IJCSET – Volume 6, Issue 1 JANUARY 2020. ISSN: 2455-9091 Pages: 11-18

A SURVEY ON HEART DISEASE PREDICTION USING MACHINE LEARNING

[1]Ms. C. Keerthana, [2]Dr.B.Azhagusundari
[1] Assistant professor, [2] Associate professor,
[1, 2] Department of computer science,
[1, 2] NallamuthuGounderMahalingam College, Pollachi.

ABSTRACT - Heart disease is the one of the most common disease. This disease is quite common now days we used different attributes which can relate to this heart diseases well to locate the better method to predict and we additionally used algorithms for prediction. This survey paper explores a different models based on such algorithms and techniques and analyse their performance. Models based on supervised learning algorithms, Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaïveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models are discovered very mainstream among the researchers.

Keyword: [Machine Learning algorithm, K- Nearest Neighbour, Naïve Bayes, Decision Tree, Random Forest,Heart disease Prediction.]

1. INTRODUCTION

Heart disease is one of the significant causes of death all through the world. It can't be easily predicted by the medical practitioners as it is a troublesome task which demands expertise and higher knowledge for prediction. Cardiovascular Diseases (CVDs) are the principle reason for a huge number of deaths on the planet over the most recent couple of decades and has emerged as the most life-threatening disease, in India as well as in the whole world. In this way, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to different medical datasets to automate the examination of large and complex data. An automated system in medical diagnosis would enhance medical efficiency and furthermore reduce costs. The prediction of heart disease requires a huge size of data which is excessively complex and massive to process and analyse by conventional techniques.

Machine Learning

Machine learning is an emerging region of artificial intelligence. Its essential center is to design systems, permit them to learn and make predictions based on the experience. It trains machine learning algorithms utilizing a training dataset to create a model. The model uses the new information data to predict heart disease. Utilizing machine learning, it detects hidden patterns in the info dataset to manufacture models. It makes accurate predictions for new datasets. The dataset is cleaned and missing values are filled. The model uses the new info data to predict heart disease and afterward tested for accuracy. Machine learning techniques are classified as:

i) Supervised Learning

The model is trained on a dataset that is labelled. It has input data and its outcomes. Data are classified and split into training and test dataset. Training dataset trains our model while testing dataset capacities as new data to get accuracy of the model. The dataset exists with models and its yield. The classification and regression are its example.

ii) Unsupervised Learning

Data used to prepare are not classified or labelled in the dataset. Point is to discover hidden patterns in the data. The model is trained to develop patterns. It can easily predict hidden patterns for any new info dataset, yet after exploring data; it makes determination from datasets to describe hidden patterns. In this technique, no responses in the dataset are seen. The clustering method is an example of an unsupervised learning technique.

In this survey, extract hidden patterns by applying data mining techniques, which are noteworthy to heart diseases and to predict the presence of heart disease in patients where the presence is valued on a scale and is to discover the suitable machine learning technique that is computationally efficient just as accurate for the prediction of heart disease. Data mining combines Statistical investigation machine learning and database technology to extract hidden patterns and relationships from large databases. The implementation of work is done on Cleveland heart diseases data set from the University of California Irvine (UCI) machine learning repository to test on different data mining techniques. This paper explores different models based on such algorithms and techniques andanalyze their performance. Models based on supervised learning algorithms, for example, Support Vector Machines (SVM), K-Nearest Neighbor (KNN), NaïveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models are discovered very well known among the researchers.

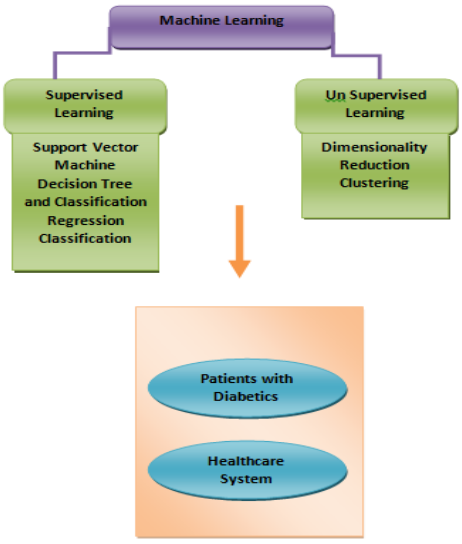

Figure 1: Machine Learning Techniques

2. LITERATURE SURVEY

1. M. Ganesan, Dr. N. Sivakumar (2019), et.al proposed IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. In this research, an efficient framework is utilized for heart disease is created using the UCI Repository dataset just as the healthcare sensors to predict the public who suffer from heart disease. Moreover, classification algorithms are used to order the patient data for the identification of heart disease. In the training phase, the classifier will be trained utilizing the data from benchmark dataset. During the testing phase, the genuine patient data to identify disease is used to identify the presence of disease. For experimentation, a benchmark dataset is tested utilizing a set of classifiers namely J48, logistic regression (LR), multilayer perception (MLP) and support vector machine (SVM). The re-enactment results ensured that the J48 classifiers shows superior performance in terms of different measures, for example, accuracy, precision, recall, F-score and kappa value.

