

A Supervised Learning Approach for Cardiovascular Disease Using Clinical Parameters by Machine Learning

^[1]Dr Chandrakala V, ^[2]Shashwat AK, ^[3]Sinchan S, ^[4]Somesha HM, ^[5]Sharan Kumar, ^[6]Nandeesh K
Dr Ambedkar Institute of Technology Bengaluru 560056

Abstract—The heart disease cases are rising day by day and it is very Important to predict such diseases before it causes more harm to human lives. The diagnosis of heart disease is such a complex task i.e., it should be performed very carefully. The work done in this research paper mainly focuses on which patients has more chance to suffer from this based on their various medical feature such as chest pain etc. We proposed a system of heart disease prediction that is used to diagnose whether the patient is a victim or not by using the previous medical features of the patient. Support vector machine and k- nearest neighbor algorithms of machine learning are used to predict and classify the patient with heart disease. The models gave satisfactory results and were capable for predicting a heart disease by using k-nearest neighbor and support vector machine which gave a good accuracy in contrast to the algorithms that were used in the previous research such as naive bayes etc.

Keywords—Cardiovascular disease (CVD), Artificial intelligence, Machine Learning (ML), Support Vector Machine (SVM)

I. INTRODUCTION

The heart or blood vessels are effected by a disease called heart disease. Another name for this disease is cardiovascular disease and it is the main reason of death in adults in this era. This is the major cause of deaths in the world and the identification at the early stages is key to successful treatment and cure. Cardiovascular disease (CVD) in all its forms, is a matter of life and death in the planet. This disease is a combination of multiple disease and injuries that effects the heart of a person. [1].

AI has several domains, and one is known as machine learning (ML). It learns from the data given to it as input and it go through that data and find patterns. Human intervention is not too much requires for this. Tasks are not told to the computer instead they are given with the patterns, they detect them and comes to a conclusion. Moreover, ML plays a very

important role in the healthcare centers. The data is generated in electronic from in the healthcare [2].

The use of machine learning in the center of healthcare, the patterns are being found by the algos of machine learning that is very difficult to be detected individually. As machine learning is being more utilized in the field of health care, its providers must need to use a very effective approach so that they have better results[3]. Machine learning algos takes significant features from the data that has more effect and it also helps the doctors, physicians to make better decisions. It also helps them to take steps for the improvement of the health of the patient [4].

The use of machine learning in field of health care has much more effects than a human can even think. The doctors cannot be replaced but the use of machine learning gives a better solution for the heath of a person. It is very important area as we are moving towards smart cities and smart health systems [5]. It is used for the better diagnosis of severe diseases. It also helps in our project to detect a human would have chances of a heart disease through their medical features [6].

If someone has chest pain or high blood pressure, then with fewer test he/she can be diagnosed with disease. In the past, when machine learning was not introduced, the doctors do a lot of tests just to confirm whether the patient is a victim of a certain disease or not. It was such a difficult task but now machine learning algorithms are used instead. Health care offers the care of millions or billions of people that is why it is becoming the highest earners in many of the countries [7].

II. LITERATURE REVIEW

The leading cause of the heart disease is high blood pressure and stroke because it damages the lining of the arteries, that makes more susceptible for the

making of plaque, which narrows the arteries. Early detection of this disease is essential for timely cure. Chang et al. [6] proposed a system in which they proposed a detection system which used Random Forest classifier to identify heart disease. The accuracy of this model is 83%. It predicts heart disease using python. It was published in November 2022. Sonam Nikhar et al. [7] tries to provide a detailed description of the naïve bayes and the decision tree algos, which are applied in research especially for the prediction of heart disease. They concluded that decision tree has much better accuracy than Naïve bayes.

Rishab et al. [8] has used four machine learning algorithms in research, which are Logistic Regression with accuracy 72.89%, naïve Bayes with 70.43% and decision tree with 70.43% also and support vector machine with 71.57%. Irfan Javid et al. [9] in their research used many algos such as random forest, SVM and KNN, LSTM and GRU models of the deep learning as their proposed methodology. Its accuracy was 80.71% accuracy. Jyoti Soni et al. [10] in which three algorithms was used. These were naïve bayes with 76.53% accuracy, decision tree with 79% and KNN with 75.53% accuracy.

