HEART DISEASE PREDICTION USING MACHINE LEARNING

S. M. Sasikala, AP/CSE Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.K.V.Sridhar Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.S.M.Sriram Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.D.Tamilarasan Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.J.S.Vittal Computer Science and Engineering & Dhirajlal Gandhi College of Technology

---

Abstract - Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. The traditional approach to predicting heart disease involves a comprehensive evaluation of multiple risk factors such as age, gender, smoking status, blood pressure, cholesterol levels, and family history. However, recent advances in machine learning have enabled healthcare professionals to develop more accurate and efficient models for heart disease prediction using large datasets and advanced algorithms. This prediction demonstrated several classification mechanisms to build the prediction model. The data was collected and cleaned from any missing values and extreme outliers. The results show that SVM and random forest models are highly accurate and effective in identifying patients who are at risk of heart disease. Further many unprocessed machine learning algorithm techniques will be processed with the dataset by firstly collecting the dataset and cleaning it by training and testing the data, and the missing values are removed. Once the model is trained and validated, deploy it in a web application that can be used by doctors or patients to predict the risk of heart disease based on their medical information.

***Key Words*:** Heart disease prediction, Machine learning, Support vector machine, Multilayer perceptron, Naïve bayes, Random forest.

1.INTRODUCTION

Machine learning is a subset of synthetic intelligence that involves education pc systems to research from facts and improve their performance on a particular task. Machine studying includes various strategies consisting of supervised getting to know, unsupervised gaining knowledge of, and reinforcement getting to know. In supervised learning, the set of rules is trained the usage of labeled information, and the intention is to are expecting an output for brand spanking new, unseen input records correctly. In unsupervised learning, the algorithm is trained using unlabeled statistics, and the aim is to find out styles and shape in the records. Reinforcement learning includes schooling an agent to engage with an environment to obtain a specific purpose thru trial and blunders. The assignment makes use of numerous supervised system gaining knowledge of algorithms and has been applied as an internet software the usage of Flask. Cardiovascular Disease (CVD) is a group of conditions that affect the coronary heart and can result in critical fitness complications, including strokes and heart attacks. Risk factors for CVD encompass an bad weight-reduction plan, a sedentary way of life, tobacco use, and immoderate alcohol intake. These can cause improved blood strain, high blood glucose, high blood lipids, and weight problems, that may serve as early warning signs and symptoms of CVD. Symptoms such as shortness of breath, coughing, swelling of the ankles and feet, fatigue, loss of urge for food, and impaired wondering also can suggest the onset of CVD. The COVID-19 virus also can boom the danger of growing heart sickness. Early prognosis is essential for effective remedy to save you in addition deterioration of the coronary heart's health. Machine getting to know models have shown promise in predicting the probability of developing CVD primarily based on diverse danger elements consisting of age, sex, circle of relatives history, blood pressure, and cholesterol levels. Predictive models can assist identify high-threat people and enforce preventive measures, including lifestyle changes and medical interventions. Ultimately, heart disease prediction aims to lessen the worldwide burden of CVD and promote healthier existence to prevent coronary heart sickness.

2.LITRETURE SURVEY

Heart disease prediction was addressed in the literature using several methods. In [7], Naïve Bayes, SVM, and Functional Trees were used to predict the possibility of heart diseases with an accuracy of 84.5%, using measurements from wearable mobile technologies with the same inputs used in our work. Furthermore, Naïve Bayes was solely used in [8] with a slightly better accuracy of 86.4%, using the same dataset.

Another work have included several algorithms[9] such as Logistic Regression, KNN, NN, SVM, NB, Decision Tree, and RF, with three feature selection algorithms: Relief, mRMR, and LASSO to predict the existence of heart disease. The Logistic Regression a algorithm had the best performance and yielded predictions with an accuracy as high as 89%.

Moreover, a work done in 2020 [10] applied 4 algorithms with a very high accuracy of 90.8% for the KNN model, and minimum accuracy of 80.3% for the other models.

A work done in 2021 [16], Naïve bayes, Random Forest, and SVM, MLP were used for the prediction. All the previous is very promising for the future of heart diseases and failure prediction, especially with the current advances in portable electronic measurement devices.

