Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

PREDICTING HEART DISEASE BY USING MACHINE LEARNING& ,KAGGLE DATASET

[1]MR.B.SURESH REDDY, [2]G.VENKATA NAGARAJU

[1] Assistant Professor in the department of MCA at QIS College of Engineering & Technology, Vengamukkapalem, Ongole, AP, India

[2] PG Scholar in the department of MCA at QIS College of Engineering & Technology, Vengamukkapalem, Ongole, AP, India.

ABSTRACT

The World Health Organization (WHO) estimates that 17.7 million people die unexpectedly each year as a result of cardiovascular diseases. People may profit from the ability to anticipate the complexity of their health at an early stage, owing to the heart disease prediction system. Traditional approaches for predicting heart illness include a doctor's examination and a number of medical tests, such as an ECG, stress test, cardiac MRI, and so on. The existing health care data contains a tremendous amount of hidden information. Having access to this hidden information helps you make better decisions. For satisfactory results, computer-based data and advanced data mining techniques are used.

INDEX TERMS Machine learning, heart disease prediction, feature selection, prediction model, classification algorithms, cardiovascular disease (CVD).

1. INTRODUCTION

Heart disease is an important public health issue that affects millions of people globally. It is the main cause of death in numerous countries, including the United States. Early diagnosis of cardiac disease is crucial for successful treatment and management, but existing diagnostic approaches are frequently expensive, time-consuming, and necessitate specialized equipment and knowledge.

Machine learning algorithms have the ability to help in early identification of heart disease by properly predicting the likelihood of the disease in patients based on their medical history and other relevant characteristics. Among these strategies, Artificial Neural Networks (ANNs) have shown great accuracy in forecasting the risk of heart disease.

The goal of this project is to create an ANN model that can accurately predict the likelihood of heart disease in patients using a set of medical features. The model will be trained and tested on a huge dataset of patient information before being evaluated for accuracy and performance.

The goal of this research is to provide healthcare practitioners with an accurate and accessible tool for forecasting the likelihood of heart disease in patients. The established model has the potential to aid in the early detection and treatment of cardiac disease, resulting in better patient outcomes and more efficient healthcare resource management.

The remaining sections of this paper are organized as follows: The next part will provide an overview of related research in the field of heart disease prediction using machine learning techniques. The following parts will cover the project's methodology and implementation, including dataset preparation, ANN model building, and performance evaluation. Finally, the conclusion will summarize the findings and emphasize any future improvements to the project.

2. LITERATURE SURVEY

[1] A. S. Abdullah and R. R. Rajalaxmi, "A data mining model for predicting the coronary heart disease using random forest classifier," 2012.

The proposed work is for the most part worried about the improvement of an information mining model with the Irregular Woods characterization calculation. The new model will be able to predict the various events that are associated with each patient record, prevent risk factors with the associated cost metrics, and improve overall prediction accuracy. Because of this, the causes and symptoms of each event will be determined in accordance with the patient's record, allowing for a significant reduction in CHD. Coronary heart disease (CHD) is a prevalent heart disease that is a significant contributor to premature death. From a medical science perspective, data mining is involved in the discovery of various metabolic syndromes. Prediction and data exploration are significantly aided by classification methods utilized in data mining. Decision trees and other classification methods have been used to predict CHD-related events and accuracy. A Random Forest classifier-based data mining model has been developed in this paper to increase prediction accuracy and investigate various CHD-related events. This model can help the clinical professionals for anticipating CHD with its different occasions and how it very well may be connected with various sections of the populace. Angina, Acute Myocardial Infarction (AMI), Percutaneous Coronary Intervention (PCI), and Coronary Artery Bypass Graft surgery (CABG) are the events being investigated. Exploratory outcomes have shown that order utilizing Irregular Woods Characterization calculation can be effectively utilized in anticipating the occasions and hazard factors connected with CHD.

Constant fatigue, physical disability, mental stress, and depression are all consequences of CHD. The development of a Random forest classification algorithm-based data mining model for the evaluation and prediction of various CHD-related events is the primary focus of this paper. A portion of these examinations, has made with the execution of information mining

calculation like K-NN, Credulous bayes, K-implies, ID3, and Apriority calculations. The design and development of novel methods in data mining technologies and allied medical informatics disciplines can significantly reduce the growing healthcare burden and suffering caused by life-threatening diseases like heart disease and the rising cost of drug development. In CHD, two types of issues can be resolved if the risk factors are anticipated in advance. First, angioplasty, coronary stents, coronary artery bypass grafting, and heart transplant can all be avoided to a large extent. Second, the related expense with each hazard element can be diminished .

