonstrated that when exercise was fake, it was a decent pointer of an individual being sound . This examination has exhibited the utilization of mining to decide information.

V.Krishnaiah has composed about the Healthcare exchange normally clinical conclusion is finished commonly by specialist's information and practice. Choice Support System assumes a significant job in clinical field. Information mining gives the strategy and innovation to adjust these information into helpful data for deciding. Information mining systems sets aside less effort for the expectation of the ailment with more precision. Among the expanding research on coronary illness anticipating framework, it has happened to huge to classifications the exploration results and gives perusers with a blueprint of the current coronary illness forecast procedures in every classification. Information mining instruments can respond to exchange addresses that customarily being used a lot of time superseding to choose. In this paper we study various papers in which at least one calculations of information digging utilized for the expectation of coronary illness. As of the investigation it is seen that Fuzzy Intelligent Techniques increment the exactness of the coronary illness forecast framework. The by and large utilized strategies for Heart Disease Prediction and their complexities are condensed in this paper.

Subrata Kumar Mandal has wrote about a mainstream saying goes that we are living in a “data age”. Information mining is the transformation of an assortment of information into information. The medicinal services industry produces an enormous measure of information ordinary.. Proficient instruments to separate information from these databases for clinical discovery of maladies or different reasons for existing are very little predominant. The point of this paper is utilized to abridge a portion of the exploration on foreseeing heart maladies utilizing information mining strategies, break down the different mixes of mining calculations utilized and finish up which technique(s) are viable and effective.

Vardhaman has composed about an information is produced by the clinical business. This information is unpredictable in nature like electronic records, written by hand contents, and so forth since it is produced from numerous sources. Because of the Complexity of this information, methods that can separate knowledge from this information in a brisk and effective manner. These experiences analyze the sicknesses as well as foresee and forestall the illness .One such utilization of these methods is cardiovascular maladies. Coronary illness is one of the significant reasons for death everywhere throughout the world. the exploration on single information mining systems have not brought about a satisfactory precision.

3. PROPOSED SYSTEM

A. Overview

The analysis is developed by using python and pandas operations to perform heart disease classification of the Cleveland UCI repository.

ML process is from a pre-preparing

N

information stage and highlight determination is done dependent on information cleaning, grouping of demonstrating execution assessment, and the outcomes with improved exactness. The outcomes are contrasted for the exhibition and exactness and DT,SVM,RF and KNN calculations. By utilizing these calculations, the precision level is 88.7%.

a. Overall System Architecture

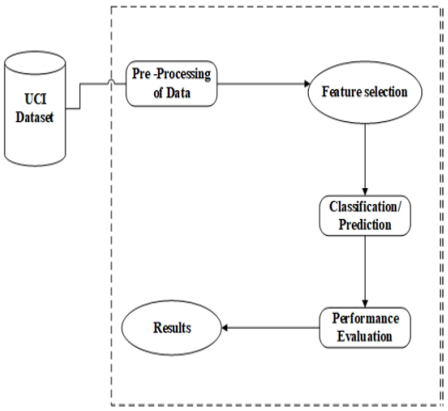


Fig 1: System Architecture

B. Workflow

The strategies which is utilized are information mining and neural systems that have been utilized to give out the seriousness of heart disease people. The seriousness of the infection is characterized dependent on different techniques like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic calculation (GA), and Naive Bayes (NB).