2. K. M. ZubairHasan, ShourobDatta, MdZahidHasan, NusratZahan (2019), et.al proposed Automated Prediction of Heart Disease Patients using Sparse Discriminant Analysis. In this paper, we propose a novel classifier SDA - Sparse Discriminant Analysis method for heart disease detection. The time complexity will be reduced in this algorithm by ideal scoring investigation of LDA and will be comprehensive to execute sparse separation through the blend of Gaussians if limits between classes are nonlinear or if subgroups are available inside every class. On the whole, compared to previous techniques, our proposed technique is more appropriate for the diagnosis of heart disease patients with higher accuracy.

3. ShivendraKaura, AssemChandel, Nitin Kumar Pal (2019), et.al proposed Heart disease-Sinus arrhythmia prediction system by neural network using ECG analysis. The processed data involves the attributes or the physiological variables in the data set to discover a prediction likelihood or future state values of an alternative attributes. The

Description will emphasize on the discovering a typical recognizable patterns that describes that data that can be understood by humans. The project we observed that our project can be induced to detect the disease called sinus arrhythmia which is most prevailing these days. With the help of our trained neural network we can more easily and efficiently predict the sinus arrhythmia and its behavior in a person.

4. Halima EL HAMDAOUI, Saïd BOUJRAF, Nour El Houda CHAOUI, Mustapha MAAROUFI (2019), et.al proposed A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques. Heart disease is a leading cause of death worldwide. However, it remains hard for clinicians to predict heart disease as it is a complex and expensive task. Hence, we proposed a clinical support system for predicting heart disease to help clinicians with symptomatic and make better decisions. Machine learning algorithms, for example, Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, Random Forest, and Decision Tree are applied in this examination for predicting Heart Disease utilizing risk factors data retrieved from medical files. Several experiments have been conducted to predict HD utilizing the UCI data set, and the outcome reveals that Naïve Bayes outperforms utilizing both cross-validation and train-test split techniques with an accuracy of 82.17%, 84.28%, respectively. The second end is that the accuracy of all algorithm decrease after applying the cross-validation technique. At last, suggested multi validation techniques in prospectively collected data towards the endorsement of the proposed approach.

5. TülayKarayilan, ÖzkanKiliç (2017), et.al proposed Prediction of Heart Disease Using Neural Network. The proposed heart disease prediction system has been designed as a Multilayer Perceptron Neural Network. For the system Cleveland dataset was used. The neural network in the system used 13 clinical data which are obtained from Cleveland Dataset as info. It was trained with Backpropagation Algorithm to predict whether heart disease present or not in the patient. In this paper, a heart disease prediction system which uses artificial neural network backpropagation algorithmis proposed. 13 clinical features were used as contribution for the neural network and afterward the neural network was trained with backpropagation algorithm to predict absence or presence of heart disease with accuracy of 95%.

6. K. Prasanna Lakshmi, Dr.C.R.K.Reddy (2015), et.al proposed Fast Rule-Based Heart Disease Prediction using Associative Classification Mining. This work presented our experiences mining stream association rules from medical data to predict diseases. We used a novel dynamic tree to handle streaming data. As there is an increasing rate of death worldwide due to heart disease we designed a decision support system called SACHDP which helps to identify the risk score for predicting the heart disease. In this paper we proposed an efficient technique for heart disease prediction. This research uses associative classification which manufactures a classifier with prediction rules of high interestingness values. Experimental results show that this work helps doctors in their diagnosis decisions.

7.Can Xiao, Yi Li, Yimin Jiang (2020), et.al proposed Heart coronary artery segmentation and disease risk warning based on a deep learning algorithm. In the experiment, it found that simple data expansion might be detrimental to the test data. From the training curve, it is believed that with the improvement of the nature of training data, the performance of coronary artery segmentation can be further improved, and it is of great significance to provide doctors and patients with more accurate and efficient conclusions and suggestions in clinical practice to improve the nature of diagnosis and treatment. The purpose of helping experts in realtime diagnosis and investigation achieved.

