The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853

Development of Automated Heart Disease Diagnosis System Using Machine Learning Algorithm: Current Status And Future Prospects

[1]Dr.Somnath B. Thigale
Head & Associate Professor,
CSE Dept.FTC College of Engineering & Research,Sangola.Dist Solapr.413307
drsomnaththigale@gmail.com
[2]Dr. Vaishnaw G. Kale ,
Department of Computer Engineering, Sinhgad Institute of Technology, Lonavla-410401, Maharashtra, India
Email-vaishnaw25@gmail.com
[3]Sakshi M Hosamani
Assistant Professor, Dept. of E&TC,
Pune Institute of Computer Technology, Pune, India, 411043. shridevikjamage@gmail.com
[4]Mr.Abhiman B.Honkalas
Assistant Professor,CSE Dept. FTC COER,Sangola
[5]Mr.Gajanan G. Kale
Assistant Professor,CSE Dept. FTC COER,Sangola

Abstract— Nowadays Heart disease is one of the most common and serious disease as it is the one of major cause of death globally. Heart disease prediction is very critical and challenging task. Machine Learning (ML) an important technique in the field of Health care applications. Such systems assist (Not replaces) doctors in the interpretation of diseases. Automated Heart Disease Diagnosis System using Machine Learning Algorithm is amongst popular systems which have attracted the attention of numerous researchers, making it a thrust area for further investigations. In last decade, extensive investigations have been contributed to design Automated Heart Disease Diagnosis System using Machine Learning Algorithm. Designing a heart disease diagnosis system with high accuracy can help to save lives. A comprehensive survey of the developments and current trends in area of Automated Heart Disease Diagnosis System using Machine Learning algorithms is presented in this paper. The survey details the overall advancements in diagnosis or prediction of heart disease and an effective review of pre-processing of data, feature extraction algorithms and the classifiers used in prediction system. Few unaddressed issues and challenges that have comparatively received meagre attention are discussed highlighting the future prospects of Heart Disease Diagnosis and providing pointers to the further research.

Keywords— Machine learning, Heart Disease Diagnosis, Accuracy

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853

Introduction

A heart is vital organ, it is a main muscular organ of the human body. The heart purposes as the central point assortment and blood dispatch from the lungs to the rest part of body and vice versa. Heart ailments patients are increasing now a days because of the lifestyle, usage of tobacco, work environment, COVID-19, lack of physical activity across all spectrums of age. The high death rates are due to undiagnosed till a critical stage. If a heart does not execute its functions in well manner, then it will affect other organs like brain, kidney, etc. in the human body [3]. Due to cardiovascular diseases (CVDs) an estimated 17.9 million people died in 2019, representing 32% of all global deaths. 85% of these deaths, were due to heart attack and stroke [4].

A. Heart disease

The risk factor of heart related disease comprises of high blood pressure, high Cholesterol, being obese, diabetics, smoking, more alcohol intake or coronary illness family background. Heart attack symptoms are squatness of breath, pain and uneasiness in chest, the pain may feast to the right hand or left hand to the neck, back, jaw, or stomach, fatigue, cold sweat and instability fast or uneven heartbeat, heart burn or atypical pain [3]. The high death rates are due to undiagnosed till a critical stage. For identifying a heart related issues traditional method includes analyzing the results of multiple tests and arriving at a conclusion and there may be change of error as this process includes multiple human interventions.

One of main reason of heart disease becomes very serious conditions because of blockages which prevent flow of blood to the heart. The reason for blockages is fats are deposited on the inner walls of the blood vessels. So, if the blockades are predicted in the early stage, then treatment is provided and the fatality rate could be decreased proportionally. Generally, a medical dataset consists of thousands of records and a number of attributes to be considered to diagnose or predict a particular disease. To achieve this machine learning technique is used. So, the objective of this paper is to discover machine learning algorithms which gives the relationships between various diseases related to heart and patient characteristics in order to assist physicians and which can provide correct solution to predict heart disease at a primary stage with high accuracy and precision.

