

PREDICTING CANCER AND HEART DISEASE AT EARLY STAGES USING MACHINE LEARNING

Arjun C , Anusha , Meghana B Kamath , Nireeksha S Amin,
Under graduate Scholars, Dept of CSE,
Mangalore Institute of Technology & Engineering, Moodabidri, INDIA
Dr.Pradeep B S
Professor, Dept of CSE,
Mangalore Institute of Technology & Engineering, Moodabidri, INDIA
pradeepbs@mite.ac.in



CrossMark



Publication History

Manuscript Reference No: IJIRAE/RS/Vol.09/Issue05/MYAE10081

Research Article | Open Access

Peer-review: Double-blind Peer-reviewed

Article ID: IJIRAE/RS/Vol.09/Issue05/MYAE10081

Received Date: 04, May 2022 | Accepted Date: 14, May 2022 | Available Online: 31, May 2022

Volume 2022 | Article ID MYAE10081 <http://www.ijirae.com/volumes/Vol9/iss-05/01.MYAE10081.pdf>

Article Citation: Arjun, Anusha, Meghana, Nireeksha & Dr.Pradeep (2022). PREDICTING CANCER AND HEART DISEASE AT EARLY STAGES USING MACHINE LEARNING. International Journal of Innovative Research in Advanced Engineering, IX, 105-109. <https://doi.org/10.26562/ijirae.2022.v0905.001>

BibTeX key [arjun2022predicting](#)

Academic Editor-Chief: Dr.A.Arul Lawrence Selvakumar

Copyright: ©2022 This is an open access article distributed under the terms of the **Creative Commons Attribution License**; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: People today suffer from a variety of diseases as a result of their living habits and their surroundings. As a result, predicting sickness at an early stage becomes a critical responsibility. Machine learning has been demonstrated to be useful in assisting with decision-making and prediction from enormous amounts of data generated by the health sector. As a result, machine learning's use in medical diagnostics is rapidly rising. Heart disease is one of the leading causes of death in the modern world. Predicting and detecting cardiac disease has always been a difficult issue for medical professionals. To treat cardiac disorders, hospitals and other clinics are offering costly medicines and operations. As a result, anticipating cardiac disease in its early stages will be beneficial to people all over the world, allowing them to take the essential steps before it becomes serious. Cancer has been described as a diverse illness with numerous subgroups. Early detection and prognosis of a cancer type has become a necessity in cancer research since it can help with patient clinical treatment. The health industry nowadays plays a key part in curing patient's problems, thus this is also a way for the health sector to enlighten users. It is useful for the user if he or she does not want to go to the hospital or any other clinic, because by simply entering the symptoms and other relevant information, the user can learn about the disease they are suffering from. The health industry can also benefit from this system by simply asking the user for symptoms and entering them into the system, and in a matter of seconds, they can tell the exact and, to a degree, accurate diseases.

Keywords: Machine Learning, disease prediction, symptoms, heart disease, cancer

I. INTRODUCTION

Nowadays some of the biggest causes of death are heart disease and cancer. Such chronic diseases become the major healthcare issue all over the world. Therefore the analysis of chronic disease is more important to reduce the risk of people life. In the biomedical and healthcare communities, precise prediction is critical for determining the patient's disease. Early disease detection is one of the advantages of the prediction. Medicine and healthcare are two of the most important aspects of a human's life. For the prevention and treatment of illness, an accurate and timely analysis of any health-related issue is critical. In the case of a critical illness, the standard method of diagnosis may not be sufficient.

The development of a medical diagnosis system based on machine learning (ML) algorithms for disease prediction can aid in a more accurate diagnosis than the current method. Using numerous machine learning techniques, we developed a disease prediction system. A disease predictor, often known as a virtual doctor, may accurately forecast the sickness of any patient without the need for human intervention. Our diagnosis model can operate like a doctor in the early detection of a disease, allowing for timely treatment and the saving of lives. Chronic diseases are a major problem in the healthcare industry all around the world. According to the medical statement, the death rate of humans is increasing as a result of chronic disorders. The cost of treating this disease consumes more than 70% of the patient's income. As a result, it is necessary to minimize the patient's chance of dying.