[3] N. Al-milli, "Backpropagation neural network for prediction of heartdisease'' 2013

In this work, we present an approach that based on back propagation neural network to model heart disease diagnosis. In this research paper, a heart disease prediction system is developed using neural network. The proposed system used 13 medical attributes for heart disease predictions. The experiments conducted in this work have shown the good performance of the proposed algorithm compared to similar approaches of the state of the art

Moreover, new algorithms and new tools are continued to develop and represent day by day. Diagnosing of heart disease is one of the important issue and many researchers investigated to develop intelligent medical decision support systems to improve the ability of the physicians. Neural network is widely used tool for predicting heart disease diagnosis. In this research paper, a heart disease prediction system is developed using neural network. The proposed system used 13 medical attributes for heart disease predictions. The experiments conducted in this work have shown the good performance of the proposed algorithm compared to similar approaches of the state of the art.

[2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease," 2017.

The trial results show that the PSO calculation accomplished higher prescient precision and a lot more modest rule list than C4.5. In this paper we proposed PSO calculation for creation of best standards in expectation of coronary illness. The trials show that the standards found for the dataset by PSO are by and large with higher exactness, speculation and fathomability. In light of the typical exactness, the precision of the PSO technique is 87% and the exactness of C4.5 is 63%. By utilizing the PSO, one can separate compelling grouping rules with OK exactness. Besides, we presume that PSO calculation in rule creation has great execution for rule revelation on consistent information. For future work we consider utilizing further developed PSO calculation for creating the best guidelines in coronary illness informational collection.

Coronary illness is as yet a developing worldwide medical problem. In the medical services framework, restricting human experience and skill in manual analysis

prompts mistaken conclusion, and the data about different sicknesses is either deficient or ailing in exactness as they are gathered from different kinds of clinical gear. Since the right expectation of an individual's condition is vital, furnishing clinical science with smart instruments for diagnosing and treating disease can diminish specialists' slip-ups and monetary misfortunes. In this paper, the Molecule Multitude Enhancement (PSO) calculation, which is perhaps of the most remarkable transformative calculation, is utilized to produce rules for coronary illness. First the irregular principles are encoded and afterward they are enhanced in light of their exactness utilizing PSO calculation. At long last we contrast our outcomes and the C4.5 calculation.

The errand of arrangement turns out to be extremely challenging when the quantity of conceivable different mixes of boundaries is so high. The self-flexibility of developmental calculations relied upon populace is exceptionally valuable in rule extraction and choice for information mining.

[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya,and R. Deepika, ``Analysis of neural networks based heart disease predictionsystem,'' 2018.

We have presented an artificial neural network (ANN)-based heart disease prediction system (HDPS) in this research paper. The system is built using a multilayer perceptron neural network and a back propagation algorithm derived from the ANN. Because the MLPNN model produces realistically good results even without retraining, it helps domain experts and even people related to the field plan for a better diagnosis and provide the patient with early diagnosis results. The exploratory outcome shows that utilizing brain networks the framework predicts Coronary illness with almost 100 percent precision.

This secret data is helpful for settling on powerful choices. For appropriate outcomes, computer-based information and cutting-edge data mining methods are utilized. A popular method for predicting the diagnosis of heart disease is the neural network. A neural network-based Heart Disease Prediction System (HDPS) is developed in this research paper. The HDPS system can predict whether a patient will develop heart disease. For expectation, the framework utilizes sex, pulse, cholesterol like 13 clinical boundaries. For improved precision, obesity and smoking are added as two additional parameters.

From the outcomes, it has been seen that brain network foresee coronary illness with almost 100 percent exactness. Predication ought to be finished to lessen hazard of Coronary illness. A patient's signs, symptoms, and physical examination are typically used to make a diagnosis. Based on their education and experience, almost all doctors can predict heart disease. In the medical field, diagnosing a disease is a laborious and challenging process. Foreseeing Coronary illness from different variables or side effects is a diverse issue which might prompt misleading assumptions and capricious impacts. Medical services industry today creates a lot of complicated

information about patients, emergency clinics assets, sickness determination, electronic patient records, clinical gadgets and so forth. The large amount of data is an essential resource that must be processed and analyzed in order to extract knowledge that enables cost-saving support and decision making. For accurate diagnosis, human intelligence alone is insufficient..