B. Machine Learning

Machine learning (ML) is a subgroup of Artificial Intelligence (AI) which is advanced from the pattern recognition study and computational learning theory which uses different statistical and analytics techniques for improving the performance of particular machine learning from old data. Basically, it is a collection of techniques and algorithms employed for creating computational systems which learn from the data for making the forecasts and implications. Machine learning application area is abounding.

Machine learning (ML) stands creates an artificial knowledge which is generated through past experience. Initially, the system undergoes learning stage through study of so many examples and it can be generalizing those afterwards. During this learning phase patterns and

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853
other regularities are memorized, while inspecting new or unknown data patterns will be discovered, that is called transfer of learning (25). There are so many types of machine learning algorithms, the most common ones are supervised, unsupervised and reinforcement learning. Basic block diagram of ML algorithm is as shown in fig 1.

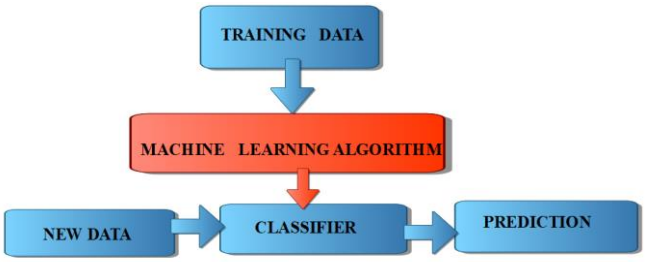

Fig 1: A Machine Learning Algorithm

There are so many types of machine learning algorithms, the most common ones are supervised, unsupervised and reinforcement learning.

Supervised learning: It includes direct supervision and developer label the dataset limits the algorithm boundaries. The patterns study in the is the purpose of the algorithm and construct widespread guidelines for mapping the input to the class or event. Supervised learning machine learning models can be built using three stages Training, Testing or validation, and Prediction. Supervised learning furthermore categorized into classification and the regression problems.

Unsupervised learning: The overall understanding of data available at hand is nothing but the unsupervised learning. The data is not labeled and also not structured so; the output can't be predicted. There are conditions wherein the preferred output event or class is not known for historic records. The goal in such instances is to examine the patterns withinside the input dataset with a purpose to get accurate expertise and to perceive comparable patterns. These patterns may be congregated in explicit events or classes. There is no any necessity of interventions for such algorithms, so they are known as unsupervised learning. Applications of this type of learning are data compression and clustering.

Clustering is nothing but regularities and patterns with likenesses are grouped together withinside an identical group. For identification of new groups, clustering can be employed within the data. However, in the data compression, a big quantity of data is abridged or compressed during operation. Accordingly, memory is abridged and time required for transmission is reduced.

Reinforcement Learning: It may be associated to humans learning method. In the system, there is separate learning for exploiting its performance either by receiving reward or by punishment reaction from the environment. Reinforcement learning algorithms are employed for mapping the situations to actions which will lead to supreme final reward. When an action

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853
is mapped, the algorithm should take into account instantaneous as well as the next and all subsequent rewards [24].

Machine Learning (ML) technique will be used to find different patterns and it can provide required information from them. Machine learning contains large amount of data therefore it can be employed as an efficient support system in the health diagnosis. More time and resources are consumed for analyzing a huge amount of data. Also, altogether the structures existing in the dataset do not support in to solve given problem. Hence, an effective feature selection algorithm can be used for contributing additional in diagnosis of the diseases for finding the more important features. [5]

Literature Review

To tackle the main problem of heart disease prediction system accurately extensive research has been carried out worldwide in last decade. Different machine learning algorithms for prediction of heart disease were examined and they are summarized in this paper. Although some acceptable levels of accuracy have been reached in the presented work there is some drawbacks and challenges leaving scope for research's to still achieve better performance.

Noura Ajam et al. [7] employed Artificial Neural network for diagnosing the Heart Disease with greater accuracy. The algorithm used was Feed forward Back Propagation learning for testing the capability to diagnose heart related disease. 88% classification accuracy was reported by taking into account an appropriate activation function for hidden layer and 20 neurons in hidden layer.

Sairabi H. Mujawar et al. [8] used Naïve Bayes algorithm improved k-mean for Diagnosis of disease related to heart. Cleveland Heart Disease Database data set is for this work. Predictor has achieved 93 % accuracy in forecasting disease related to heart and 89% accuracy for patient who don't have any heart related disease.

