Machine Learning-Based Heart Patient Scanning, Visualization, and Monitoring

Ahmed Al Ahdal
Department of Machine Learning and Arterial intelligence
Lovely Professional University
Jalandhar, Phagwara,
Punjab, India
Ahmedlpu20@gmail.com

Dr. Deepak Prashar
Department of Computer Science and Engineering
Lovely Professional University
Jalandhar, Phagwara,
Punjab, India
Deepak.prashar@lpu.co.in

Manik Rakhra
Department of Computer Science and Engineering.
Lovely Professional University
Jalandhar, Phagwara,
Punjab, India
Rakhramanik786@gmail.com

Ankita Wadhawan
Department of Computer Science and Engineering
Lovely Professional University
Jalandhar, Phagwara,
Punjab, India
Ankita.23891@lpu.c.in

***Abstract* - heart diseases leading most causes of death globally according to World Health Organization cardiovascular or all heart related disease are responsible for 17.9 million death every year. An early detection and diagnosis of the disease is very important and maybe it's the key of cure. The major challenge is to predict the disease in early stages therefor most of scientists and researches focus on Machine learning techniques which have the capability of detection with accurate result for large and complex data and apply those techniques to help in health care. The purpose of this work is to detect heart diseases at early stage and avoid consequences by implementing different Machine Learning Algorithm for example, KNN Decision Tree (DT), Logistic Regression, SVM, Random Forest (RF), and Naïve Bayes (NB).**

Index Terms* - *Machine Learning, Cardiovascular Disease, Decision Tree, Heart Disease Prediction

I. Introduction

This paper focuses mainly on various Machine Learning techniques that are employed in heart disease prediction, cardiac (heart) is very important organ on the body which is responsible for regular blood flow throughout the body, therefore any irregularity to heart can cause distress in other parts of the body. Today, there is more reasons for heart diseases such as unhealthy lifestyle, smoking etc. Alcohols are major causes of heart disease. Good health style and early detection are the most way to prevent heart disease. Machine learning, a subfield of artificial intelligence, can learn from massive datasets and predict similarly previously unseen or new data based on its methods of learning or training. There are various kinds of cardiovascular diseases, within each variety of symptoms, such as: 1 - cardiovascular disease caused by an irregular heartbeat, low heartbeat, anxiety, and chest pain. 2- Blood vascular disease in the heart that causes chest discomfort and breathlessness. There are several causes of heart diseases, such as high blood pressure, hypertension, and drugs. Heart diseases include, heart infections, heart failure, cardiac arrest, hypertension, slow heartbeat, and stroke. Many factors for heart diseases are age, family history of coronary illness, blood pressure, and Cholesterol level.

II. Related Work

Many researchers have applied machine learning techniques that help in healthcare industry in past few years. and specialists in the detection of heart-related illnesses. K-Nearest Neighbor (KNN), Decision Tree, Logistic Regression(LR), SVM, Random Forest, and Nave Bayes are some examples, Lot of work has been carried out in the existing system, by using Machine learning techniques and different datasets different accuracy have been attained.

Senthil Kumar et al, [1] improved prediction of cardiovascular disease using composite machine learning techniques that includes a methodology which seeks to find significant implication through the applying machine learning, leading to enhanced accuracy within prediction of cardiovascular illness.

Abhay Kishore et al,[2] used Deep Learning to develop heart attack predictions Recurrent Neural Network which is a revolutionary characterisation approach that makes use of the Deep Learning technique in Artificial Neural Network, The paper goes into great detail on the framework's primary modules, and also the associated assumption. This suggested methodology uses deep learning with data mining to achieve the most accurate results with the lowest failures. This work provides a foundation and reference point for the building of a different variety of heart attack prediction platform.

Vembandasamy et al,[3] work was preformed by using Naïve Bayes Algo which is a powerful independence assumption, the data was obtained from diabatic research institute and it consists of 500 patients record and Naïve Bayes Algorithm offers 86.919% of accuracy.

Mr. Santhana Krishnan. J and Dr. Geetha. S, [4] used classification techniques to predict cardiac disease in male patients. This work provides comprehensive data regarding Heart Diseases, covering Background, Prevalent Type, and Factors Associated. All three weak interfaces are used there; the key data mining methodologies are Naive Bayes, Artificial Neural Networks, and Decision Trees, and these techniques are used to predict heart disease.

The table below shows that most of Machine Learning Algorithm has been used for heart disease with their author, year, and accuracy.