3.PROPOSED SYSTEM

The proposed system employs machine learning, specifically artificial neural networks (ANNs), to create a predictive model capable of accurately identifying patients at risk of heart disease. The system makes use of the well-known Cleveland Heart Disease dataset, which is available on the UCI machine learning repository/Kaggle, and employs ANNs to predict the likelihood of heart disease. ANNs can learn from large datasets. An ANN also called as neural network is a mathematical model based on biological neural networks. Artificial neural network is based on observation of a human brain. Human brain is very complicated web of neurons. Neuron has axons, dendrites and synapses. The designed ANN has three layers: namely an input layer, a hidden layer and an output layer. The proposed system is developed using Flask web framework where the trained ANN model is deployed as a web application so that any user or healthcare professional, patients can be accessed to predict the likelihood of heart disease..

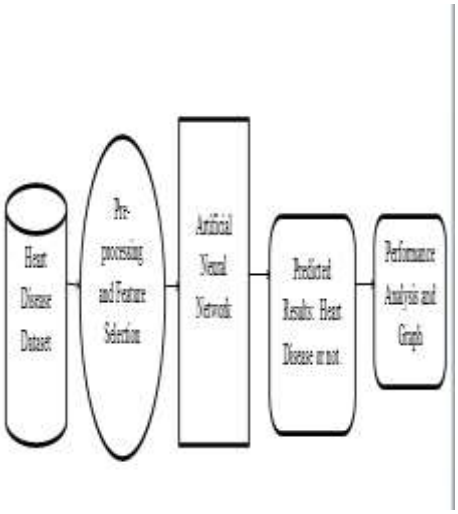


Figure 1: Overview of the proposed system.

3.1 IMPLEMENTATION

3.1.1 Data Collection:

We create the data gathering procedure in the first module of the Heart Disease Prediction utilizing Machine Learning. Gathering data is the first significant stage in the actual creation of a machine learning model. This is a crucial stage that will have a cascading effect on the model's quality; the more and better data we collect, the more capable our model will be.

There are various methods for gathering the data, including manual interventions and online scraping. The model folder contains the dataset.

3.1.2 Splitting the dataset:

This module involves splitting the preprocessed dataset into training and testing sets for model development and evaluation. Split the dataset into train and test. 80% train data and 20% test data.

3.1.3 Neural network:

This module involves developing the ANN model with multiple hidden layers, activation functions, and optimization

algorithms. For classification problems, it is composed of:

- A scaling layer.
- A perceptron layer.
- A probabilistic layer.

For the scaling layer, the base and most extreme scaling techniques are set.

We set one perceptron layer, with 3 neurons as a first conjecture, having the calculated enactment capability.

Brain Organization is a progression of calculations that are attempting to mirror the human mind and track down the connection between the arrangements of information. It is being utilized in different use-cases like in relapse,arrangement, Picture Acknowledgment and some more.

As we have talked over that brain networks attempts to emulate the human mind then there may be the distinction as well as the likeness between them. Allow us to discuss it.

A few significant contrasts between them are organic brain network matches handling though the Counterfeit brain network does series handling likewise in the previous one handling is more slow (in millisecond) while in the last one handling is quicker (in a nanosecond).

3.1.4 Design of ANN

This module includes planning the design of the ANN model, including the quantity of info and result neurons, the quantity of secret layers, and the initiation capabilities. A brain network has many layers and each layer carries out a particular role, and as the intricacy of the model expands, the quantity of layers likewise builds that why it is known as the multi-facet perceptron.

The most flawless type of a brain network has three layers input layer, the secret layer, and the result layer. The information layer gets the information signals and moves them to the following layer lastly, the result layer gives the last expectation and these brain networks must be prepared with some preparation information too like AI calculations prior to giving a specific issue. Presently, we should see more about perceptron.

3.1.5 Apply the model and plot the graphs for accuracy and loss:

This module involves applying the selected model to predict the likelihood of heart disease in new patients and plotting the graphs for accuracy and loss during training. We will compile the model and apply it using fit function. The batch size will be 25. Then we will plot the graphs for accuracy and loss. We got average validation accuracy of 92% and average training accuracy of 85%.