II. RELATED WORK

Purushottam et al. [1] proposed an approach for efficient heart disease prediction. They developed a data mining model to predict heart disease efficiently. It mainly helps the medical practitioners to make efficient decisions way based on the given parameters. The author has used Cleveland dataset from UCI, and they have used age, sex, resting blood pressure, chest pain, serum cholesterol, fasting blood sugar, etc. as attributes. Furthermore, they have divided the datasets into two parts one is for testing, and the other one is for training. They have used a 10-fold method to find accuracy. Deepika [2] proposed an approach for predictive analytics to prevent and control chronic diseases. The proposed predictive analytics to prevent and control the chronic disease with the help of machine learning techniques such as naive Bayes, support vector machine, decision tree, and artificial neural network and they have used UCI machine learning repository datasets to calculate the accuracy. Among them, Support vector machine gives the best accuracy of 95.55%. Moataz M. Abdelwahab et al. [3] proposed a Four Layers Image Representation for Prediction of Lung Cancer Genetic Mutations Based on 2DPCA. They predicted that Lung cancer is among the most common genetic diseases that thousands of people die from each year. It develops in human body due to damage of normal genes caused by many reasons such as smoking cigarette, which chemically activate the oncogenes and deactivate the tumor suppressor the normal lung cell and produce mutations that result in tumors. This creates a critical need for robust, fast, sensitive diagnostic tests. Therefore, it has become very important to develop a new genetic mutation-based technique for cancer prediction to win the battle against cancer.

Shanjida Khan Maliha et al. [4] focused on an approach where algorithms such as K-Nearest Neighbor, Naive Bayes, and J48 algorithm were used for the prediction of cancer disease. Medical datasets which help in the prediction of cancer are used in this paper. When large datasets are involved, Naive Bayes was the chosen one due to its simplicity in implementation. For predicting the categorization of new instances, the K-nearest neighbor is used by taking a dataset and dividing it into multiple groups. The decision Tree from training datasets is formed in case of the J48 Classifier, and it also makes use of the fact that the datasets are used for the smallest subset decision. The correctness of a cancer illness in the datasets which include nine other cancer types are assessed using the Weka tool. When the three algorithms are compared, all three are found to be effective, but it is to be noted that in case of precision K-Nearest Neighbour was found to be better than the rest. Senthilkumar Mohan et al. [5] proposed a novel method for efficient heart disease prediction by using hybrid machine learning techniques. By applying such techniques, it finds significant features and thus improves the accuracy in the prediction of cardiovascular disease. This prediction model was introduced with different combinations of features and several classification techniques. They produced an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease.

III. WORKING METHODOLOGY

The proposed idea is used for predicting the severe diseases like heart disease or cancer at an early stage. It's a web-based interface where the user has to provide the symptoms which the user is going through based on which we have several algorithms which predicts the disease and also displays the percentage of accuracy. The user needs to enter all the columns of symptoms to get the accurate result. If the user is diagnosed with the disease, then the user can interact with the chatbot and could also be suggested with the nearest doctor or a specialist whom the user can contact and can the carry on with the treatment.

A. Data pre-processing

The data is pre-processed after collection of various records. The datasets contain huge number of records and they are in the .csv format. The symptoms are used in the dataset and they help to take the prediction. The pre-processing is carried out by converting the medical records into diagnosis values. Once it has been completed, it is then subjected to feature extraction followed by disease prediction.

B. Feature selection and Reduction

From among the several attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining attributes are considered important as they contain vital clinical records are vital to diagnosis and learning the severity of the disease. The technique that we used for predicting heart disease is support vector machine and Naive Bayes for predicting cancer.

C. Classification modelling

1) Support vector machine (SVM):

A support vector machine is a supervised machine learning algorithm. It is utilized as a part of medical diagnosis for both classification as well as regression. If we give any labeled training data to support vector machine algorithm, it will produce a classifier which will divide the labelled data into different classes.