TABLE 1
A COMPARATIVE STUDY OF VARIOUS ML

Author	Disease	Methods Algo	Year	Accuracy
Shan Xu et al	Heart disease	SVM	2017	98.9%
Kamran Farooq	Heart	DT	2014	78.4604%

2021 International Conference on Computing Sciences (ICCS) | 978-1-6654-9445-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCS54944.2021.00049

978-1-6654-9445-8/21/$31.00 ©2021 IEEE
DOI 10.1109/ICCS54944.2021.00049

et al [2]	disease			
Otoom et al[3]	Heat disease	SVM	2015	88.3%
Syed Muhammad Saqlain Shah et al	Heat disease	SVM	2017	91.30%
Megha Shahi et al.	Heat disease	SVM	2017	85%
Vembandasamy et al	Heat disease	Naïve bayes	2018	86.41
Abhu Kidhorel et al	Heart attack	RNN	2019	92%
Mr. Santhana Krishnan. J	Heart disease	NB, DT	2019	91%
Senthil kumar Mohan	Heart disease	DT, SVM	2019	88.4%

As we see in the table most of Machine Learning algorithm has been used for heart disease with their author, year, and accuracy. There are many other studies, but our aim was for latest paper from last 8 years. The main idea after reviewing the above table and the purpose of all articles was to compare the accuracy and f-measure scores of classification methods such as Decision Tree, Random Forest, Logistic Regression, and Nave Bayes,[5] The main objective is to effectively predict if the patient suffers from heart disease or not having any heart related illness. data on various health issues is being collected all across the world. These data may be used to acquire significant information by implementing a number of machine learning approaches. Unfortunately, the size of the data gathered is massive, and this data is frequently incorrect and noisy and too massive for mind to comprehend, may be investigated by applying machine learning methodologies. As a result, in recent times, these algorithms have shown to be extremely useful in accurately confirming the presence or absence of heart-related diseases There are several techniques for cleaning data. removing the nose from it in order to obtain exact results, such as Dimensionality reduction involves choosing a mathematical representation that can correlate the most, though not all, of a variance within the data presented, leading in just the most important information being included. A massive number of features, or elements, can increase computing complexity and therefore can cause in overfitting, resulting in bad results. As a result, Dimensionality Reduction is an important stage in the creation of any model. Dimensionality reduction is often achieved by one of two methods: feature extraction or dimension reduction.

III. Methodology

The below block diagram shows the basic steps which have been taken for all machine learning model, firstly data cleaning to convert the raw data into a manageable form because raw data cannot be used directly, then determine important features those steps are applied to each machine learning model predictions.

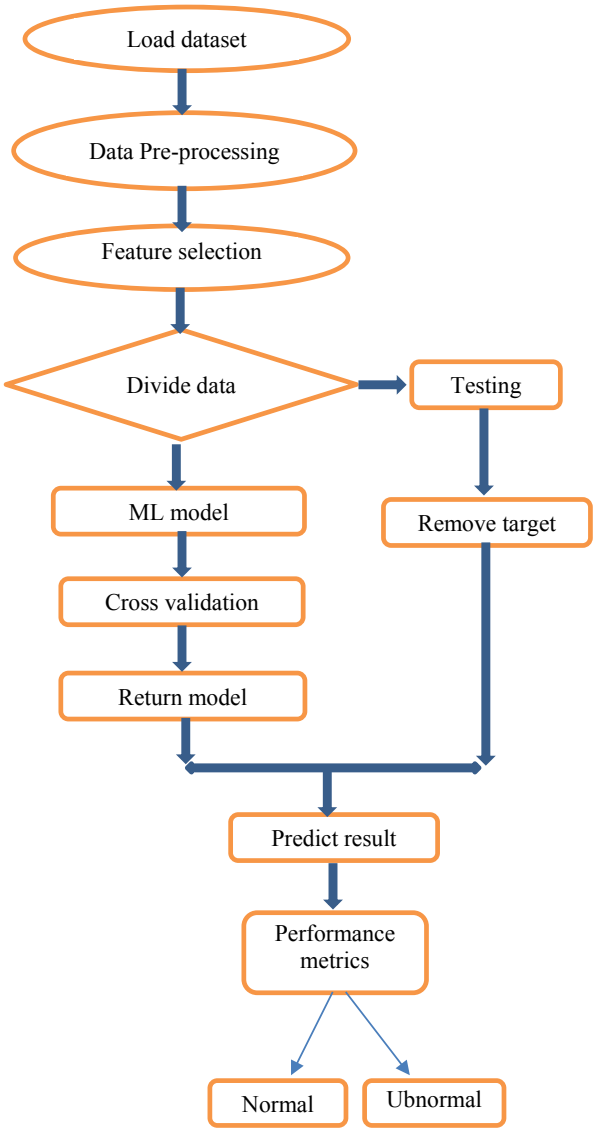

Fig. 1: Flow of the proposed methodology

After cleaning and analyzing all dataset, apply machine learning models such as SVM Logistic Regression, Naïve. Bayas(NB), (DT)Decision Tree, and RD.

