Heart Disease Prediction Based on Machine Learning Techniques

G. Chakravarthi[1], S.MD. Jabeer [2]

[1]*PG Scholar, Department of CSE, Global College of Engineering and Technology, Kadapa*
[2]*Assistant Professor, Department of CSE, Department of CSE, Global College of Engineering and Technology, Kadapa*

Abstract - **The latest statistics of the World Health Organization anticipated that cardiovascular diseases including Coronary Heart Disease, Heart attack, vascular disease as the biggest pandemic to the world due to which one-third of the world population would die. With the emerging AI trends, applying an optimal machine learning model to target early detection and accurate prediction of heart disease is indispensable to bring down the mortality rates and to treat cardiac patients with the best clinical decision support. This stems from the motivation of this paper. This paper presents a comprehensive survey on heart disease prediction models derived and validated out of popular heart disease datasets like the Cleveland dataset. The main keywords for the search were Heart Disease, Prediction, Coronary disease, Healthcare, Heart datasets, and Machine Learning. Results: This review explores the shortcomings of various approaches used for the prediction of heart diseases. It outlines the pros and cons of different research methodologies along with the validation parameters of each reviewed publication.**

Index Terms - **Heart diseases, Machine learning, Deep learning, Health care, heart disease dataset.**

I.INTRODUCTION

Heart Disease/Disorders (HD) have been recognized as one of the convoluted and fatal human illnesses in the world. Due to this disease, the heart functions abnormally leading to blocked blood vessels and get affected by angina, heart attack, and stroke. The most common types of heart diseases are Coronary Vascular Disease (CVD), coronary artery disease (CAD), Congestive Heart Failure (CHF), and Abnormal Heart Rhythms. There are many challenges in predicting such HD at the early stages due to the involvement of several conventional risk factors like age, sex, hypertension, high cholesterol, abnormal pulse, and many other factors [1]. Despite wide diversity in the existence of cardiovascular risk factors across different sectors of society, CVD has been noticed to be one of the major causes of death all over India including economically backward states and rural areas. The global statistics also showed that the premature mortality in terms of years of life lost because of CVD climbs to 37 million (2010) from 23.2 million (1990) with an incremental rise of 59 % every year, which serves as the prime motivation of this paper. The need for heart disease diagnosis has compelled towards invention few invasive clinical techniques like angiogram, which in spite of being expensive also induces some side effects for the diagnosed patients. This has motivated several researchers to use data mining techniques to diagnose CVD safely. Machine Intelligence is a type of intelligence exhibited by machines to interconnect with the physical world [2]. Machine learning and deep learning technologies are two subsets of AI, which are likely to be used as the model to predict and ascertain the data. Both these technologies are very powerful and worthy for medical data analytics. Application of different types of machine intelligence paradigms is an ideal approach for heart disease diagnosis but as well serves as an aid for prediction, illness monitoring, and its other related clinical management aspects [3,4]. The related works of machine/deep learning in the medical field related to heart disease predictions have been explored elaborately in forthcoming sections and the generalized framework opted by most of the researchers for the prediction of heart disorders is shown in Figure 1. A prelude on the heart disease datasets commonly used by the researchers is presented in the subsequent section. This article provides the benefits and shortcomings of the

reviewed publications in the results section and highlights the salient points in the discussion section.