The execution of anticipating heart disease utilizes a few procedures, for example, (1) get dataset,(2) preprocess dataset ,(3) examine dataset ,(4) approve and (5) show exactness

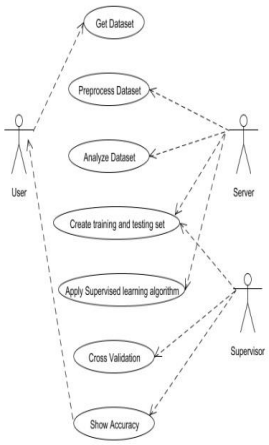
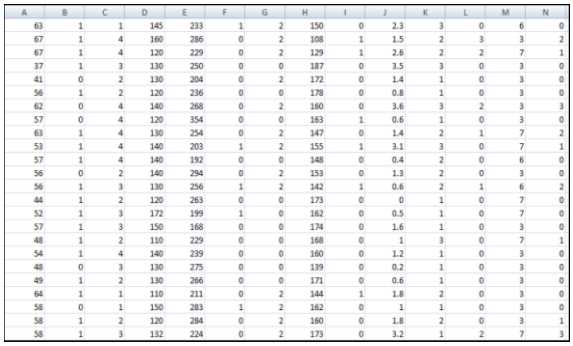


Fig 2.usecase diagram

4. METHODOLOGIES:

➢ Data Pre-Processing:

The collection of datasets are preprocessed. The patients record has been collected of about 303 datasets, where 6 records are with some missing values. Those 6 records have been expelled from the dataset and the staying 297 patient records are utilized in pre-processing.



63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
44	1	2	120	263	0	0	173	0	0	1	0	7	0
52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
48	1	2	110	229	0	0	168	0	1	3	0	7	1
54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
58	0	1	150	283	1	2	162	0	1	1	0	3	0
58	1	2	120	284	0	2	160	0	1.8	2	0	3	1
58	1	3	132	224	0	2	173	0	3.2	1	2	7	3

Fig 3

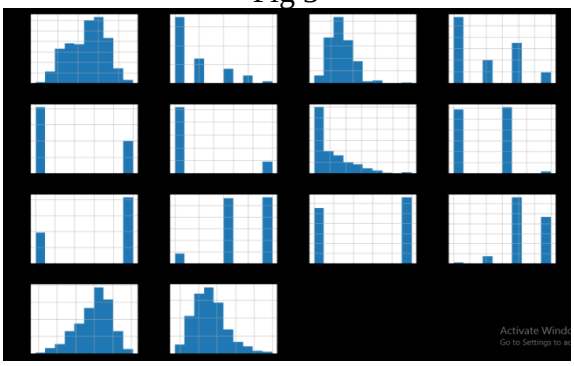


Fig 4

- **Feature Selection and Reduction:**

The age and sex are the two attributes which is used to identify the personal information of the patient among the 13 attributes of datasets.The staying 11 traits are considered as significant on the grounds that they contain clinical records. The severity of heart disease are learning and diagnosis with clinical records.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	num
2	67	1	4	160	286	0	2	108	1	2
3	67	1	4	120	229	0	2	129	1	1
4	37	1	3	130	250	0	0	187	0	0
5	41	0	2	130	204	0	2	172	0	0
6	56	1	2	120	236	0	0	178	0	0
7	62	0	4	140	268	0	2	160	0	3
8	57	0	4	120	354	0	0	163	1	0
9	63	1	4	130	254	0	2	147	0	2
10	53	1	4	140	203	1	2	155	1	1
11	57	1	4	140	192	0	0	148	0	0
12	56	0	2	140	294	0	2	153	0	0
13	56	1	3	130	256	1	2	142	1	2
14	44	1	2	120	263	0	0	173	0	0
15	52	1	3	172	199	1	0	162	0	0
16	57	1	3	150	168	0	0	174	0	0
17	48	1	2	110	229	0	0	168	0	1
18	54	1	4	140	239	0	0	160	0	0
19	48	0	3	130	275	0	0	139	0	0
20	49	1	2	130	266	0	0	171	0	0
21	64	1	1	110	211	0	2	144	1	0
22	58	0	1	150	283	1	2	162	0	0
23	58	1	2	120	284	0	2	160	0	1

Fig 5

- **Classification Modeling:**

The grouping of datasets is done based on the factors and standards of Decision Tree (DT) highlights. At that point, the classifiers are applied to each bunched dataset so as to appraise its exhibition. The best performing models are recognized from the above outcomes dependent on their low rate of error.

