IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Survey On Cardiovascular Disease Prediction Using Deep Learning and Feature Selection

[2]Shrivatsa . S . Desai, [2]Suraj Bhosale, [3]Kunal Gajmal, [4]Aniket Manjare

Department Of Computer Engineering, RMD Sinhgad School of Engineering, Savitribai Phule Pune University, Warje, Pune, India

Abstract: Cardiovascular diseases are one of the leading causes of death worldwide, killing 17.9 million people annually. Coronary heart disease, cerebrovascular disease, rheumatic heart disease and other diseases are one of the categories of cardiovascular diseases known as cardiovascular diseases. More than four out of five deaths from myocardial infarction are related to heart attack and stroke, with one-third occurring before the age of 70. Unhealthy diet, lack of exercise, use of tobacco products and alcohol consumption are the main behavioral risk factors for the heart. Disease and stroke People may experience high blood pressure, high blood sugar levels, high blood lipids, and even overweight or obesity as a result of behavioral risk factors. Approximately 90 percent of cardiovascular disease is preventable if predicted in advance, providing clinicians with important insights that allow them to tailor diagnosis and treatment on a patient-by-patient basis. Machine learning techniques are one of the most successful techniques for forecasting. Various studies have only mentioned the use of ML systems to predict heart disease. One of the biggest challenges today is predicting disease and its causes. This study presents a new method for identifying large datasets using deep learning algorithms to improve the accuracy of cardiovascular disease prediction. This model used artificial neural networks and feature selection to predict cardiovascular diseases. ANN is an ML technique that can be used to produce useful predictive results and feature selection to select the most appropriate features from a dataset.

***Index Terms* – Artificial Neural Network(ANN), Feature Selection, Deep Learning, Cardiovascular disease prediction(CVD).**

I. Introduction

Cardiovascular disease describes several conditions that affect the heart. Cardiovascular disease (CVD) is also referred to as heart disease. Over the past decade, heart disease has been the leading cause of death worldwide. It is necessary for the diagnostic system to generate accurate results so that the correct treatment is available to the patient. In earlier methods, different procedures were tried to determine the cause of the disease. Elements that can be edited and elements that cannot be edited are divided into two distinct groups. Risk factors for heart disease include age, smoking, heredity, gender, high blood pressure, poor diet, alcohol consumption and lack of exercise, although the results were not as precise.

Machine learning is proving to be effective in making decisions and making predictions from the vast amount of data generated by the healthcare sector. Machine learning techniques can be beneficial in this regard. Although heart disease can come in many forms, there is a common set of underlying risk factors that influence whether or not someone will ultimately be at risk for heart disease. It was now important to develop a new system that can predict cardiovascular problems in an easy and less expensive way due to the increasing number of heart diseases.

II. Literature Survey

[1] Tulasi Krishna Sajja and Hemantha Kumar Kalluri proposed a model for heart disease prediction using CNN on the UCI dataset and got an accuracy of 94%. They afterwards compared the model accuracy of various traditional machine learning algorithms like SVM, KNN, and Naive Bayes with the proposed model.

[2] Dinesh Kumar G, Santhosh Kumar D, Arumugaraj K, and Mareeswari V perform various data preprocessing techniques like removal of noisy data, removal of missing data, and filling default values, and then compare the model accuracies on various machine learning algorithms like SVM, KNN, Gradient Boosting, Random Forest and Naïve Bayes. In the future, different ensemble methods of these algorithms will be explored in the following work in order to improve performance with additional parameter settings like accuracy, sensitivity, and specificity analysis.

[3] Tsatsral Amarbayasgalan, Van-Huy Pham, Nipon Theera-Umpon, Yongjun Piao, and Keun Ho Ryu proposed two models for heart disease prediction. The first one is using the PCA and variational autoencoder, and the second one is a comparison of the proposed model against different machine learning algorithms using Naïve Bayes, Random Forest, K-Nearest Neighbour, Decision Tree, Support Vector Machine, and Adaptive Boosting. The first one outperforms the second one with an accuracy of 89%, a specificity of 0.840, a precision of 0.911, a recall of 0.920, an f-measure of 0.915, and an AUC of 0.882.

