Smart Detection of Diseases Using Machine Learning

Miss. Harika koormala,

Mtech, CSE branch,

Holy Mary Institute of technology and science, `E-mail.vharika.juluri@gmail.com

Dr. S. Srinivas,

Associate professor,

Holy Mary Institute of technology and science

ABSTRACT

A human being may get many various ailments in our planet. Diseases may affect a person physically as well as psychologically. In contemporary culture, it is the duty of medical experts to identify and classify the proper diseases, as well as to provide medical therapies or treatments to either cure or control the condition. While certain diseases may be cured with therapy, chronic illnesses are never cured; nonetheless, treatment can stop chronic illnesses from becoming worse over time. Therefore, it is crucial to identify and treat diseases at an early stage.

This project aims to provide a GUI for users to obtain a preliminary diagnosis or a second opinion regarding certain health issues, such as diabetes, breast cancer, heart disease, liver disease, and kidney disease. Users can be doctors, healthcare workers, or anyone else awaiting a diagnosis after testing. Based on the likely patient's body check-up & test data, we utilise machine learning models to identify these health concerns and indicate to them whether they may be at risk for a health condition.

Key Words: Diabetes, Heart, Liver, Knn, Random Forest,

1. INTRODUCTION

A human illness is an abnormal condition that adversely impacts the human body and may result in discomfort, agony, malfunction, and even death. Diseases may affect a person physically as well as psychologically. Additionally, there are several varieties of illnesses throughout the globe. The four basic causes of disease are (I) infection, (II) deficiency, (III) heredity, and (IV) malfunction of bodily organs.

Doctors and other medical experts in modern society are in charge of identifying and diagnosing the proper diseases as well as offering medical therapies or treatments to treat the diseases. Chronic conditions, however, are still present and might become worse with time. Therefore, it is

crucial to identify and treat diseases at an early stage. It is crucial that today's population be conscious of their health and does not exhibit either extreme sensitivity or indifference to it.

The healthcare industry now depends increasingly on computer technologies. Intelligent systems play a vital role in medical diagnostics. In order to detect irregularities at an early stage of illness diagnosis, machine learning techniques are applied. The best possible diagnosis is essential for choosing the right course of action.

A Disease Detection System (DDS) is presented that users, whether they be physicians, medical professionals, or regular people, may use to identify different illnesses in patients utilising the DDS's Graphical User Interface (GUI). DDS was created to find various ailments, including kidney disease, breast cancer, heart disease, and liver issues.

2. LITERATURE REVIEW

Priyanka Sonar, Prof. K. JayaMalini et.al: One of the most fatal illnesses in the world is diabetes. Additionally, it is the creator of a number of different types of illnesses, including cardiac failure, blindness, diseases of the urinary system, etc. In this situation, the patient must go to a diagnostic facility to get their findings after consultation. They have to devote their time and money each time because of this. However, with the development of machine learning techniques, we now have the freedom to look for a solution to the present problem. We now have sophisticated systems that use information processing that can predict if a patient has polygenic disease or not. Additionally, predicting the illness before it manifests is crucial for giving the patients with care. With information withdrawal, it is possible to exclude hidden information from a large collection of diabetes-related data. The objective of this investigation is to create a system that might more accurately estimate a patient's degree of diabetes risk. Decision Tree, ANN, I Bayes, and SVM algorithms are used as the foundation for classification approaches in model construction. The models provide precisions of 85% for Decision Tree, 77% for I Bayes, and 77.3% for Support Vector Machine. Results demonstrate a substantial level of procedure accuracy.

Samrat Kumar Dey, Ashraf Hossain and Md. Mahbubur Rahman et.al: Diabetes is brought on by an excessive concentration of sugar in the blood. It is now regarded as one of the most deadly illnesses in the world. Unknowingly or not, this serious sickness affects people all around the world. Diabetes may also lead to other conditions including heart attacks, paralysis, renal problems, blindness, etc. For detecting and analysing diabetes, several computer-based detection methods have been developed. Regular diabetes patient identification requires additional time and money. However, with the development of machine learning, we now have the capacity to

provide a solution to this serious problem. As a result, we have created an architecture that can predict if a patient has diabetes or not. Our primary goal in doing this investigation is to develop a web application based on certain potent machine learning algorithms that have greater prediction accuracy. We utilised the Pima Indian benchmark dataset, which uses diagnostic methods to predict the development of diabetes. Artificial Neural Network (ANN) predicts with an accuracy of 82.35%, a considerable gain in accuracy that motivates us to create an interactive web application for diabetes prediction.