II LITERATURE SURVEY

There is number of works has been done related to disease prediction systems using different machine learning algorithms in medical Centres. Senthil Kumar Mohan et al,[10] proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which strategy that objective is to finding critical includes by applying Machine Learning bringing about improving the exactness in the expectation of cardiovascular malady. The expectation model is created with various blends of highlights and a few known arrangement strategies. We produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model (HRFLM) they likewise educated about Diverse data mining approaches and expectation techniques, Such as, KNN, LR, SVM, NN, and Vote have been fairly famous of late to distinguish and predict heart disease. Sonam Nikhar et al [11] has built up the paper titled as Prediction of Heart Disease Using Machine Learning Algorithms by This exploration plans to give a point-by-point portrayal of NaÃ¯ve Bayes and decision tree classifier that are applied in our examination especially in the prediction of Heart Disease. Some analysis has been led to think about the execution of prescient data mining strategy on the equivalent dataset, and the result uncovers that Decision Tree beats over Bayesian classification system. Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD), [3] Prediction of Heart Disease Using Machine Learning, In this paper proposed system they used the neural network algorithm multi-layer perceptron (MLP) to train and test the dataset. In this algorithm there will be multiple layers like one for input, second for output and one or more layers are hidden layers between these two input and output layers. Each node in input layer is connected to output nodes through these hidden layers. This connection is assigned with some weights. There is another identity input called bias which is with weight b, which added to node to balance the perceptron. The connection between the nodes can be feedforwarded or feedback based on the requirement. Abhay Kishore et al,[4] developed Heart Attack

Prediction Using Deep Learning in which This paper proposes a heart attack prediction system using Deep learning procedures, explicitly Recurrent Neural System to predict the probable prospects of heart related infections of the patient. Recurrent Neural Network is a very ground-breaking characterization calculation that utilizes Deep Learning approach in Artificial Neural Network. The paper talks about in detail the significant modules of the framework alongside the related hypothesis. The proposed model deep learning and data mining to give the precise outcomes least blunders. This paper gives a bearing and point of reference for the advancement of another type of heart attack prediction platform. Prediction stage. Lakshmana Rao et al,[14] Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more (circulatory strain, diabetes, current smoker, high cholesterol, etc..). So, it is difficult to distinguish heart disease. Different systems in data mining and neural systems have been utilized to discover the seriousness of heart disease among people. The idea of CHD ailment is bewildering, in addition, in this manner, the disease must be dealt with warily. Not doing early identification, may impact the heart or cause sudden passing. The perspective of therapeutic science furthermore, data burrowing is used for finding various sorts of metabolic machine learning a procedure that causes the framework to gain from past information tests, models without being expressly customized. Machine learning makes rationale dependent on chronicled information. Mr. Santhana Krishnan.J and Dr. Geetha.S, [15] Prediction of heart disease using machine learning algorithm This Paper predicts heart disease for Male Patient using Classification Techniques. The detailed information about Coronary Heart diseases such as its Facts, Common Types, and Risk Factors has been explained in this paper. The Data Minin tool used is WEKA (Waikato Environment for Knowledge Analysis), a good Data Mining Tool for Bioinformatics Fields. The all three available Interface in WEKA is used here; Naive Bayes, Artificial Neural Networks and Decision Tree are Main Data Mining Techniques and through this techniques heart disease is predicted in this System. The main Methodology used for prediction is Decision Trees like CART, C4.5, CHAID, J48, ID3

III. BACKGROUND METHODS

Machine learning is a hot topic in research and industry, with new methodologies developed all the time. The speed and complexity of the field makes keeping up with new techniques difficult even for experts and potentially overwhelming for faster analysis.

Logistic regression

Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. It is one of the supervised learning and is used to estimate the target object value's possibility. It is a tool to calculate the statistical values and make results on binary output. In the linear method, which is calculated by the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts $P(Y=1)$ as a function of X. Here, y is the linear model's output trained with logistic regression produce value between zero and one.

Naive Bayes

In the Naïve Bayes network, all features are independent. Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles. When there is a change in one feature, it does not affect another. This is suitable for large datasets. The assumption from Conditional independence is that an attribute value is independent of the values, which are from other attribute values in a class. Bayes' Theorem is based on probability theory.

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:
Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is

identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

Support Vector Machine (SVM)

SVM is used both for regression and classification tasks. The SVM model represents the data in the space described so that the examples in various categories are divided by a distance as large as possible. That divides sensitive information with the maximum separable space between them and is calculated so that many of the points belong to one group fall on the plane's one side.