[4] Chaurasia Vikas and Pal Saurabh provide a model of three classifiers such as Naive Bayes, Decision Tree, and Bagging algorithms, and feature selection techniques are used to reduce the number of features to train the data, which will enable us to further improve the accuracy of the prediction of heart disease.

[5] Bhuvaneswari Amma N.G. combined the Neural Network and Genetic algorithm for the classification problem of predicting heart disease on the UCI. The multilayer feed-forward network is used as it is best suited for complex classification problems, and the genetic algorithm accepts a good set of weights in a small number of iterations. The proposed combined model of a neural network and a genetic algorithm achieves an accuracy of 94.17%.

[6] Aniruddha Duttaa, Tamal Batabyal, Meheli Basue and Scott T. Acton proposed a heart disease prediction model based on Convolutional Neural Network on the class imbalanced National Health and Nutritional Examination Survey (NHANES) dataset with an accuracy of 85.7%.

[7] Gihun Joo, Yeongjin Song, Hyeonseung IM, and Junbeom Park developed ML-based models on Korean National Health Insurance Service-National Health Sample Cohort (KNHSC) data using logistic regression, deep neural networks, Random Forests, and LightGBM and validated them against different metrics such as receiver operating characteristic curves, precision-recall curves, sensitivity, specificity, and F1 score. The ML-based algorithms outperform the baseline method derived from the ACC/AHA guidelines for estimating the 10-year risk for cardiovascular diseases.

[8] Furqan Rustam, Abid Ishaq, Ashif Munir, Mubarak Almutairi, Naila Aslam and Imran Ashraf try to build a model by proposing a novel use of feature selection with a Convolutional Neural Network model. A CNN model is used to widen the feature set for training linear models such as the stochastic gradient descent classifier, logistic regression, and support vector machine, which form the ensemble model based on soft voting.

[9] Mallikarjuna Shastry Pm and Ashok Kumar P S proposed a model that combines AI and information mining to achieve accurate results with the fewest errors.

[10] L. J. Muhammad, Ibrahem Al Shourbaji, Ahmed Abba Haruna, I. A. Mohammed, Abdulkadir Ahmad, and Muhammed Besiru Jibrin applied various ML based algorithms like support vector machine, K nearest neighbour, random tree, Naïve Bayes, gradient boosting, and logistic regression to CAD datasets.Machines based on random forests are more accurate. The best model was the learning model, which had an accuracy rate of 92%. A random forest based machine learning model came out as the best model, with 92.40%, while a support vector machine-based machine learning model emerged as the best model, with 87.34%, and a ROC-based machine learning model emerged as the best model, with 92.20%.

[11] Sina Dami and Mahtab Yahaghizadeh developed the LSTM-DBN model, which is a combination of long-short-term memory and deep belief networks, and that is compared with various deep learning and machine learning models. The results show that the proposed LSTM DBN model outperformed the various ML and deep learning algorithms.

[12] A. K. Paul, P. C. Shill, M. R. I. Rabin, and M. A. H. Akhand developed a fuzzy decision support system for predicting the degree of heart disease risk based on a genetic algorithm. Our suggested fuzzy decision support system analyses datasets, chooses a few key features based on several techniques, and then develops a weighted fuzzy decision support system (FDSS) that aids in the prediction of cardiac illnesses.

[13] Mienye, Domor & Sun, Yanxia & Wang, and Zenghui proposed an improvised machine learning approach on Cleveland and Framingham datasets that involves randomly sampling data into smaller subsets using a mean-based splitting approach, and then various subsets are modelled using classification and regression trees. An ensemble model is then created from different CART models using an accuracy based weighted ageing classifier ensemble, which is a modification of the weighted ageing classifier ensemble (WAE). The approach gives an accuracy of 91% and 93% on the Cleveland and Framingham datasets, respectively.