Md.Julker Nayeem et al. [11] in their research used KNN, naïve bayes and random forest algo. Out of which the best accuracy was given by random forest which as 85.63%. Malavika et al. [12] utilized naïve bayes, logistic regression, random forest, SVM, decision tree, and k- nearest neighbor algos in the research. The accuracy of Logistic regression is 76.88%, k-nearest neighbor has 76.88%, support vector machine has 78.52%, decision tree has 73.68%, naïve bayes has 80.52% and random forest has 84.80% respectively.

Chithambaram et al [13] algorithms that were used are KNN, SVM, random forest and decision tree. The best accuracy was given by decision tree algo which is 98.83%. Pooja Anbuselvan [14] used various algorithms which are as follows: logistic regression, naïve bayes SVM, KNN, decision tree, random fore, and the technique of ensemble which is Boosting are used. The accuracy given by logistic regression was 75.41%, naïve bayes 77.05%, support vector machine gave 73.77%, k-nearest neighbor had 57.83%, decision tree gave 77.05%, random forest algo has 86.89%, XGBoost algo which gave 78.69% accuracy.

Gunturu Deepthi et al. [15] in this research they used various algorithms like SVM, naïve bayes, decision tree, random forest, logistic regression, Adaboost, XG-boost were used .The dataset was taken from UCI in which 76 attributes were taken, out of which only 14 attributes were selected for the prediction. The best accuracy was given by XGBoost which was 81.3%. The accuracy of SVM was 80.2%, naïve bayes 76.9%, logistic regression 79.1%, decision tree 75.8%, random forest 79.1% and Adaboost 73.6%.

To enhance security and risk mitigation in heart disease detection, control systems and convolutional neural networks can be utilized to improve the accuracy and efficiency of diagnosis and treatment [67-69]

Animesh hazra et al. [16] naïve bayes, kNN and decision tree was used. These algo gave 52.33%, 45.67 and 52% respectively.

III METHODOLOGY AND IMPLEMENTATION

The proposed model provides a forecast for the earlier identification of the heart disease. The dataset is taken from Kaggle website, and it is easily available. As the dataset contains no missing value, outliers or any categorical data so the only one step of preprocessing “feature selection” is performed.

Support Vector Machine (SVM) is one of the most popular supervised learning algo. Its primary use is for the classification and regression problems. In ML its primary function is to use it for the classification problems. Better accuracy can be predicted using SVM. It is a new type of ML methods which is totally based on statistical learning theory. Because of its higher accuracy, Support vector machine became the main focus in the ML community SVM is recently developed within the statistical learning theory. It is applied on a lot of applications such as from time series prediction to the recognition of faces and in the medical for the prediction purposes .

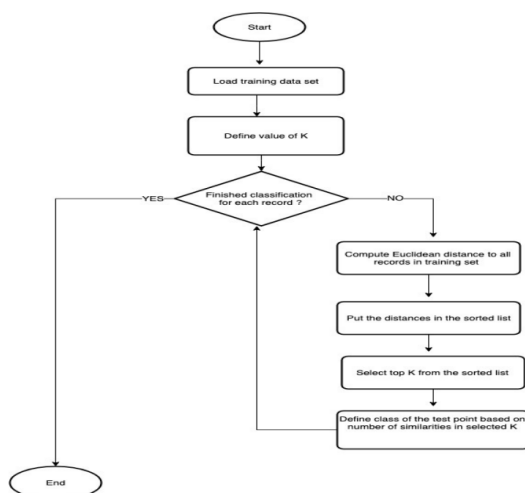
SVM has many advantages both of its problems such as separable and non-separable. Linear separable and non- linear separable problems are included in the separable problems .It works on a high dimensional space by mapping the data so that the points of data can be easily categorized even if the given data is not even linearly separable. There is a separator between the categories, then a separator is drawn as a hyperplane by transforming the data. If there is a clear gap between classes that are categorized, SVM

performs well in such cases. SVM is memory efficient it has more dimensional space than the total number of samples in it, then its performance is very efficient in such cases.