2) Naïve Bayes:

Naïve Bayes is one of the supervised machine learning classification algorithm based on the Bayes theorem. It deals with the dataset that have the highest dimensionality. It is based on the assumption that attributes are independent of each other which means attributes in one class that is independent of any other attributes that are present in the same class. The Naive Bayes calculation depends on the conditional probabilities. The main advantage of using the Naive Bayes classifier is that it requires a little measure of training data to assess the parameters vital for classification.

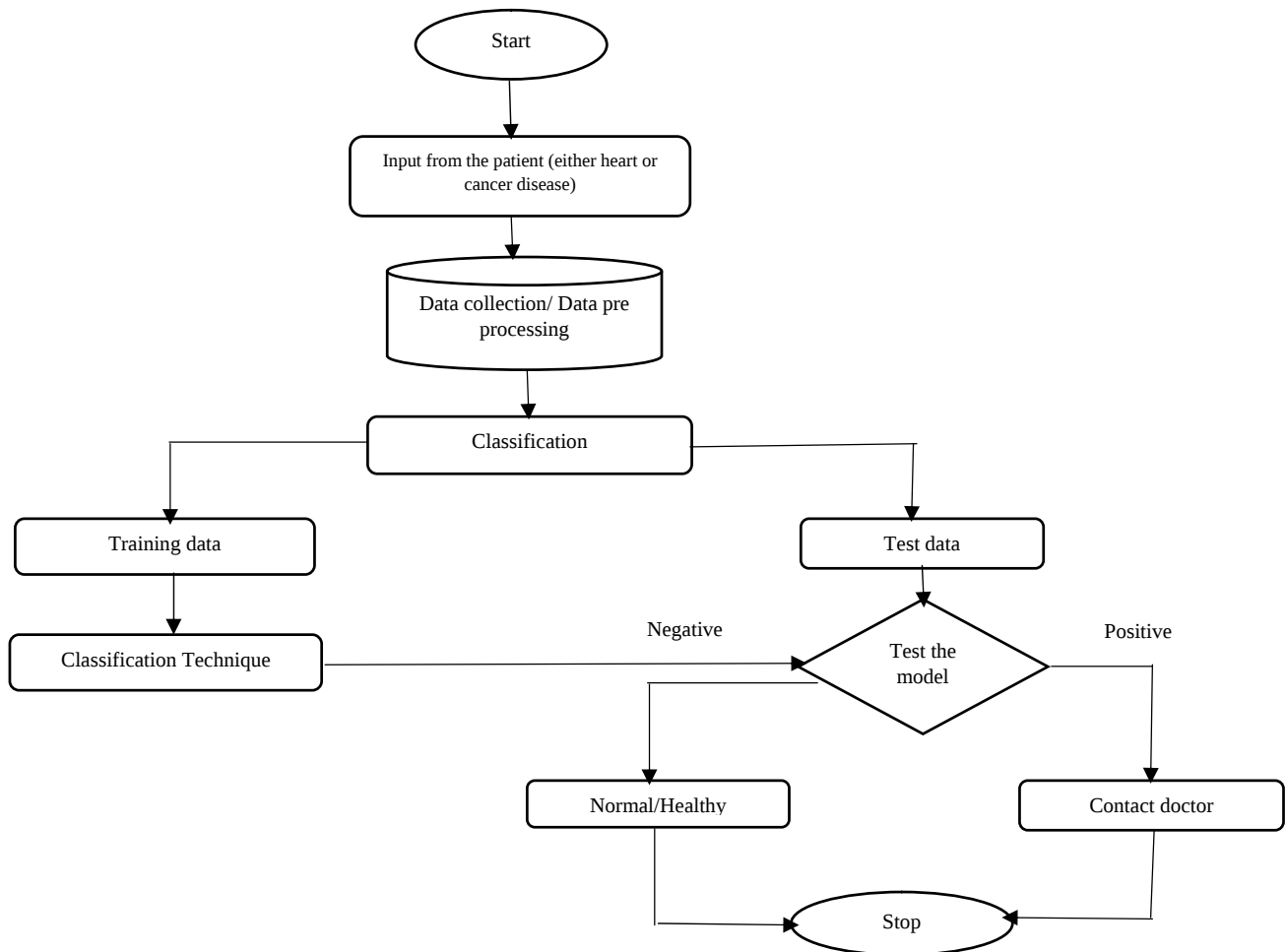


Figure 1: General model of predicting cancer and heart disease

IV. EXPECTED OUTCOMES

The system helps in predicting whether the patient is having the risk of heart disease or lung cancer according to the inputs provided. When found that the patient is suffering from the respective disease, the system also provides the patient with the information of the nearest doctors in his locality.