Radial Basis Function (RBF)

An Artificial Neural Network that uses non linear radia basis function as activation functions and gives linear output using combination of radial basis functions of the inputs and neuron parameters. RBF is mainly used in SVM classification, which maps input space in new dimensional space. In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. It is the default kernel used within the sklearn's SVM classification algorithm. A kernel is a function that takes the original non-linear problem and transforms it into a linear one within the higher dimensional space.

KNN

K-Nearest Neighbor is an anti-parametric method, which is used for regression and classification. It is essentially a grouping method, consider the distance between a point and the coordinates (x, y) and its neighbors. The distance between the Euclidean its neighbors are determined from the point and eventually located in the region nearest to its neighboring points.

IV. EXISTING ANALYSIS

HD classification system by using machine learning classification techniques and the performance of the system was 77% in terms of accuracy. Cleveland dataset was utilized with the method of global

evolutionary and with features selection method. In another study Gudadhe et al. [22] developed a diagnosis system using multi-layer Perceptron and support vector machine (SVM) algorithms for HD classification and achieved accuracy 80.41%.

Drawbacks:
- In the existing work, the system is poor performance in which the low performance LOCAL LEARNING BASED FEATURES SELECTION ALGORITHM is used.
- This system is less performance due to Lack of Heart disease classification Techniques

V. PROPOSED WORK

The system is proposed a machine learning based diagnosis method for the identification of HD in this research work. Machine learning predictive models include ANN, LR, K-NN, SVM, DT, and NB are used for the identification of HD. The standard state of the art features selection algorithms, such as Relief, mRMR, LASSO and Local-learning-based features-selection (LLBFS) have been used to select the features.

Benefits:
- The system is fast and reliable due to presence of feature selection algorithm (FCMIM).
- The system is more effective due to STANDARD STATE OF THE ART FEATURES SELECTION ALGORITHMS
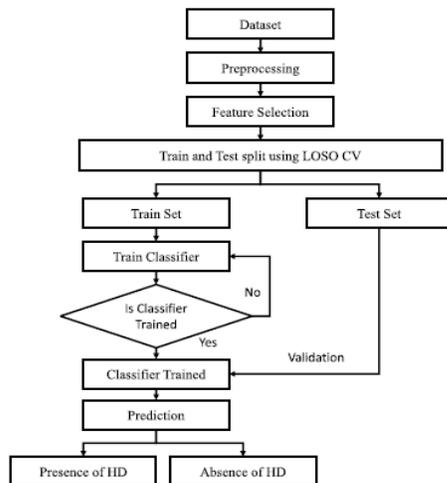