Fischer linear discrimination and quadratic discrimination can be utilized in heart disease prediction. These algorithms can analyze patterns in medical data and provide insights to doctors for better diagnosis and treatment planning [56]. Further, heart disease prediction may be improved while using mobile agents, blockchain technology and machine learning as well



Fig 1.Expected Result



K- Nearest Neighbor a simple and very efficient classification method. It is very productive algorithm of ML. It gives efficient results in both the regression and classification problems. But it is mainly used for the classification problems in which there are two categories. KNN makes a cluster of data, when new

data comes to it. It checks and gives the result. Entry is given to the cluster which has a minimum distance from its neighbor. It finds the distance between all the queries in the data, It selects a number of k neighbors that is closest to the query, then it do voting in the case of classification problems or averaging in terms of regression problems. It is very easy to understand and implement. There is no assumption about the underlying data in KNN so it is very ideal for the use of non-linear problems. There are multiple classes in KNN and it handles it naturally

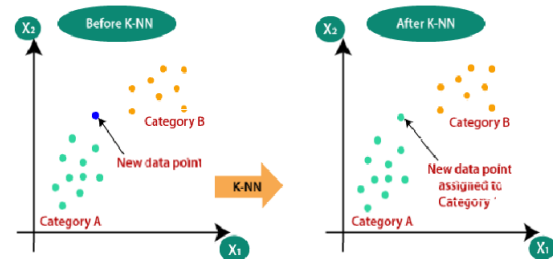


Figure 2: KNN

Following are the libraries used:

- Math Library permitted to use common math constants and functions
- Numpy library facilitates the use of arrays for storing data
- Pandas library assist in reading of data from CSV file
- Plotly and matplotlib are used to create plots for input and output data

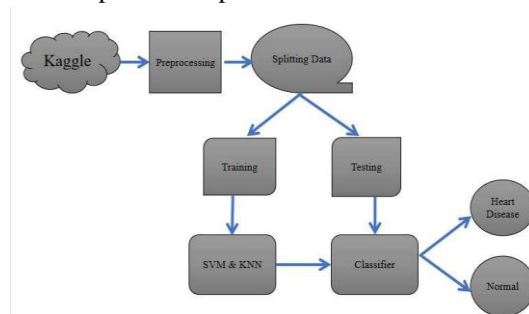


Figure 3: Proposed Methodology

Once the libraries and dataset are imported, the system is first trained on the data, followed by a testing phase. To read the CSV files, we utilize the Pandas library, and then visualize the data as plots using Matplotlib and Plotly.

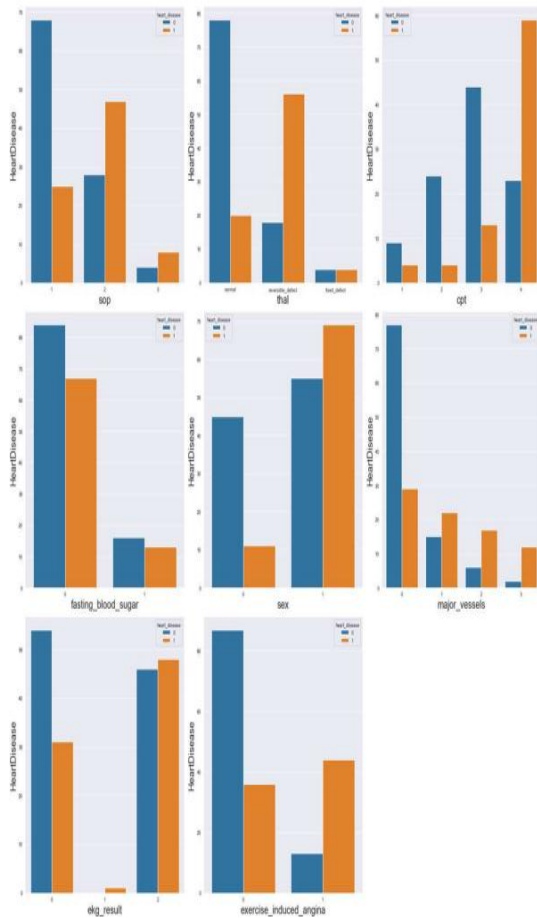


Table 3: RESULTS & ANALYSIS OF KNN

Total Samples = 257	Output	
Expected output	Ec	Enc
c=123 positive	101	22
nc = 134 negative	13	121