[14] Abdeldjouad, Fatma Zahra, Menaouer, Brahami, and Nada, Matta, suggested a hybrid model to predict cardiovascular disease via multiple machine learning techniques such as Logistic Regression (LR), Adaptive Boosting (AdaBoostM1), MultiObjective Evolutionary Fuzzy Classifier (MOEFC), and Fuzzy Unordered Rule Induction (FURIA), Genetic Fuzzy System-LogitBoost (GFSLB), and Fuzzy Hybrid Genetic Based Machine Learning (FH-GBML), and then the accuracies and results of each classifier are compared.

[15] Javid, Irfan, Zager, Ahmed, and Ghazali, Rozaida proposed a model that uses a number of machine learning and deep learning techniques, such as Random Forest, KNN, and Long Short-Term Memory (LSTM), which provides satisfactory accuracy. However, to increase accuracy, the weak classifiers are integrated based on the voting-based model, which is the model's strength, which in turn helps enhance the accuracy of the model by 2.1%.

[16] Christalin, Beulah and Jeeva, Carolin proposed an ensemble model of various classifiers that is modelled to improve the accuracy of the model. The comparative analysis demonstrates how ensemble methods contribute to improving the accuracy of models. The proposed model further improves accuracy by a maximum of 7% on weak classification algorithms.

[17] Sumit Sharma and Mahesh Parmar proposed the model based on Deep Neural Network which in turn helps to improve the classification accuracy to predict heart disease.

III. METHODOLOGIES USED

Artificial Neural Network:

The term "artificial neural network" comes from biological neural networks that evolve the structure of the human brain. Artificial neural networks also have interconnected neurons at different levels of the network, like the human brain's interconnected neurons. Nodes are the names of these neurons.

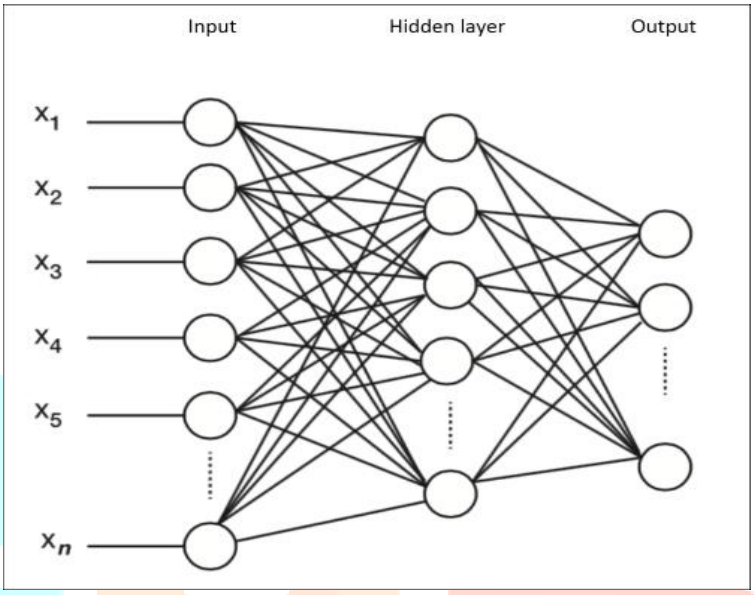

Fig 1: Structure of ANN

Artificial Neural Network primarily consists of three layers:

The input layer can receive input from programmers in a variety of formats. Between the input layer and the output layer, there is a hidden layer. The input is modified using hidden layers through a sequence of calculations to uncover hidden features and patterns, and then the output is transmitted to the output layer using this layer. When inputs are received, an artificial neural network computes a weighted sum of the inputs, taking biases into account. A transfer function is used to represent the results of this calculation. In order to produce the required result, a group of transfer functions is referred to as the activation function. Numerous activation function types exist, most of which are linear or comprise a group of non-linear functions like binary, linear, tangent-hyperbolic, and sigmoidal.