Sonam Nikhar et al. [9] author has introduced heart disease prediction system with Naïve Bayes classifier and decision tree classifier. The accuracy of decision tree has reported better in comparison with naïve Bayes classifier. The enhancement in the performance of the Naïve Bayesian classifier by eliminating unrelated characteristics from the dataset and picking characteristics which are furthermost edifying for the classification was reported.

Purshottam et al. [19] was performed research using data mining for efficient Heart Disease Forecast System. By employing 10 fold method, the system has been trained and tested and achieved 86.3 % accuracy in the phase of testing and 87.3 % in the phase of training.

Singh, Yeshvendra K et al. [10] Cleveland heart disease dataset has been used and using Random Forest algorithms an accuracy of 85.81% is achieved

Seyedamin Pouriyeh et al. [11] reported related research for finding and comparing the accuracy of dissimilar data mining classification arrangements so as the accurate Machine Learning Techniques can be employed for the forecast of heart disease. Accuracy obtained

are Decision Tree: 77.55%, Na¨ıve Bayes: 83.49%, SCRL: 69.96%, SVM :84.15%. It was also reported that the SVM method was more accurate in comparison with other schemes.

Senthil Kumar Mohan et al, [1] author projected a method for finding noteworthy structures by employing machine learning techniques so as to improve the accuracy in cardiovascular disease forecast. The hybrid random forest with a linear model (HRFLM) was developed with 88.7% accuracy.

Liaquat Ali el al. [12] author employed a system which stacks two support vector machine (SVM) models for the efficient forecast of HF. The proposed hybrid grid search algorithm (HGSA) is proficient of augmenting the two models concurrently. The different metrics evaluated are sensitivity, accuracy, MCC (Matthews correlation coefficient), specificity, area under curve (AUC) and ROC charts. The proposed method given the enhancement in the performance by 3.3% of a conventional SVM model. Accuracies achieved are in the range of 57.85%–91.83%.

Senthilkumar Mohan et al. [1] proposed hybrid HRFLM approach for uniting of Linear Method (LM) and Random Forest (RF). HRFLM attained 88.7% accuracy in the forecast of heart related disease.

TABLE I. SUMMARY OF LITERATURE REVIEW

Author	**Year**	**Purpose**	**Description**	**Parameters**
Noura Ajam [7]	2015	Heart Diseases Diagnoses using Artificial Neural Network	Technique employed: ANN (Artificial neural network) Algorithm used: Feed forward Back Propagation learning Dataset employed: Cleveland Heart Disease Database taken from UCI machine learning Repository.	Achieved parameter: Mean Square Error (MSE) = 0.1071 Regression = 0.73166 Accuracy = 88%
Sairabi H. Muj	2015	Prediction of Heart relate	Naïve Bayes algorithm modified k-mean for Diagnosis of	93 % accuracy in forecasti

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853

awar[8]		d Disease employing Modified k-means and by using Naive Bayes	heart disease predicted.	ng a heart related disease and 89% accuracy for patient who doesn't have a heart disease.
Sonam Nikhar and A.M. Karandikar [9]	2016	Prediction of Heart Disease Using Machine Learning Algorithms	Techniques used: Decision tree and Naïve Bayes Classifier Dataset used: UCI MachineLearning Dataset	Using Decision tree technique accuracy is increased as compared naïve Bayes classifier
Purushottam [19]	2016	Efficient Heart Disease Prediction System	Techniques used: Decision tree algorithm is used for Classification rules generation V.A. Therapeutic Center created Database.	Attained 86.3 % accuracy in phase of testing and 87.3 % in phase of training
Seyedamin Pouriyeh, Sara	2017	To find and compare the accura	Techniques used: : Na¨ıve Bayes (NB), Decision Tree (DT), Multilayer Perceptron (MLP), Single	Accuracy Decision Tree: 77.55% Na¨ıve Bayes: 83.49