Table 4: RESULTS & ANALYSIS OF KNN

Algorithm	Accuracy %	Miss rate %	Precision %	Recall %
KNN	85.23	14	84	90

Table 5 Previous result

Study	Methodology/Model	Accuracy	Proposed Accuracy
Harshit Jindal et al. 2021 [17]	KNN, logistic regression, random forest	KNN 70.56% Lr 72.65% RF 80.56%	KNN 73.6%
Rustem Yilmaz et al. 2021 [19]	Random forest, SVM, Logistic regression	RF 0.929, and LR 0.861	Lr 78.68%
J. Jeyaganesan et al. 2020 [20]	KNN, random forest	KNN 72% RF 83%	RF 85.24 %

IV CONCLUSION AND FUTURE WORK

The major organ in human body is Heart if it stops working human dies. As the death rate is increasing from this disease it has become compulsory to develop a system which is accurate and make the exact prediction of the presence or absence of disease. As it is a matter of life and death the system must needs to be perfect. Heart disease is a type of disease which is not detected at early stages as its symptoms are very normal. We proposed a system which is accurate and do identification of disease at early stages. ML algos are required for the implementation of this system. The analysis is done on the confusion matrix and it compares the accuracies between the models and also find KNN is the best algorithm as its accuracy is higher than the applied SVM. Many research articles written by researchers they used different methods such as KNN, logistic regression, SVM, decision tree, random forest algo. All these algos gave better accuracy. In our research KNN and SVM gave accuracy which is greater than 82% which is almost equal or greater than the accuracies given by different research. In short ,our accuracy can be made more higher by including more medical features of a patient.

V. REFERENCES

- [1] Dr. Seth Flam and EDWARD NASON CEO.” Benefits of Machine Learning in Healthcare”.an overview of cardiovascular disease and research Jan 2007.
- [2] Andrey Koptelov et al.” Machine learning in healthcare: Acomplete overview”. August 24, 2022.
- [3] Anand Nayyar et al. “Machine learning and the Internet of Medical Things in Healthcare” 2021
- [4] Jain, Vishal & Chatterjee, Jyoti. (2020). Machine Learning with Health Care Perspective Machine Learning and Healthcare: Machine Learning and Healthcare.10.1007/978-3-030-40850-3.
- [5] Chang, V.R. Bhavani, A.Q. Xu et al Healthcare Analytics 100016sss 2022.
- [6] Sonam Nikhar, and A.M. Karandikar et al.” Prediction of Heart Disease Using Machine Learning Algorithms”. International Journal of Advanced Engineering, Management and Science (IJAEMS). [Vol- 2, Issue-6, June-2016]. ISSN: 2454-1311.
- [7] Rishabh Magar, Rohan Memane, Suraj Raut et