Deeraj Shetty, Kishor Rit, Sohail Shaikh and Nikita Patil et.al: Software engineering has an area called data mining. It is a systematic process that involves methods at the intersection of artificial intelligence, machine learning, insights, and database systems to uncover instances in large data sets. The purpose of the data mining process is to analyse data from a data collection and transform it into an appropriate structure for further usage. Our research focuses on this area of medical conclusion learning design using the diabetic data that has been obtained in order to develop an intelligent therapeutic choice, emotionally supporting network to assist the doctors. The main goal of this investigation is to create an Intelligent Diabetes Disease Prediction System that analyses the diabetes illness using a database of diabetes patients. In this system, we suggest using algorithms like Bayesian and KNN (KNearest Neighbor) to a database of diabetes patients and analysing them while taking different diabetic features for diabetes illness prediction.

ZhilbertTafa, NerxhivanePervetica and BertranKarahoda et.al: With diabetes now affecting 346 million people worldwide, more than one-third of whom are undiagnosed in the early stages, there is an urgent need to enhance medical decision-making. Many studies have focused on either applying an algorithm or comparing the performances of algorithms on supplied, often preset and static datasets that are available online. In order to increase the reliability of computer-supported diagnosis, this work focuses on the combined implementation of support vector machine (SVM) and naive bayes statistical analysis in the dataset obtained from the medical exams of 402 patients. Some of the characteristics in the dataset have never been utilised in computer-based analyses before. The findings indicate that combining the two methods considerably increases the system's overall dependability, which is essential in the computer-supported diabetes diagnosis process.

3. SYSTEM ANALYSIS

3.1 Existing System

Disease detection is still in the research stage at the moment, and the majority of diseases diagnoses and forecasts are based on machine learning models without a graphical user interface,

which restricts the users to just tech-savvy individuals.

Additionally, the few illness detection GUIs that are accessible, such diabetes detection, are not particularly user-friendly. The GUI fails to clearly communicate what has to be entered, which causes uncertainty in the units or data and, sometimes, produces incorrect results.

There are several models that may be applied to each system, and each system employs distinct datasets and parameters.

3.2 Proposed System

A online tool called "Illness DETECTION USING MACHINE LEARNING" is being developed to aid in the early detection of disease by both people and medical professionals. It has an interactive GUI that offers several illness detection options in one location. A machine learning model is used in each illness detection module to get highly accurate findings. In this project, only text-based datasets are being utilised.

Scope of the Project

All operating systems are compatible with the designed system. The newest browsers that are installed on each machine are the extent of the analysis.