Some of the screenshots of the system so far developed are provided below:



Figure 2: Home Page



Heart Disease Prediction



Cancer Disease Prediction

Figure 3: Selection of Disease

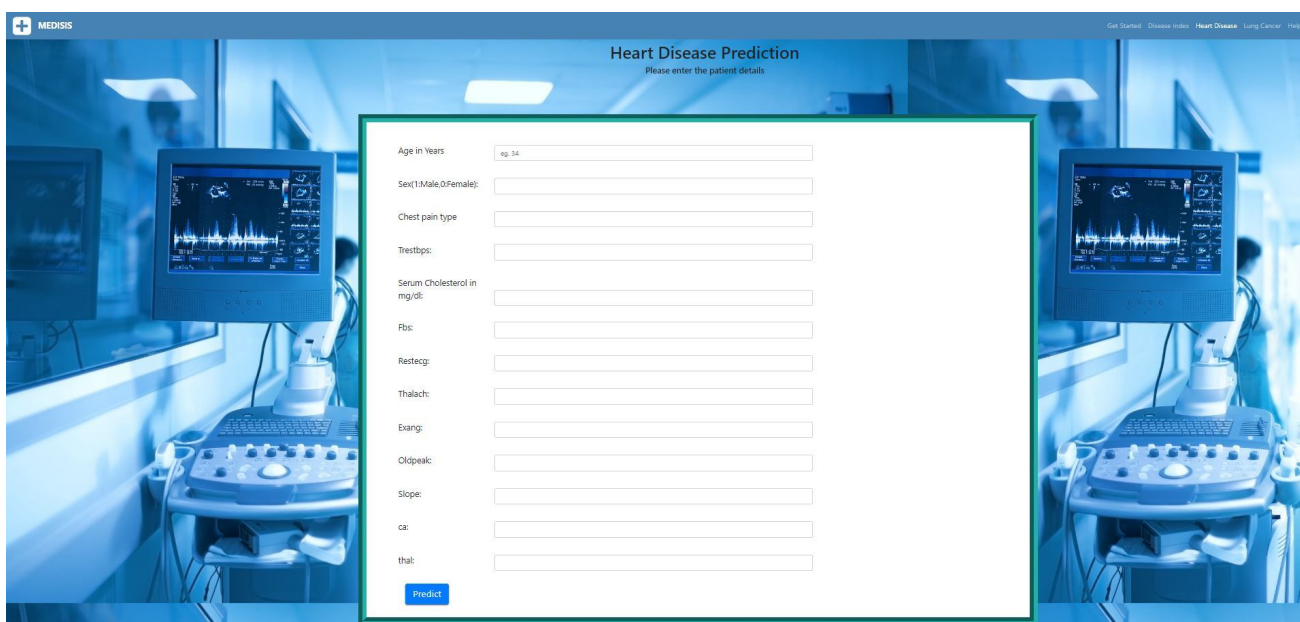


Figure 4: Taking inputs for Heart Disease Prediction

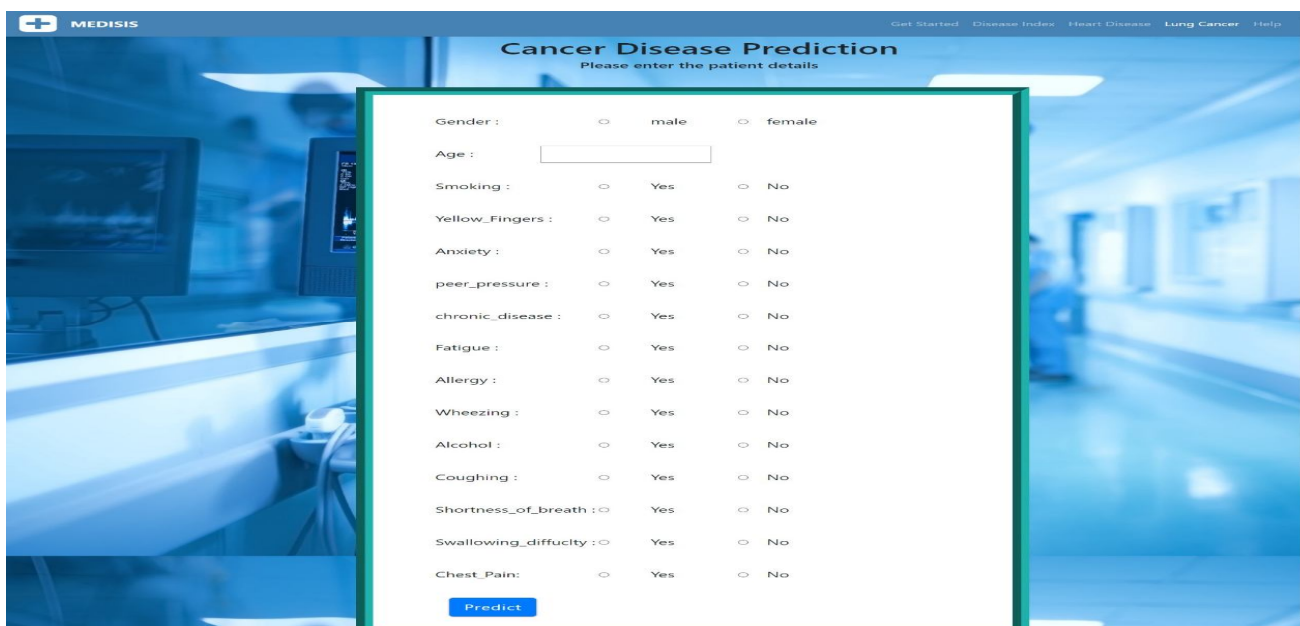


Figure 5: Taking inputs for Lung Cancer prediction

V. CONCLUSION

In today's rapidly changing world, cancer and heart disease are major concerns. As a result, an automated system to predict these at an early stage is required. So that it will be valuable to physicians because their workload will be decreased and they will be able to forecast the patient's sickness more quickly, and it will also be useful to the general public since they will be able to track their health issues at home using this automated system. The goal of the prediction system is to provide early warnings for heart disease and cancer, which if left untreated or disregarded in their early stages can be fatal and cause a slew of problems for the patient and their family.

REFERENCE

1. Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," *Procedia Computer. Sci.*, vol. 85, 2016. <https://doi.org/10.1016/j.procs.2016.05.288>