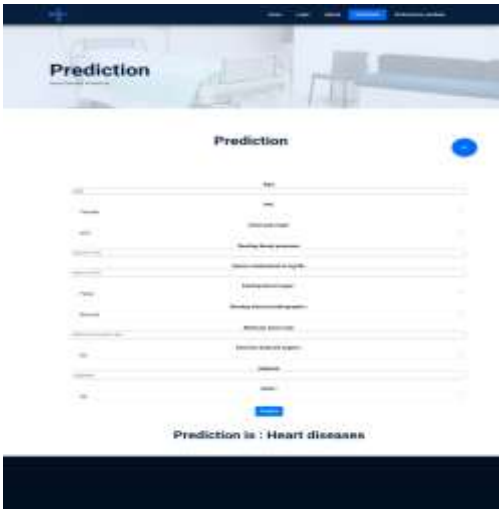
3.1.6 Analyze and Prediction:

This module involves analyzing the results of the ANN model and making predictions about the likelihood of heart disease in patients. In the actual dataset, we chose only 11 features:

1. *Age*
2. *Sex*
3. *Chest-pain type(*cp*)*
4. *Resting Blood Pressure(*trestbps*)*
5. *Serum Cholestrol(*chol*)*
6. *Fasting Blood Sugar(*fbs*)*

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

7. *Resting ECG (*restecg*)*
8. *Max heart rate achieved*.
9. *Exercise induced angina* :
10. *ST depression induced by exercise relative to rest*
11. *Peak exercise ST segment*



3.1.7 Accuracy on test set:

After training and evaluating the model on the validation set, the accuracy of the model will be assessed on the test set. The accuracy on the test set will be an important metric for evaluating the model's performance. We got an accuracy of 85% on test set.

3.1.8 Saving the Trained Model:

This module involves saving the trained ANN model for future use, such as deploying the model as a web application that can be accessed by healthcare professionals and patients to predict the likelihood of heart disease.

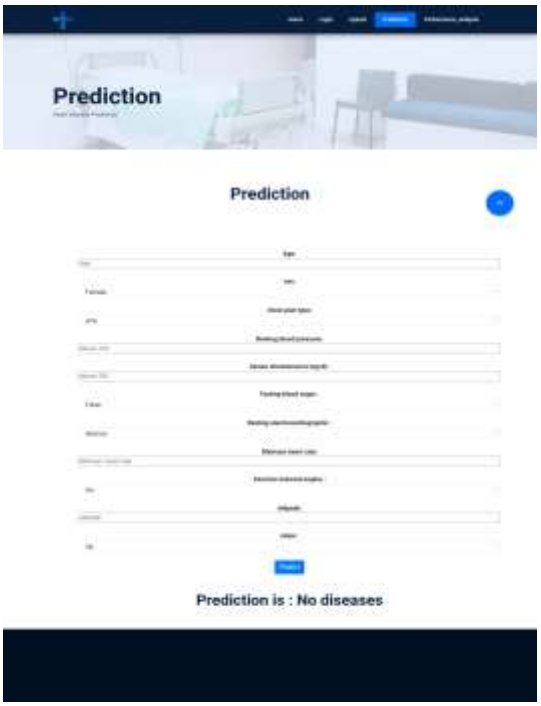
Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle .

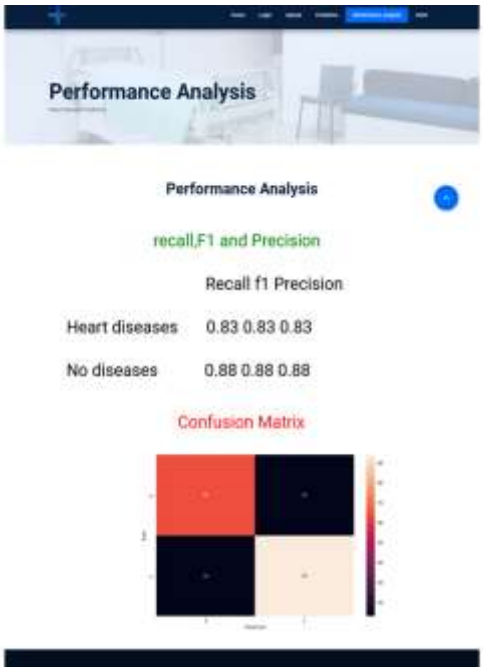
Make sure you have pickle installed in your environment.

Next, let's import the module and dump the model into .pkl file.

4.RESULTS AND DISCUSSION

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024





5. CONCLUSION

Finally, the study "Heart Disease Prediction using Artificial Neural Network (ANN)" demonstrates the feasibility of utilizing machine learning techniques to effectively forecast the likelihood of heart disease in patients. The project entailed gathering a big dataset of patient information, preprocessing it, and creating an artificial neural network (ANN) model that can learn from the data and accurately forecast the possibility of heart disease. The suggested approach has various advantages over current methods, including early identification of cardiac illness, high accuracy, scalability, and customizability. The system may be continuously enhanced by adding new data and changing the model as new information becomes available.