IV. Algorithm and Technique

A. *Support Vector Machine*

Support Vector Machine (SVM) is extremely popular in supervised learning algorithm which is used for Classification and Regression technique. It is considered as the most successful method in Machine learning. SVM uses hyper planes that gives maximum distance between two classes in chosen SVM is high dimensionality overcomes, high dimensionality means the number of input variables is relatively more than number of observations. SVM can be used as classifiers as well as predictor the hyper plane differentiates between the classes.

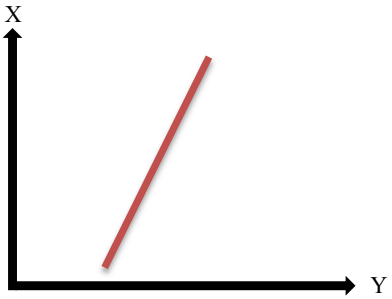

Fig. 2: Support Vector Machine

Support vectors are the closest data point set at the Hyperplane and are very important points in the data classification process because the best Hyperplane is determined in the data separation process, thus removing or changing its location requires another high level to be set again.

Margin is the distance between the Hyperplane and the closest data set point so that the larger this distance, the greater the likelihood of classifying new data correctly. So the importance of the margin is to get the best position for the Hyperplane.

SVM in People's Hospital datasets achieved accuracy of 98.9% by Shan Xu et al,[6] SVM used with boosting technique to give 84.81% [7].

Limitations of SVM is slow

Advantage is more robust

B. Naïve Bayes

Naive Bayes algorithm is simple and powerful supervised learning algorithm, and it depends on the Bayes theorem below. NB based on likelihood and probability and required small data for training in NB the existence class is independent from other classes this independence is very important in making a classification. Naïve Bayes algorithm is simplified predictive modeling it usually have high dimension training datasets.

$$p\left(\frac{A}{c}\right) = \frac{p\left(\frac{A}{c}\right)p(c)}{p(A)}$$

$$p(\frac{c}{A}) = p(\frac{A_1}{c})p(\frac{A_2}{c}) * \dots * P(\frac{A_n}{c})\,p(c)$$

As shown in the previous equation, P(c) (A) is given some prior probability of occurrence c.

C. Decision Tree Algorithm

The Decision Tree algorithm is represented by a flowchart or a tree-like structure. It's a supervised learning algorithm that's used to solve classification problems. It divides the data into smaller subsets where the inner subset (node) represents the dataset attributes, and the outer subset (branch) represents the outcome. This algorithm uses entropy to calculate homogeneity of the sample and information gain, then chooses the attribute with the highest information gain. Decision Tree is fast, dependable, and easy to use.[9]

$$Entropys\ (S) = \sum_{i=1}^{c} -p_i\ log2\ \ p_i$$

Information Gain:

$$Gain(S, A) = Entropys\ (S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|}\ Entropy(S_v)$$

In [7], M.A. Jabbar et al. utilized alternate decision trees (DT) in combination plus PCA to achieve 92.2% accuracy. In [8,] Kamran Farooq et al., the best results were obtained by combining a decision tree-based classifier with forward selection. and got 78.4604 % accuracy.

D. Logistic Regression

The logistic regression model is amongst the most powerful statistical model in predicting the likelihood of a specific class or event, such as success or failure. The logistical regression uses several predicted variables that can be digital or class. Logistical regression is also known by other names as logit or the general work of entropy. Logistics regression is part of supervised machine learning algorithms dedicated to "classification" tasks. Logistics regression is a simple and more efficient way for binary and linear classification problems that perform very well with linear separation layers. The logistics regression is used to predict the value of data based on the prior observation of a data set, it predicts a dependent data attribute by analyzing the relationship between one or more independent attribute.[10]

E. Random Forest

Random Forest algorithm is used for classification as well as for Regression and it is one of the best algorithms for classification, it has the ability for classifying huge amount of data[8], Random Forest technique, as the names indicates, is mainly composed of a huge number of independent decision trees that work as a group. From all the trees which have the same distribution, and an individual tree is unique. the prediction is averaged using bootstrap aggregation and random feature selections.