4. SYSTEM DESIGN

Architecture Diagram

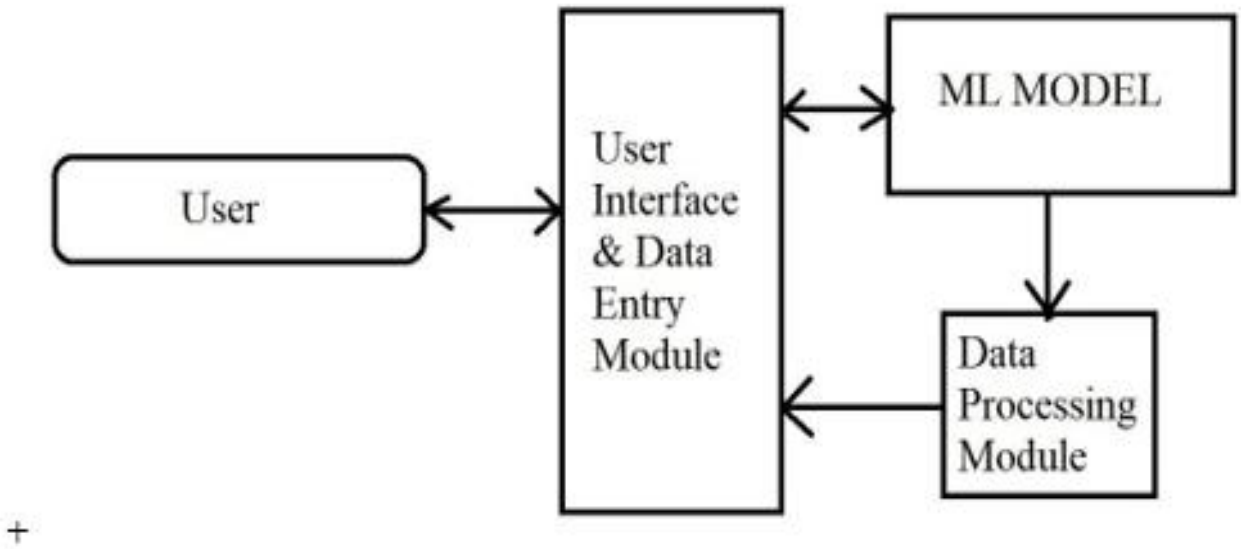

Fig1. Architecture of Disease Detector using Machine Learning Model

Use Case Diagram

Fig2. Use Case Diagram

Sequence Diagram

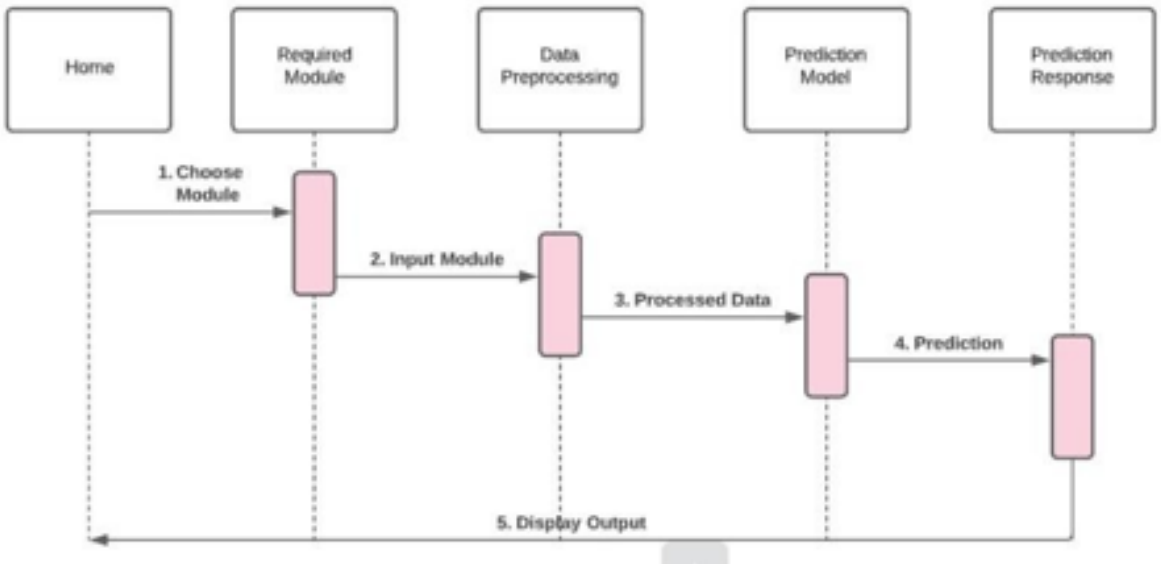

Fig3. Sequence Diagram

Class Diagram

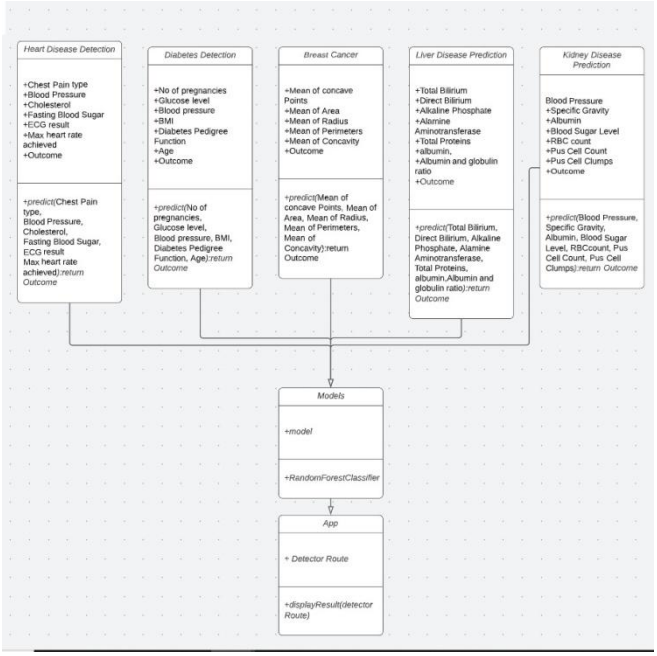

Fig4. Class Diagram

5. IMPLEMENTATION

Data Preparation:

Step 1: Import the required files

Step 2: Examine the data in the diabetes.csv and heart.csv files.

Step 3: Cleaning up the data, which entails reducing string values and substituting acceptable values for those that are lacking.