3.SYSTEM IMPLEMENTATION:

EXISTING SYSTEM:

A model have been developed to predict the existence of heart disease in patients based on specific health measurements using machine learning algorithms. A predictive model was developed using four different classification mechanisms: Multilayer Perceptron (MLP), Support Vector Machines (SVM) with linear kernel, Naive Bayes (NB), and

Random Forest (RF). The performance of each algorithm was evaluated using a pre-processed dataset that was cleaned of missing values and outliers to ensure accuracy. The dataset underwent various phases, including visualizing the imbalances, obtaining the correlation matrix, and employing dimensionality reduction techniques before being split using Hold-out to train and test the model for each machine learning algorithm. The algorithms used were able to extract complex relations between the symptoms and the disease. The system demonstrated the potential of machine learning algorithms to be applied to other types of diseases. The model has been trained for the dataset and not for the real time data and other data. If crossvalidation was performed, it would be helpful to know the average performance across different folds to get a better sense of how well the models generalize to new data.

PROPOSED SYSTEM:

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

COLLECTION OF DATASET:

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model.

SELECTION OF ATTRIBUTES:

Attribute or feature selection is an important step in building a machine learning model as it involves identifying the most relevant and informative attributes that can be used to predict the target variable. In the case of a heart disease prediction model, various attributes such as gender, chest pain type, fasting blood pressure, serum cholesterol, are commonly used as input features.

PRE-PROCESSING OF DATA:

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model.

MODEL SELECTION:

Select a suitable machine learning algorithm to train the model, such as logistic regression, decision tree, or random forest. Compare the performance of different algorithms and choose the one that provides the best results. It involves evaluating different models based on their performance metrics and selecting the best one for the specific task at hand. Validate the model using the testing data and tune the hyperparameters to improve the model's accuracy.

DISEASE PREDICTION:

Various machine learning algorithms like SVM, Naive Bayes, Decision Tree, Random Tree, Logistic Regression, MLP are used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction. Once the model is trained and validated, deploy it in a web application or mobile app that can be used by doctors or patients to predict the risk of heart disease based on their medical information.

4.SYSTEM ARCHITECTURE

An architecture for heart disease prediction using machine learning was developed. The first step involved collecting data on risk factors for heart disease, such as age, sex, blood pressure, and cholesterol levels. The data was preprocessed to remove missing values and normalize the features and it was splitted into two models as training model and the testing model. Next, a suitable machine learning algorithm was selected, such as logistic regression or decision tree. The algorithm was trained on the preprocessed data to learn patterns and relationships between risk factors and heart disease. The trained model was then evaluated on a test dataset to assess its accuracy and performance. Finally, the model was deployed in a clinical setting, where it could be used to predict the likelihood of heart disease in patients based on their risk factors. The architecture showed promising results and had the potential to assist in early diagnosis and prevention of heart disease.

Fig -1: System Architecture

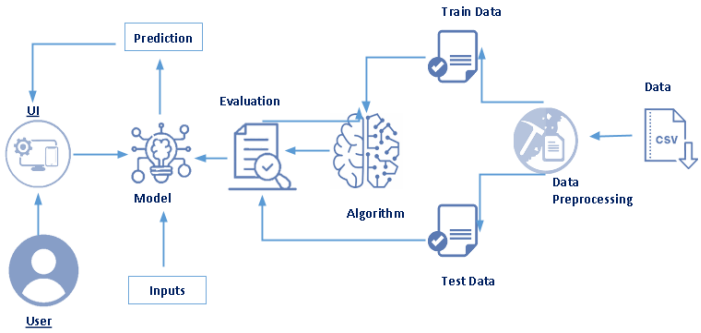

Table -1: Model Accuracy after testing

MODEL	ACCURACY
SVM	81.9%
Naïve Bayes	81.9%
Random Forest	75.4%
Decision Tree	78.6%
Logistic Regression	81.9%
MLP	70.4%

5. CONCLUSION

The heart disease prediction project is a practical and useful application of machine learning algorithms that can assist medical professionals in diagnosing heart disease. The project uses several supervised machine learning algorithms and has been implemented as a web application using Flask. The project has several advantages, including the ability to save time and reduce the risk of errors, and is easily scalable. However, the project also has limitations, including the use of a small dataset and simple models, which could be improved to enhance the accuracy of the predictions. This prediction demonstrated several classification mechanism to build the prediction model. The data was collected and cleaned from any missing values and extreme outliers. The results show that SVM and Naïve bayes models are highly accurate and effective with the accuracy of 82% in identifying patients who are at risk of heart disease. Finally the model is deployed on Microsoft Azure Cloud platform. The future work can enhanced in the form of Improving the model's accuracy, Adding more features.