8. Seyed Mohammad JafarJalali, Mina Karimi, Abbas Khosravi, Nahavandi (2019), et.al proposed An efficient Neuroevolution Approach for Heart Disease Detection.In this paper, we use these techniques for early detection of CAD by

applying them on a well-known CAD dataset named Z-Alizadeh sani. Hence, an effective natureinspired optimization algorithm named Multi-verse optimizer (MVO) based on Multilayer perceptron (MLP) training just as nine states of the craftsmanship supervised learning techniques are employed for CAD prediction. As this dataset has 54 features, before applying the supervised learning algorithms, we used a feature selection method to identify the best features. This procedure enhances the prediction ability of the utilized algorithms. The classification rates of all algorithms are compared with each other utilizing the most usable evaluation metrics including accuracy and area under the curve. Eventually, the experimental results show that the most appropriate model to characterize CAD patients is the MLP model trained by MVO among all other nine supervised learning methods.

9. Ali Mirza Mahmood, Mrithyumjaya RaoKuppa (2010), et.al proposed Early Detection Of Clinical Parameters In Heart Disease By Improved Decision Tree Algorithm. In this paper, we propose a new pruning method which is a mix of pre-pruning and post-pruning, pointing on both classification accuracy and tree size. Based upon this method, we induce a decision tree. The experimental results are computed by utilizing 18 benchmark datasets from UCI Machine Learning Repository. The results, when compared to benchmark algorithms, indicate that our new tree pruning method considerably reduces the tree size and increases the accuracy in general. We have likewise conducted a case investigation of heart disease dataset by utilizing our improved algorithm. This examination suggests that, type of defect in heart is the main predictor for affirming the presence of heart disease.

10. Aanshi Gupta, ShubhamYadav, ShaikShahid, VenkannaU (2019), et.al proposed HeartCare: IoT based heart disease prediction system. This paper expects to develop a ML based model to detect heart diseases. In this case, KNN outstands as the best algorithm in contrast with other algorithms, for example, Random Forest, Decision Tree, Support Vector Machine and Naive Bayes. Furthermore, a prototype is developed to validate the results. The prototype consisted of a set of sensors to screen the health of a person that causes heart diseases. It is at last predicted whether a person is prone to suffer from heart disease or not based on the model trained previously. In this way, our answer provides a critical human benefit as well as enables proactive health checking data with a predicting accuracy of 88.52%.

11. ElzhanZeinulla, Karina Bekbayeva, Adnan Yazici (2019), et.al proposed Effective diagnosis of heart disease imposed by incomplete data based on fuzzy random forest. This research presents data preprocessing and attribution techniques for creating a model from medical sensor data. We intend to solve the problem of creating a framework to diagnose heart diseases with an incomplete and messy data, which is basic with medical data. The medical dataset is often incomplete and messy due to its little size, imbalance and many missing, false, inaccurate data. In this examination, we utilize the synthetic minority oversampling technique with the blend of Tomek links to increase the size and eliminate the imbalance of the dataset. We performed a number of experiments and measurements on the Cleveland dataset and conducted a comparative investigation of different prediction models with recent algorithms in the literature. To process extra data from Budapest, Zurich and Basel, we apply the technique of semisupervised pseudo-labeling, which means that the model has been trained on unlabeled data and combined with labeled data by predicting unlabeled values and making them pseudo-labeled. The last accuracy of the methodology proposed in this investigation is 93.4%, with the specificity and sensitivity values of 96.92% and 89.99%, respectively, which is superior to previous models included in the literature.

12. Mamatha Alex P and Shaicy P Shaji (2019), et.al proposed Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique. The point of this project is to diagnose different heart diseases and to make all possible precautions to

prevent at early stage itself with affordable rate. We follow 'Data mining' technique in which attributes are fed in to SVM, Random forest, KNN, and ANN classification Algorithms for the prediction of heart diseases. The preliminary readings and studies obtained from this technique is used to know the chance of detecting heart diseases at early stage and can be completely cured by proper diagnosis.