Fig 3: Flow Showing Proposed Architecture showing Stacking Classifiers

VI. METHODOLOGY

Heart disease (HD) is the critical health issue and numerous people have been suffered by this disease around the world. The HD occurs with common symptoms of breath shortness, physical body weakness and, feet are swollen. Researchers try to come across an efficient technique for the detection of heart disease, as the current diagnosis techniques of heart disease are not much effective in early time identification due to several reasons, such as accuracy and execution time. The diagnosis and treatment of heart disease is extremely difficult when modern technology and medical experts are not available. The effective diagnosis and proper treatment can save the lives of many people. According to the European Society of Cardiology, 26 million approximately people of HD were diagnosed and diagnosed 3.6 million annually. Most of the people in the United States are suffering from heart disease. Diagnosis of HD is traditionally done by the analysis of the medical history of the patient, physical examination report and analysis of concerned symptoms by a physician. But the results obtained from this diagnosis method are not accurate in identifying the patient of HD. Moreover, it is expensive and computationally difficult to analyze.
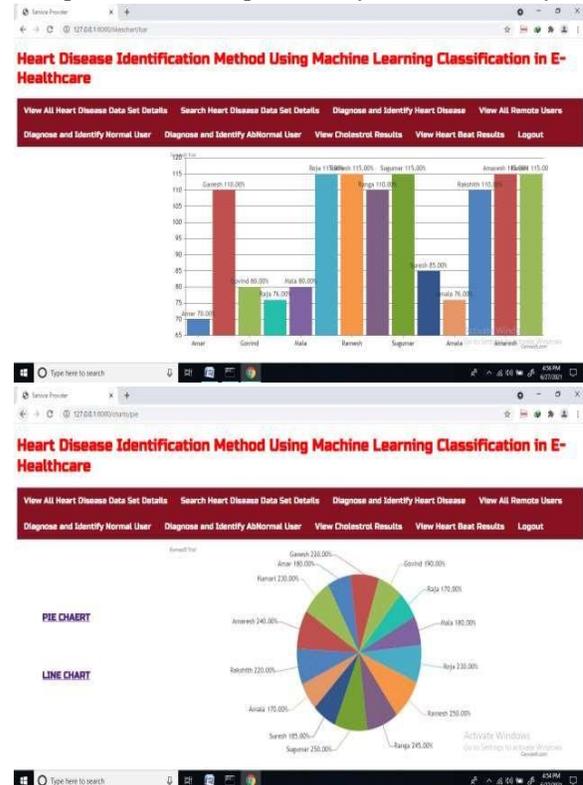
VI. CONCLUSION

Machine intelligence can serve as a genuine alternative diagnostic method for prediction, which will, in turn, keep the patients well aware of their illness state. This article presents a comprehensive study of heart prediction systems based on machine learning, ensemble, and deep learning approaches. From the reviewed literature, it is obvious that the Cleveland heart disease dataset that contains only 303 instances with 14 features is mostly used. This is mainly because of the tiny and restricted sample size. Any study that uses other data sources also concentrated on a single dataset with a limited number of features. Consequently, high accuracies obtained in the prediction models with the removal of irrelevant features or removal of highly correlated factors or by using feature selection/ optimization techniques cannot be generalized, which is a major shortcoming. Despite the researcher's efforts, still uncertainty exists towards the standardization of prediction models. To get a more generalized classification and prediction accuracy, other multiple heart disease datasets from different sources with more features should be considered. An efficient predictive framework model which eliminates most of the shortcomings reported in this paper is the cardinal intent of our future research. Furthermore, real-time data should be analyzed on the working learning model to get it standardized and ensure its reliability with the clinical correlation and validation.

REFERENCE

[1] M. Durairaj and V. Revathi, ``Prediction of heart disease using back propagation MLPalgorithm,'' Int. J. Sci. Technol. Res., vol. 4, no. 8, pp. 235-239, 2015.

[2] 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.

[3] 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.

[4] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[5] Mangesh Limbitote , Dnyaneshwari Mahajan , KedarDamkondwar , Pushkar Patil, 2020, A Survey on Prediction Techniques of Heart Disease using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 06 (June 2020),

[6] A. S. Abdullah and R. R. Rajalaxmi, ``A data mining model for predicting the coronary heart disease using random forest classifier,'' in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[7] 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.

[8] N. Al-milli, ``Backpropogation neural network for prediction of heart disease,'' J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[9] 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.

[10] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019

[11] S.P. Bingulac, On the Compatibility of Adaptive Controllers, Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings)

[12] 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.

[13] 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- 5386-0965-1

[14] Abhay Kishore1, Ajay Kumar2, Karan Singp, Maninder Punia4, Yogita Hambir5, Heart Attack Prediction Using Deep Learning, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 | Apr-2018.

[15] 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.

[16] Mr.Santhana Krishnan.J, Dr.Geetha.S, Prediction of Heart Disease Using Machine Learning Algorithms,2019 1st International Conference on Innovations in Information and Communication Technology(ICIICT),doi:10.1109/ICIICT1.2019. 8741465.