REFERENCES

1. S. Rehman, E. Rehman, M. Ikram, and Z. Jianglin, “Cardiovascular disease (CVD): assessment, prediction and policy implications,” BMC Public Health, vol. 21, no. 1, p. 1299, 2021, doi: 10.1186/s12889-021-11334-2.
2. O. Atef, A. B. Nassif, M. A. Talib, and Q. Nassir, “Death/Recovery Prediction for Covid-19 Patients using Machine Learning,” 2020.
3. A. B. Nassif, I. Shahin, M. Bader, A. Hassan, and N. Werghi, “COVID-19 Detection Systems Using Deep-Learning Algorithms Based on Speech and Image Data,” Mathematics, 2022.
4. H. Hijazi, M. Abu Talib, A. Hasasneh, A. Bou Nassif, N. Ahmed, and Q. Nasir, “Wearable Devices, Smartphones, and Interpretable Artificial Intelligence in Combating COVID-19,” Sensors, vol. 21, no. 24, 2021, doi: 10.3390/s21248424.
5. O. T. Ali, A. B. Nassif, and L. F. Capretz, “Business intelligence solutions in healthcare a case study: Transforming OLTP system to BI solution,” in 2013 3rd International Conference on Communications and Information Technology, ICCIT 2013, 2013, pp. 209–214, doi: 10.1109/ICCITechnology.2013.6579551.
6. A. Nassif, O. Mahdi, Q. Nasir, M. Abu Talib, and M. Azzeh, “Machine Learning Classifications of Coronary Artery Disease.” Jan. 2018. [7] A. F. Otoom, E. E. Abdallah, Y. Kilani, A. Kefaye, and M. Ashour, “Effective diagnosis and monitoring of heart disease,” Int. J. Softw. Eng. its Appl., vol. 9, no. 1, pp. 143–156, 2015, doi: 10.14257/IJSEIA.2015.9.1.12.
7. A. F. Otoom, E. E. Abdallah, Y. Kilani, A. Kefaye, and M. Ashour, “Effective diagnosis and monitoring of heart disease,” Int. J. Softw. Eng. its Appl., vol. 9, no. 1, pp. 143–156, 2015, doi: 10.14257/IJSEIA.2015.9.1.12.
8. K. Vembandasamyp, R. R. Sasipriyap, and E. Deepap, “Heart Diseases Detection Using Naive Bayes Algorithm,” IJISETInternational J. Innov. Sci. Eng. Technol., vol. 2, no. 9, 2015, Accessed: Dec. 11, 2021. [Online]. Available: www.ijiset.com.
9. A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, R. Sun, and I. GarciáMagarinõ, “A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms,” Mob. Inf. Syst., vol. 2018, 2018, doi: 10.1155/2018/3860146.
10. D. Shah, S. Patel, · Santosh, and K. Bharti, “Heart Disease Prediction using Machine Learning Techniques,” vol. 1, p. 345, 2020, doi: 10.1007/s42979-020-00365-y.
11. K. Pahwa and R. Kumar, “Prediction of heart disease using hybrid technique for selecting features,” 2017 4th IEEE Uttar Pradesh Sect. Int. Conf. Electr. Comput. Electron. UPCON 2017, vol. 2018-January, pp. 500–504, Jun. 2017, doi: 10.1109/UPCON.2017.8251100.
12. H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, “Heart disease prediction using machine learning algorithms,” doi: 10.1088/1757- 899X/1022/1/012072.
13. “Heart Disease UCI | Kaggle.” https://www.kaggle.com/ronitf/heart-disease-uci (accessed Jan. 10, 2022).
14. D. Murphy, “Using Random Forest Machine Learning Methods to Identify Spatiotemporal Patterns of Cheatgrass Invasion through Landsat Land Cover Classification in the Great Basin from 1984 - 2011,” 2019.
15. S. Liu, Z. Fang, and L. Zhang, “Research on Urban Short-term Traffic Flow Forecasting Model,” J. Phys. Conf. Ser., vol. 1237, no. 5, Jul. 2019, doi: 10.1088/1742-6596/1237/5/052026.
16. Boukhatem, C., Youssef, H. Y., & Bou Nassif, A. (2021). Heart disease prediction using machine learning. Proceedings of the 2021 International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2021), Sharjah, UAE, 8-10 March 2022, pp. 1-6. DOI: 10.1145/3456257.3456285.