- al.” HEART DISEASE PREDICTION USING MACHINE LEARNING”. Journal of Emerging Technologies and Innovative Research (JETIR). June 2020, Volume 7, Issue 6.
- [8] Irfan Javid, Ahmed Khalaf Zager Alsaedi and Rozaida Ghazala et al.” Enhanced Accuracy of Heart Disease Prediction using Machine Learning and Recurrent Neural Networks. Ensemble Majority Voting Method”. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 3, 2020.
- [9] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni et al. Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”, International Journal of Computer Applications (0975 – 8887)Volume 17– No.8, March 2011, pp .43-48.
- [10] Md. Julker Nayeem, Ohel Rana, Md. Rabiul Islam et al “Prediction of Heart Disease Using Machine Learning Algorithms”, European Journal of Artificial Intelligence and Machine Learning. Vol 1 | Issue 3 | November 2022, pp .22-26.
- [11] Malavika G, Rajathi N, Vanitha V and Parameswari P et al, PG Scholar, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India. Professor, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India. Assistant Professor (SRG) Department of MCA, Kumaraguru College of Technology, Coimbatore, India. “Heart Disease Prediction Using Machine Learning Algorithms” Biosci. Biotech. Res. Comm. Special Issue Vol 13 No 11 (2020) Pp-24-27.
- [12] Chithambaram T, Logesh Kannan N and Gowsalya M et al.” Heart Disease Detection Using Machine Learning.pp .1-5.
- [13] Pooja Anbuselvan Student Bangalore Institute of Technology Bengaluru, Karnataka, India. “Heart Disease Prediction using Machine Learning Techniques”. International Journal of Engineering Research and Technology (IJERT)ISSN: 2278-0181.Vol. Issue 11, November-2020.
- [14] Gunturu deepthi, Cherukuri Shivani, Koruprolu nagavinith and Kesuboyina hanudeep et al.”HEART DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS”.BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE ENGINEERING.2017- 2021.
- [15] Animesh hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee et al.”Machine Learning and Data Mining Techniques: A Review”. Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 7 pp. 2137-2159.
- [16] Harshit Jindal et al “heart disease prediction using machine learning Algorithms” 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072.
- [17] Raniya R. Sarra, Ahmed M. Dinar, Mazin Abed Mohammed, and Karrar Hameed Abdulkareem.” Enhanced Heart Disease Predictio Based on Machine Learning and χ^2 Statistical Optimal Feature Selection Model”.
- [18] Rüstem Yilmaz,Fatma Hilal Yağın et al,”Early Detection of Coronary Heart Disease Based on Machine Learning Methods”.MEDICAL RECORDS-International Medical Journal.
- [19] J.Jeyaganesan,A.Sathiya, S.Keerthana, Aaradhyaniidhi Aiyer et al.”Diagnosis And Prediction Of Heart Disease Using Machine
- [20] Rati Goel et al.”Heart Disease Prediction Using Various Algorithms of Machine Learning”.Computer Science and Engineering Inderprastha Engineering College.Electronic copy available at: <https://ssrn.com/abstract=3884968>
- [21] Zhang, Y. (2012). Support Vector Machine Classification Algorithm and Its Application. In: Liu, C., Wang, L., Yang, A. (eds) Information Computing and Applications. ICICA 2012. Communications in Computer and Information Science, vol 308. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34041-3_27.
- [22] Theodoros Evgeniou and Massimiliano Pontil Center for Biological and Computational Learning, and Artificial Intelligence Laboratory.”WORKSHOP ON SUPPORT VECTOR MACHINES: THEORY AND APPLICATIONS”.
- [23] Huibing Wang,Jinbo Xiong,Zhiqiang Yao,Jun Ren,Mingwei Lin et al.”Research Survey on Support Vector Machine”.
- [24] Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). KNN Model-Based Approach in Classification.

- [25] K. Taunk, S. De, S. Verma and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.
- [26] Ahmad, G., Alanazi, S., Alruwaili, M., Ahmad, F., Khan, M.A., Abbas, S. and Tabassum, N., 2021. Intelligent ammunition detection and classification system using convolutional neural network.
- [27] Ali, N., Ahmed, A., Anum, L., Ghazal, T.M., Abbas, S., Khan, M.A., Alzoubi, H.M. and Ahmad, M., 2021. Modelling Supply Chain Information Collaboration Empowered with Machine Learning Technique. *Intelligent Automation & Soft Computing*, 30(1).
- [28] Naz, N.S., Abbas, S., Khan, M.A., Abid, B., Tariq, N. and Khan, M.F., 2019. Efficient load balancing in cloud computing using multi-layered mamdani fuzzy inference expert system. *International Journal of Advanced Computer Science and Applications*, 10(3).
- [29] Saleem, M., Abbas, S., Ghazal, T.M., Khan, M.A., Sahawneh, N. and Ahmad, M., 2022. Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques. *Egyptian Informatics Journal*, 23(3), pp.417-426.
- [30] Asif, M., Abbas, S., Khan, M.A., Fatima, A., Khan, M.A. and Lee, S.W., 2021. MapReduce based intelligent model for intrusion detection using machine learning technique. *Journal of King Saud University-Computer and Information Sciences*.