- Decision Trees Classifier
- Support Vector Classifier
- Random Forest Classifier
- K-Nearest Neighbors Classifier

Decision Tree Classifier:

Decision tree falls under the classification of managed learning calculations. Decision tree can be use for tackling relapse and arrangement issues the fundamental objective of this classifier is to make a preparation model which can use to predict or estimation of target construed from earlier preparing information**.**

Used Python packages:

1. **sklearn :** sklearn is an machine learning pack which has a tons of ML calculations.

2. **Numpy :** numpy is a numeric python module which gives quick scientific capacities to estimations. It is utilized to peruse information in numpy clusters**.**
3. **Pandas :** pandas used to peruse and compose various documents. Information control should be possible effectively with information outlines.

There are two phases for implementing decision tree:

1. Building phase
   - Preprocess the data set.
   - Split the dataset from train using python sklearn package.
   - Train the classifier.
2. Operational phase
   - Make predictions.
   - Calculate the accuracy.

Data Import: To import and manipulate the data we are using pandas package provided in python.

Data Slicing: split the dataset for training and testing data by using sklearn module before training the model.

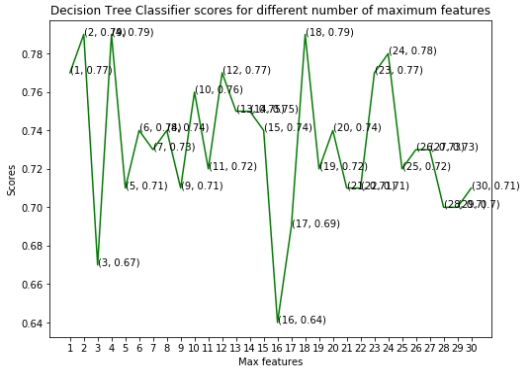


Fig 6

Support Vector Classifier:

Support vectors are a learning models with associated learning algorithms that is used to analyze data which is used for classification and regression. It gives a labeled training data.

- **Importing datasets:** The intuition of support vector ,which optimize a linear discriminant model representing the perpendicular distance between the datasets. The classifier can be trained using training data. Datasets can be imported as CSV file.

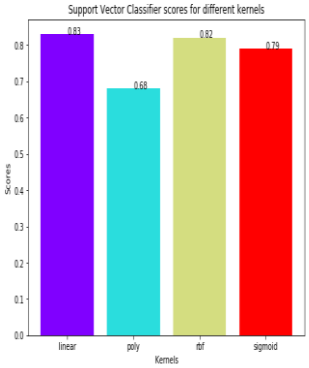


Fig 7

Random Forest Classifier:

A Random Forest is a technique that capable of performing both regression and classification tasks with the use of multiple decision tress and a technique called Bootstrap Aggregation, commonly known as bagging. The basic idea is to combine multiple decision trees in determining the final output.

Implementation of RF:

1. Import the required libraries

2. Import and print the dataset
3. Select all rows and columns from datasets
4. Fit random forest regress to dataset
5. Predicting and visualizing the result

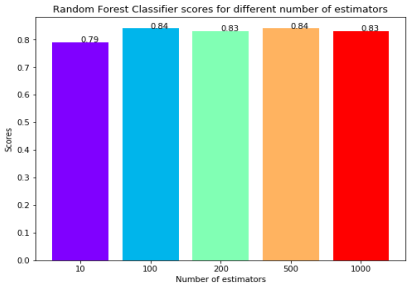


Fig 8

K-Nearest Neighbors Classifier:

KNN can be utilized for both characterization and relapse for prescient issues. It is broadly utilized in the grouping issues in the industry. To assess any strategy there are commonly 3 viewpoints

- Easy to interpret output
- Calculation time
- Predictive power

Implementation of KNN model by following steps:

- Load the data
- Initialize the value of K
- For getting predicted class, iterate from 1 to total number of training data points