3. PROPOSED METHODS, MERITS AND DEMERITS

Authors Name & Year	Proposed Methods	Merits	Demerits
1. M. Ganesan, Dr. N. Sivakumar (2019),	IoT based heart disease prediction and diagnosis model for healthcare using machine learning models	1. It is clear that J48 classifier is discovered to be the appropriate algorithm for the IoT based healthcare prediction model for heart disease compared to MLP, SVM and LR classifiers.	1. At the same time, the SVM and LR classifiers shows practically equal performance with a recall value of 84.10 and 83.70 respectively. At long last, it is reported that the worse classification performance.
2.K. M. ZubairHasan , ShourobDatta , MdZahidHasan , NusratZahan (2019),	Automated Prediction of Heart Disease Patients using Sparse Discriminant Analysis	1. Our proposed method improved the prediction accuracy of 96%. 2. Successively improve the accuracy, time complexity and memory efficiency of this SDA based machine learning model that we can entirely implementation this as the cloud-based system on a real-life scenario that it would be beneficial for the physicians to detect heart disease more accurately.	1. An earlier stage whereas their accuracy level of heart disease prediction is below 90%
3. ShivendraKaura, AssemChandel, Nitin Kumar Pal (2019),	Heart disease-Sinus arrhythmia prediction system by neural network using ECG analysis.	1. The algorithm and the procedure that we embedded in our project could be further used to prepare with different other physiological sets to predict different type of other cardiovascular diseases which are more draw out and destructive to the society.	1. The whole system ought not to be integrated into a single software so that isn't easy to for an ordinary person to operate, and economically.

4.Halima EL HAMDAOUI, Saïd BOUJRAF , Nour El Houda CHAOUI , Mustapha MAAROUFI (2019),	A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques	1. This technique is the best in our model since the used data-set isn't large so the process didn't take quite a while, at the same time we solved the problem of over fitting	1. It could improve the knowledge on the prediction of heart disease risk through better diagnoses and interpretation.
5. TülayKarayilan , ÖzkanKiliç (2017),	Heart Disease Using Neural Network	1. The proposed system gives 95% accuracy rate which means a very decent rate as indicated by related studies on this field.	1. It can't be enhanced as a half breed model with other classification algorithms to get more accurate diagnosis for heart disease.
6.K.Prasanna Lakshmi, Dr.C.R.K.Reddy (2015),	Fast Rule-Based Heart Disease Prediction using Associative Classification Mining	1. Experimental results show that SACHDP performance better when compared to other associative classification techniques.	1. Still it isn't support the performance of SACHDP by reducing the number of rules generated.
7.Can Xiao, Yi Li, Yimin Jiang (2020),	Heart coronary artery segmentation and disease risk warning based on a deep learning algorithm.	1. The results show that the model training effect of the centerlinepreprocessing is superior to the first data. 2. The experimental results show that the best effect reaches the dice coefficient of 0.8291.	1. FCN likewise has shortcomings, that is, the results obtained are still relatively fluffy and smooth, and are not sensitive to the details in the image. Therefore, it isn't suitable for more elaborate medical images.
8.Seyed Mohammad JafarJalali, Mina Karimi , Abbas Khosravi , Nahavandi(2019),	An efficient Neuroevolution Approach for Heart Disease Detection using CAD	1. The proposed approach and even its multi-class classification version of this methodology can be investigated for medical, power systems, business, environment, and mechanical applications.	1. It wouldn't be worth to compare the proposed method used in this paper with other existing mixture evolutionary neural network methods to investigate its efficacy in the area of CAD detention.
9.Ali MirzaMahmood, MrithyumjayaRaoKuppa (2010),	Early Detection Of Clinical Parameters In Heart Disease By Improved Decision Tree Algorithm	1. The proposed approach considerably reduces the size of the tree while retaining or improving the classification accuracy.	1. It doesn't to analyze a huge cancer disease database.

10. Aanshi Gupta, ShubhamYadav, ShaikShahid, Venkanna U(2019),	HeartCare: IoT based heart disease prediction system using KNN	1. It giving an efficient and real-time heart disease prediction system for proactive health observing, which can work on live data feed from the sensors.	1. Many of the sensors which were unavailable currently, to get more efficient results. 2. This project can't be extended for the observing and prediction of other diseases like diabetes.
11.ElzhanZeinulla, Karina Bekbayeva, Adnan Yazici (2019),	Effective diagnosis of heart disease imposed by incomplete data based on fuzzy random forest	1. This methodology has improved the performance of each metric, that is, essentially better than the results of the previous published results. 2. The accuracy of the model is 93.45% with a specificity and sensitivity of 96.92% and 89.99%, respectively. 3. In expansion, it tends to be adapted to keen home technologies or used as an emergency call to medical foundations.	1. The primary problem is 3 completely missing columns in the piece of the dataset.
12. Mamatha Alex P and Shaicy P Shaji(2019),	Prediction and Diagnosis of Heart Disease Patients using Data Mining Techniques (KNN, SVM & ANN).	1. These attributes are fed in to SVM, Random forest, KNN, and ANN classification Algorithms in which ANN gave the best result with the highest accuracy.	1. It is sensitive to the nearby structure of the data.