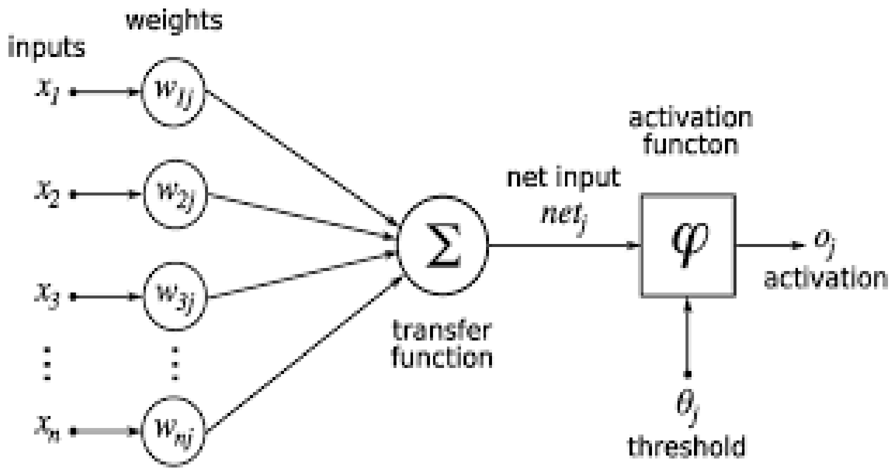

Fig 2: ANN Structure of Single Neuron

Feature Selection:

Feature selection is a method of reducing input variables in a model by using relevant data and removing noise in the data. This involves automatically selecting the features of the machine learning model that are important to the problem you are trying to solve. This is done by adding or removing important functionality without changing it. This helps reduce the amount of noise in your data and the amount of input data.
Collect large amounts of data to improve model learning and training. Data sets usually consist of noisy data, irrelevant data, and very little useful information. Also, large amounts of data slow down the model training process. Noise and irrelevant data can prevent the model from accurately predicting the future.

Therefore, it is very important to remove this noise and insignificant data from the dataset, and for this, feature selection techniques are used.

There are many techniques for feature selection, which can be classified into three groups: filtering methods, packing methods, and embedding methods. Filtering methods select features according to statistical features such as correlation, mutual information, and chi-square tests. These techniques are computationally efficient and do not use machine learning algorithms. Pearson's correlation coefficient, ANOVA F test and ReliefF are examples of filtering techniques.

Wrapper method: Wrapper method uses a machine learning algorithm to evaluate the value of each attribute. A subset of traits is used to train the algorithm and its effectiveness is evaluated using specific criteria such as accuracy and F1 score. The best elements that improve performance are selected. These strategies are applied when the dataset is small despite the high computational cost. Recursive attribute elimination and forward selection are two examples of wrapper techniques.

Embedding Method: The embedding method selects the features in the training phase of the machine learning algorithm. The algorithm automatically selects which feature is necessary for the current task. These techniques have good computational performance and are often applied to large data sets. Lasso, Ridge Regression and Decision Trees are some types of embedding techniques.

IV. CONCLUSION

In this study, it has been shown that different algorithms based on machine learning are useful in predicting heart disease, and different techniques such as feature selection, deep belief networks, CNN, or some reinforcement and ensemble techniques can also help improve the accuracy of heart prediction. Sickness. Model. The models are also validated against various measures such as exact recall curves, sensitivity, specificity, and F1 score. In the future, various models may be developed based on algorithms and machine learning techniques that are tolerant of inaccuracies and can provide accurate results for the treatment of cardiovascular diseases.

REFERENCES

[1] Kalluri, Hemantha kumar & Tulasi Krishna, Sajja. (2020). A Deep Learning Method for Prediction of Cardiovascular Disease Using Convolutional Neural Network. Revue d intelligence artificielle. 34. 601-606. 10.18280/ria.340510.