1. Calculate the distance between test data and training data in each row
2. Sort the values
3. Get K rows from the sorted array
4. Get the most frequent class of the rows
5. Return the predicted class

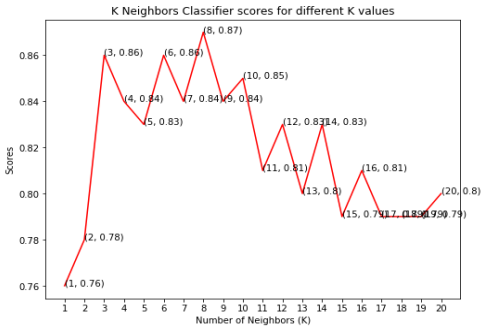


Fig 9

➢ **Performance Measures:**

A few standard exhibition measurements, for example, exactness, accuracy and blunder in grouping have been considered for the calculation of execution viability of this model.

Fig 10

Fig11. Result

3. CONCLUSION

In this paper precision of heart disease can be anticipated for the patients. Half and half AI procedures were utilized in this work to process the information and discover the exactness. Coronary illness expectation is testing and generally significant in clinical field. Anyway the profound quality rate can be controlled if the illness is distinguished at the prior stage. The hybrid machine learning algorithm approach is used the Decision Trees Classifier, Support Vector Classifier, Random Forest Classifier and K-Neighbors Classifier is proved to be accurate in the prediction of heart disease.

4. REFERENCES

[1] A. S. Abdullah and R.Rajalakshmi, ``A data mining model for predicting the coronary heart disease using random forest classi_er," in Proc. Int. Conf. Recent Trends Computer. Methods Common Controls, Apr.2012, pp. 22_25.

[2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ``Using PSO algorithm for producing best rules in diagnosis of heart disease," in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306_ 311.

[3] N. Al-milli, ``Backpropogation neural network for prediction of heart disease," J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131_135, 2013.

[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ``Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233_239.

[5] P. K. Anooj, ``Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27_40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.

[6] L. Baccour, ``Amended fused TOPSIS-VIKOR for classi_cation (ATOVIC) applied to some UCI data sets," Expert Syst. Appl., vol. 99, pp. 115_125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.

[7] C.-A. Cheng and H.-W. Chiu, ``An arti_cial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database," in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566_2569.

[8] H. A. Esfahani and M. Ghazanfari, ``Cardiovascular disease detection using a new ensemble classi_er," in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011_1014.

[9] F. Dammak, L. Baccour, and A. M. Alimi, ``The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains," in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1_8.

[10] Raja, K.S., Kiruthika, U. An Energy Efficient Method for Secure and Reliable Data Transmission in Wireless Body Area Networks Using RelAODV. Wireless Pers Commun 83, 2975–2997 (2015). https://doi.org/10.1007/s11277-015-2577-x

[11] M. Durairaj and V. Revathi , ``Prediction of heart disease using back propagationMLPalgorithm," Int. J. Sci. Technol. Res., vol. 4, no. 8, pp. 235_239, 2015.

[12] M. Gandhi and S. N. Singh, ``Predictions in heart disease using techniques of data mining," in Proc. Int. Conf. Futuristic Trends Comput. Anal. Knowl. Manage. (ABLAZE), Feb. 2015, pp. 520_525.

[13] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, ``Prediction of heart disease using machine learning," in Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Mar. 2018, pp. 1275_1278.

[14] B. S. S. Rathnayakc and G. U. Ganegoda, ``Heart diseases prediction with data mining and neural network techniques," in Proc. 3rd Int. Conf. Converg. Technol. (I2CT), Apr. 2018, pp. 1_6.
[15] N. K. S. Banu and S. Swamy, ``Prediction of heart disease at early stage using data mining and big data analytics: A survey," in Proc. Int. Conf. Elect., Electron., Commun., Comput. Optim. Techn. (ICEECCOT), Dec. 2016, pp. 256_261. 81