2. Deepika, "Predictive Analytics to Prevent and Control Chronic Diseases", 2016.
3. Moataz M. Abdelwahab, Abdel rahman, "Four Layers Image Representation for Prediction of Lung Cancer Genetic Mutations Based on 2DPCA", *IEEE* 2017. <https://doi.org/10.1109/mwscas.2017.8052994>
4. "Cancer Disease Prediction Using Naive Bayes, K-Nearest Neighbour and J48 algorithm", by IEEE access published in 2019. <https://ieeexplore.ieee.org/document/8944686>
5. "Effective Heart Disease Prediction Using Hybrid Machine learning Techniques", *IEEE Access*, published in 2019 <https://ieeexplore.ieee.org/document/8740989>
6. M. Gjoreski, A. Gradisek, B. Budna, M. Gams, and G. Pogljajen, "Machine Learning and End-to-End Deep Learning for the Detect ion of Chronic Heart Failure from Heart Sounds," *IEEEAccess*, vol. 8, 2020. <https://doi.org/10.1109/access.2020.2968900>
7. S. Divya, V. Indumathi, S. Ishwarya, M. Priyasankari, S. Kalpana Devi, "A Self-Diagnosis Medical Chatbot Using Artificial Intelligence", *J. Web Dev. Web Des.*, vol. 3, 2018.
8. Turki Turki, "An Empirical Study of Machine Learning Algorithms for Cancer Identification", *IEEE Access*, 2018.