The project included numerous modules, such as data collecting, dataset preparation, library importation, dataset splitting, ANN model development, model selection, model application, and outcomes analysis. The developed model attained great accuracy and demonstrated its ability to effectively forecast the likelihood of heart disease in patients. Overall, the experiment demonstrates the promise of machine learning approaches, notably ANN, for reliably predicting the likelihood of heart disease in patients. This method has the potential to assist healthcare practitioners in early detection and treatment of cardiac disease, resulting in improved patient outcomes and better management of healthcare resources.

FUTURE ENHANCEMENT:

Incorporating additional features: The current model uses a set of features to predict the likelihood of heart disease. However, new features may become available that could improve the accuracy of the model. For example, incorporating genetic data or wearable technology data

could provide more comprehensive information for predicting the likelihood of heart disease.

Using more advanced ANN models: The current model uses a basic ANN model with multiple hidden layers. More advanced ANN models such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs) could be used to improve the accuracy of the model.

Incorporating more data: The current model uses a relatively small dataset. Incorporating larger and more diverse datasets could improve the accuracy and robustness of the model.

Integrating with Electronic Health Records (EHRs): Integrating the model with EHRs could enable healthcare professionals to quickly and easily access patient data and predict the likelihood of heart disease without having to manually enter data into the model.

Testing on different populations: The current model was trained and tested on a specific population. Testing the model on different populations could provide valuable insights into its generalizability and applicability across different patient populations.

REFERENCES

[1] Ayatollahi, H., Gholamhosseini, L., & Salehi, M. (2019). "Predicting coronary artery disease: a comparison between two data mining algorithms."

[2] BMC Public Health. doi: 10.1186/S12889-019-6721-5. Benjamin, H., David, F., & Belcy, S. A. (2018). "Heart disease prediction using data mining techniques." ICTACT Journal of Soft Computing, 9(1), 1824-1830

[3] Dengqing Zhang, Yunyi Chen, Yuxuan Chen, Shengyi Ye, Wenyu Cai, Junxue Jiang, Yechuan Xu, Gongfeng Zheng, Ming Chen, "Heart Disease Prediction Based on the Embedded Feature Selection Method and Deep Neural Network", Journal of Healthcare Engineering, vol. 2021, Article ID 6260022, 9 pages, 2021.

[4] Dulhare, U. N. (2018). "Prediction system for heart disease using naïve bayes and particle swarm optimisation." Biomedical Research, 29 (12), 2646-2649.

[5] Gawali, M., & Shirwalkar, N. (2018). "Heart disease prediction system using data mining techniques." International Journal of Pure and Applied mathematics, 120 (6), 499-506.

[6] Haq, A. U., Li, J.-P., Memon, M. H., Nazir, S., & Sun, R. (2018). "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." Hindawi Mobile Information System. Doi: 10.1155/2018/3860146.

[7] Khan, S. N., Nawi, N. M., Shahzad, A., Ullah, A., & Mushtaq, M. F. (2019). "Comparative analysis for heart disease

prediction." International Journal on Informatics Visualization, 1 (4-2), 227-231.

[8] Nashif, S., Raiban, M., Islam, M., & Imam, M. H. (2018). "Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system." World Journal of Engineering and Technology, 6, 854-873.

[9] Prasead, R., Anjali, P., Adil, S., & Deepa, N. (2019). "Heart disease prediction using logistic regression algorithm using machine learning." International Journel of Engineering and Advanced Technology, 8 (3S), 659-662.

[10] Salhi, Dhai Eddine & Tari, Abdelkamel & Kechadi, MTahar. (2021). "Using Machine Learning for Heart Disease Prediction." 10.1007/978-3-030-69418-0_7.

AUTHER PROFILE

1]Mr.B.suresh reddy currently working as an Assistant Professor in the Department of Master of Computer Applications, QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He published two text books, two book chapters, more than 19 research papers in reputed peer reviewed Scopus indexed journals. He also attended and presented research papers in different national and international journals. His area of interest is Information Security, Computer Networks, IoT, Data Science & Machine Learning.

2]Mr.Venkata Nagaraju Galam, currently pursuing Master of Computer Applications at QIS College of Engineering and Technology(Autonomous),Ongole, Andhra Pradesh. he completed BSC from Sri Gowthami Degree & pg College , darsi, Prakasam, Andhra Pradesh. His areas of interests are Machine Learning.