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853

Vahid, Giovanna Sannino † [11]		cy of different data mining classification methods, to use Machine Learning Technique for the prediction of heart disease.	Conjunctive Rule Learner (SCRL), K-Nearest Neighbor (K-NN), Support Vector Machine (SVM) and Radial Basis Function (RBF) Dataset used: Cleveland Heart Disease data set taken from the University of California, Irvine (UCI) learning data set repository	SCRL: 69.96% SVM :84.15%. Precision Recall F-Measure ROC Area Accuracy
Singh,Yeshvendra K. [10]	2017	Heart Disease Prediction System Using Random Forest	Random forest algorithm	Accuracy 85.81%
Senthilkumar Mohan [1]	2019	Objective is to find significant features by applyi	Combining the characteristics of Random Forest (RF) and Linear Method (LM). Data set Cleveland dataset	Accuracy 88.7%.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853

		ng ML techniques so as to increase the accuracy in the prediction of cardiovascular disease	collected from a UCI machine learning repository	
Liaqat Ali et al [20]	2019	An Optimized SVM Expert system is employed for the forecast of Heart Failure	Proposed a hybrid grid search algorithm (HGSA) which is able of augmenting the SVM model 1 and SVM model 2 Data set :Cleveland dataset collected from a UCI machine learning repository	Parameters measured: Accuracy, sensitivity, specificity, MCC (Matthews correlation coefficient), ROC charts, and area under curve (AUC) The proposed method given improvement in the performance by 3.3% of a conservati

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853

				ve SVM model. Accuracies achieved in the range of 57.85%–91.83%.
Santhana Krishnan. J et al[23]	2019	Prediction of Heart Disease Using Machine Learning Algorithms	Supervised Algorithms used: Naive Bayes and Decision tree Data set: Cleveland dataset collected from a UCI machine learning repository	Accuracy of decision tree and Naive Bayes algorithm mare is 91% & 87% respectively
Senthilkumar Mohan , Chandrasegar Thirumalai , And Gautam Srivastava[1]	2019	Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques	Hybrid random forest with a linear model (HRFLM). Data set:: UCI dataset for heart disease.	Accuracy = 88.7%
Kalpana &	20	Heart Disease	Algorithm used: Random Forest	Random Forest

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 842 – 853

Snehitha[14]	21	Diagnosis System Using Machine Learning		algorithm provided 86% accuracy

Conclusion

Despite the scrupulous efforts to develop a fully automatic Automated Heart Disease Diagnosis System using Machine Learning Algorithm, it has limited success in comparison with the performance and accurate results. One of the possible areas of further expansion for improved performance includes adequate utilization of the expert knowledge and experience to develop decision making ability. Enormous efforts to develop Automated Heart Disease Diagnosis System using Machine Learning Algorithm have been made so far. This is an attempt to completely circumvent or minimize the human effort in the process of heart disease diagnosis. The system must therefore have both, expert knowledge and experience of the cardiologist from the clinical acceptability perspective. This makes the development of Automated Heart Disease Diagnosis System difficult and challenging and therefore demands additional research.

In this paper a comparative study of various machine learning algorithms which are employed for heart related disease forecast. Great efforts to develop heart related disease forecast employing Machine learning have been made so far. However, the limitations of existing methods are parameters like classification accuracy, sensitivity and prediction results are to be improved. Above mentioned ML algorithms have performed very well still there is scope that a combination of some of these techniques can help in removing the drawback of individual methods of heart disease diagnosis

Based on the above review, it can be concluded that machine learning approach can be used more effectively to predict heart related diseases. An improved system will be developed which can be employed for the selection of appropriate treatment methods for a heart disease diagnosis.

References

[1] Senthilkumar Mohan,Chandrasegar Thirumalai1,And Gautam Srivastava ," Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" Volume 7, October 2019, Pages 81542-81554.

[2] Yuvraj Nikhate,M. V. Jonnalagedda, "SURVEY ON HEART DISEASE PREDICTION USING MACHINE LEARNING", IJCRT, Volume 8, Issue 8 August 2020 | ISSN: 2320-2882, 1600–1606.