CONCLUSION

Based on the above review, it tends to be concluded that there is a huge scope for machine learning algorithms in predicting heart related diseases. Each of the above-mentioned algorithms has performed extremely well in some cases yet inadequately in some other cases. Alternating decision trees when used with PCA have performed extremely well yet decision trees have performed very inadequately in some other cases which could be due to overfitting. Random Forest and Ensemble models have performed very well because they solve the problem of overfitting by employing multiple algorithms (multiple Decision Trees in case of Random Forest). Models based on Naïve Bayes classifier were computationally very quick and have likewise performed well. SVM performed extremely well for a large portion of the cases. Systems based on machine learning algorithms and techniques have been very accurate in predicting the heart related diseases yet there is a ton scope

of research to be done on the best way to handle high dimensional data and overfitting. A ton of research should likewise be possible on the correct ensemble of algorithms to use for a specific type of data.

REFERENCES

[1]. M. Ganesan ,Dr.N.Sivakumar (2019), "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models ", DOI: 10.1109/ICSCAN.2019.8878850, ISBN: 978-1-7281-1525-2, IEEE.

[2]. K. M. ZubairHasan , ShourobDatta , MdZahidHasan , NusratZahan(2019), "Automated Prediction of Heart Disease Patients using Sparse Discriminant Analysis ", DOI: 10.1109/ECACE.2019.8679279, ISBN: 978-1-5386-9111-3, IEEE.

[3]. ShivendraKaura, AssemChandel ,Nitin Kumar Pal (2019), "Heart disease-Sinus arrhythmia prediction system by neural network using ECG analysis ", DOI: 10.1109/PEEIC47157.2019.8976829 , ISBN: 978-1-7281-1793-5, IEEE.

[4]. Halima EL HAMDAOUI, Saïd BOUJRAF , Nour El Houda CHAOUI , Mustapha MAAROUFI (2019), " A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques", DOI: 10.1109/ATSIP49331.2020.9231760, ISBN: 978-1-7281-7513-3 ,IEEE.

[5]. TülayKarayilan ,ÖzkanKiliç (2017), "Prediction of Heart Disease Using Neural Network ", DOI: 10.1109/UBMK.2017.8093512, ISBN: 978-1-5386-0930-9, IEEE.

[6]. K.Prasanna Lakshmi, Dr.C.R.K.Reddy (2015), "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining ", DOI: 10.1109/IC4.2015.7375725, ISBN: 978-1-4799-8164-9, IEEE.

[7]. Can Xiao ,Yi Li, Yimin Jiang (2020), "Heart coronary artery segmentation and disease risk warning based on a deep learning algorithm ", DOI10.1109/ACCESS.2020.3010800, ISBN: 2169-3536, IEEE.

[8]. Seyed Mohammad JafarJalali, Mina Karimi , Abbas Khosravi ,Nahavandi(2019), "An efficient Neuroevolution Approach for Heart Disease Detection ", DOI10.1109/SMC.2019.8913997, ISBN: 978-1-7281-4569-3, IEEE.

[9]. Ali MirzaMahmood, MrithyumjayaRaoKuppa(2010), "Early Detection Of Clinical Parameters In Heart Disease By Improved Decision Tree Algorithm ", DOI 10.1109/VCON.2010.12, ISBN: 978-1-4244-9628-0, IEEE.

[10]. Aanshi Gupta, ShubhamYadav, ShaikShahid, Venkanna U,(2019), "HeartCare: IoT based heart disease prediction system ", DOI**:** 10.1109/ICIT48102.2019.00022, ISBN: 978-1-7281-6052-8, IEEE.

[11]. ElzhanZeinulla , Karina Bekbayeva , Adnan Yazici (2019), "Effective diagnosis of heart disease imposed by incomplete data based on fuzzy random forest ", DOI**:** 10.1109/FUZZ48607.2020.9177531, ISBN:978-1-7281-6932-3, IEEE.

[12]. Mamatha Alex P and Shaicy P Shaji (2018), "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique ", DOI: 10.1109/ICCSP.2019.8697977, ISBN: 978-1-5386-7595-3, IEEE.

[13]. "HEART DISEASE PREDICTION USING MACHINE LEARNING", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN: 2349-5162, Vol.7, Issue 6, page no.2081-2085.

[14]. V.V. Ramalingam, AyantanDandapath, M Karthik Raja (2020), "Heart disease prediction using machine learning techniques",325116774, IJAT.