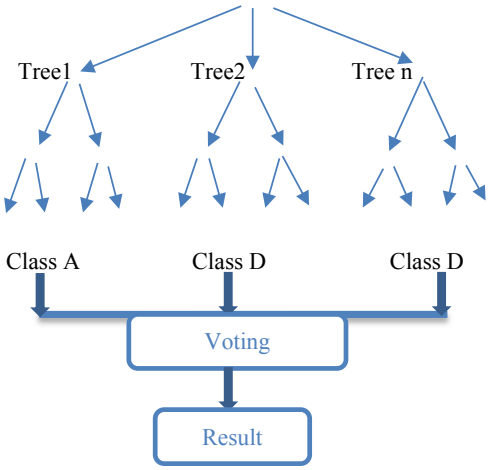

Fig. 3: Random Forest

F. K-Nearest Neighbour

K-Nearest Neighbor approach is a simple yet successful classification technique. There are no simplifying assumptions in it and is typically applied to classification issues where there is little or no prior knowledge about the distribution of the data. This approach comprises locating the nearest k data points in the training set towards the data point in which a target value is missing and allocating the average value of the obtained data points to it.

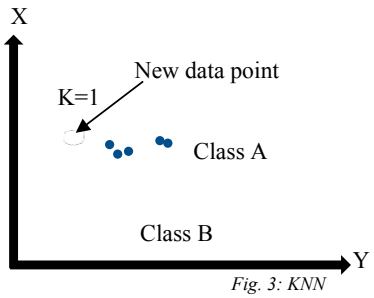

Fig. 3: KNN

Whenever the k value is equivalent to 9, and the 10-cross validation procedure is used, KNN has an accuracy of 83.16 percent. With an accuracy of 70.26 % as well as an error rate of 0.526, KNN with Ant Colony Optimization exceeds other methods in [8]. Ridhi Saini et al. achieved a very high performance of 87.5% . [11]

IV. CONCLUSION

In this work, a survey of Several Machine Learning techniques for predicting and detecting heart disease have been used., which generally is quite significant. Based on the above work, it is possible to conclude that machine learning algorithms have actually large potential for predicting and diagnosis cardiovascular illnesses or any heart-related diseases. With a large number of datasets, the Decision Tree method performs poorly. Random Forest scored extremely well as it addresses the issue of overfitting by combining numerous algorithms. (Many Decision Trees) The Nave Bayes classifier was highly quick and performed well in terms of computation. SVM improves efficiency in the vast majority of instances.

REFERENCES

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

[2] Abhay Kishore1, Ajay Kumar2, Karan Singh3,Maninder Punia4, Yogita Hambir5,"Heart Attack Prediction Using Deep Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 | Apr-2018.

[3] Vembandasamy, K., Sasipriya, R. and Deepa, E. (2015) Heart Diseases Detection Using Naive Bayes Algorithm. IJISET-International Journal of Innovative Science, Engineering & Technology, 2, 441-444.

[4] 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.

[5] Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan et al. "Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework", 2017 IEEE 2nd International Conference on Big Data Analysis.

[6] Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan et al. "Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework", 2017 IEEE 2nd International Conference on Big Data Analysis.

[7] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. " A Comprehensive Investigation and [8] Houda Mezrigui, Foued Theljani and Kaouther Laabidi et al. "Decision Support System for Medical Diagnosis Using a Kernel-Based Approach", ICCAD'17, Hammamet - Tunisia, January 19- 21, 2017.

[9] Amir Hussain, Peipei Yang, Mufti Mahmud and Jan Karasek et al. "A Novel Cardiovascular Decision Support Framework foreffective clinical Risk Assessment.", 978-1-4799-4527- 6/14/$31.00 ©2014 IEEE.

[10] D. Krishnani, A. Kumari, A. Dewangan, A. Singh, N.S. Naik, Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms, TENCON 2019 – 2019 IEEE Reg. 10 Conf., 2019, pp. 367–372

[11] Puneet Bansal and Ridhi Saini et al. "Classification of heart diseases from ECG signals using wavelet transform and kNN classifier",International Conference on Computing, Communication and Automation (ICCCA2015).