[3] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta,Arkomita Mukherjee ,Asmita Mukherjee, “Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques:A Review”, Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159

[4] CVD Details from WHO report. https://www.who.int/en/news room/factsheets/detail/cardiovascular-diseases-(cvds)

[5] J. Vijayashreea, H. Parveen Sultanaa, “A Machine Learning Framework for Feature Selection in Heart Disease Classification Using Improved Particle Swarm Optimization with Support Vector Machine Classifier”, ISSN 0361-7688, Programming and Computer Software, 2018, Vol. 44, No. 6, pp. 388–397.

[6] Mangesh Limbitote, Dnyaneshwari Mahajan, Kedar Damkondwar, Pushkar Patil, “A Survey on Prediction Techniques of Heart Disease using Machine Learning” International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 9 Issue 06, June-2020,450-453.

[7]Noura Ajam, 2015, “Heart Diseases Diagnoses Using Artificial Neural Network”, Network And Complex Systems, ISSN: 2224-610X (Paper), ISSN: 2225-0603(Online), Vol.5, No.4,pp. 7-11.

[8] Sairabi H. Mujawar, and P. R. Devale, October 2015,“Prediction of Heart Disease using Modified k-means and by using Naive Bayes”, International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10, pp. 10265-10273.

[9] Sonam Nikhar, A.M. Karandikar” Prediction of Heart Disease Using Machine Learning Algorithms” International Journal of Advanced Engineering, Management and Science (IJAEMS) Infogain Publication, [Vol-2, Issue-6, June- 2016].I.S. Jacobs and C.P. Bean,“Fine particles, thin films and exchange anisotropy,” in Magnetism,vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp.271-350.

[10] Singh, Yeshvendra K., Nikhil Sinha, and Sanjay K. Singh, "Heart Disease Prediction System Using Random Forest", International Conference on Advances in Computing and Data Sciences. Springer, Singapore, 2017

[11] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez ,”A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease”, 22nd IEEE Symposium on Computers and Communication (ISCC 2017), 22nd IEEE Symposium on Computers and Communication (ISCC 2017).

[12] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD),” Prediction of Heart Disease Using Machine Learning”, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-

[13] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar,” Machine Learning Techniques For Heart Disease Prediction”, International Journal Of Scientific & Technology Research Volume 8, Issue 11,November 2019.

[14] Kalpana, Snehitha,” Heart Disease Diagnosis System Using Machine Learning.” International Research Journal of Engineering and Technology (IRJET), Volume: 08

[15] H. A. Esfahani and M. Ghazanfari, ‘‘Cardiovascular disease detection using a new ensemble classifier,’’ in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.

[16] 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.

[17] R. Das, I. Turkoglu, and A. Sengur, ‘‘Effective diagnosis of heart disease through neural networks ensembles,’’ Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, May 2009. doi: 10.1016/j.eswa.2008.09.013.

[18] M. Durairaj and V. Revathi, ‘‘Prediction of heart disease using back propagation MLP algorithm,’’ 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.

[19] Purushottam, Kanak Saxena and Richa Sharma, "Efficient heart disease prediction system." Procedia Computer Science 85 (2016): 962-969.

[20] Ali, Liaqat, et al, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure." IEEE Access 7 (2019): 54007-54014

[21] 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.

[22] 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.

[23] Santhana Krishnan. J, Geetha S., "Prediction of Heart Disease Using Machine Learning Algorithms", 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)

[24] ”Mastering Machine Learning with Python in Six Steps”A practical impementaion guide to predictive dada analytics using Python by Manohar Swamynathan.,Apress.

[25] Machine Learning and Data Mining by Igor Kononenko, Matjaz Kukar,

[26] Reinforcement Learning by Richard S. Sutton, 2012

[27] Supervised and Unsupervised Pattern Recognition: Feature Extratraction and Computational Intelligence by Evangelia Micheli-Tzanakou, 2000

[28] http://archive.ics.uci.edu/ml/datasets/Heart+Disease

[29] Mousami V. Munot,” DEVELOPMENT OF COMPUTERIZED SYSTEMS FOR AUTOMATED CHROMOSOME ANALYSIS: CURRENT STATUS AND FUTURE PROSPECTS” International Journal of Advanced Research in Computer Science, Volume 9, No. 1, January-February 2018.