[2] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, 2018, pp. 1-7, doi: 10.1109/ICCTCT.2018.8550857.

[3] T. Amarbayasgalan, V. -H. Pham, N. Theera-Umpon, Y. Piao and K. H. Ryu, "An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets," in IEEE Access, vol. 9, pp. 135210-135223, 2021, doi: 10.1109/ACCESS.2021.3116974.

[4] Chaurasia, Vikas & Pal, Saurabh. (2013). Data Mining Approach to Detect Heart Diseases. International Journal of Advanced Computer Science and Information Technology (IJACSIT). 2. 56-66.

[5] N. G. B. Amma, "cardiovascular disease prediction system using genetic algorithm and neural network," 2012 International Conference on Computing, Communication and Applications, Dindigul, India, 2012, pp. 1-5, doi: 10.1109/ICCCA.2012.6179185.

[6] Dutta, Aniruddha & Batabyal, Tamal & Basu, Meheli & Acton, Scott. (2020). An Efficient Convolutional Neural Network for Coronary Heart Disease Prediction. Expert Systems with Applications. 159. 113408. 10.1016/j.eswa.2020.113408.

[7] G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in IEEE Access, vol. 8, pp. 157643-157653, 2020, doi: 10.1109/ACCESS.2020.3015757.

[8] Rustam, Furqan & Ishaq, Abid & Munir, Kashif & Almutairi, Mubarak & Aslam, Naila & Ashraf, Imran. (2022). Incorporating CNN Features for Optimizing Performance of Ensemble Classifier for Cardiovascular Disease Prediction. Diagnostics (Basel, Switzerland). 12. 10.3390/diagnostics12061474.

[9] Patil, Prasadgouda & Pm, Mallikarjuna & P S, Ashok. (2021). Journal of Critical Reviews MACHINE LEARNING BASED ALGORITHM FOR RISK PREDICTION OF CARDIO VASCULAR DISEASE (CVD). Journal of Critical Reviews. 7. 2020. 10.31838/jcr.07.09.157.

[10] Muhammad LJ, Al-Shourbaji I, Haruna AA, Mohammed IA, Ahmad A, Jibrin MB. Machine Learning Predictive Models for Coronary Artery Disease. SN Computer Sci. 2021;2(5):350. doi: 10.1007/s42979-021-00731-4. Epub 2021 Jun 22. PMID: 34179828; PMCID: PMC8218284.

[11] Dami, Sina & Yahaghizadeh, Mahtab. (2021). Predicting cardiovascular events with deep learning approach in the context of the internet of things. Neural Computing and Applications. 33. 1-18. 10.1007/s00521-020-05542-x.

[12] A. K. Paul, P. C. Shill, M. R. I. Rabin and M. A. H. Akhand, "Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 2016, pp. 145-150, doi: 10.1109/ICIEV.2016.7759984.

[13] Mienye, Domor & Sun, Yanxia & Wang, Zenghui. (2020). An improved ensemble learning approach for the prediction of heart disease risk. Informatics in Medicine Unlocked. 20. 100402. 10.1016/j.imu.2020.100402.

[14] Abdeldjouad, Fatma Zahra & Menaouer, Brahami & Nada, Matta. (2020). A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques. 10.1007/978-3-030-51517-1_26.

[15] Javid, Irfan & Zager, Ahmed & Ghazali, Rozaida. (2020). Enhanced Accuracy of Heart Disease Prediction using Machine Learning and Recurrent Neural Networks Ensemble Majority Voting Method. International Journal of Advanced Computer Science and Applications. 11. 10.14569/IJACSA.2020.0110369.

[16] Christalin, Beulah & Jeeva, Carolin. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked. 16. 100203. 10.1016/j.imu.2019.100203.

[17] Parmar, Mahesh. (2020). Heart Diseases Prediction using Deep Learning Neural Network Model. International Journal of Innovative Technology and Exploring Engineering. 9. 2244-2248. 10.35940/ijitee.C9009.019320.