Step 4: Extracting from the data the appropriate labels for each dataset

Step 5: For each dataset, creating a csv file from the retrieved datagram.

Detection Models:

Step 1: Import the required packages

Step 2: Reading preprocessed data in,

Step3: then separating features and labels for training for each dataset.

Step 4: Decision tree algorithm training

Step 5: Creating a.pkl file to save the decision tree model.

Step-:6 making forecasts with the model

User Interface templates:

Step-1: Download bootstrap and collect Required images.

Step-2: Write the HTML & JavaScript Code for the UI.

6. RESULTS

Fig5. Home Page

Fig6. Heart Disease Detection Safe Input page

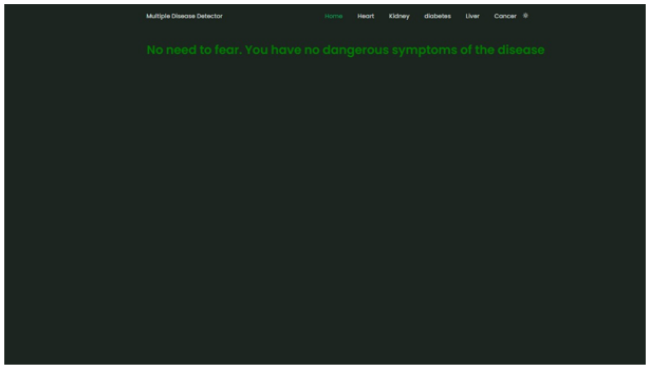

Fig7. Heart Disease Detection Safe Predicted Result

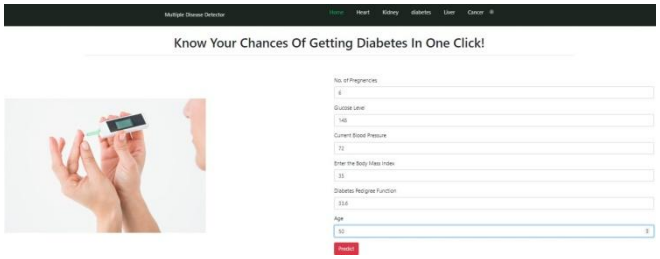

Fig8.Diabetes Detection Unsafe Input Page

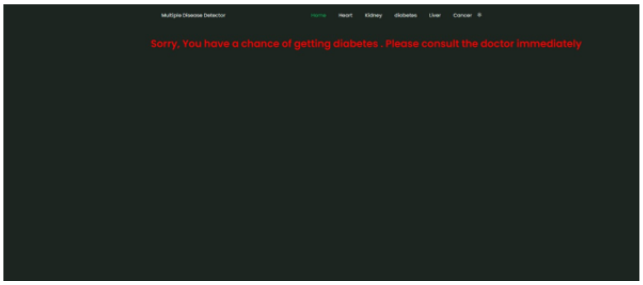

Fig9.Diabetes Detection Unsafe Predicted Result

Fig10. Breast Cancer Detection Safe Input page

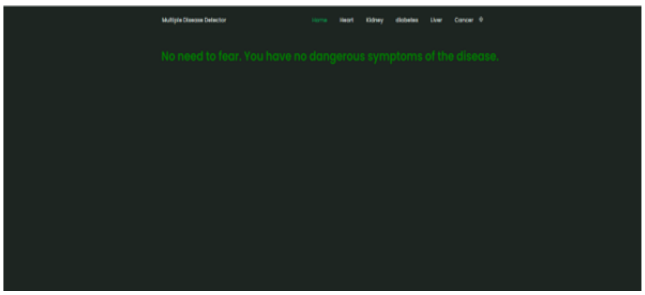

Fig11.Breast Cancer Detection Safe Predicted Result

Fig12. Breast Cancer Detection Unsafe Input page

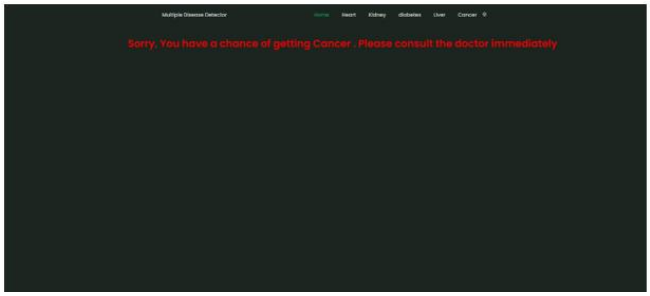

Fig 13. Breast Cancer Detection Unsafe Predicted Result

Fig 14. Liver Disease Detection Safe Input page

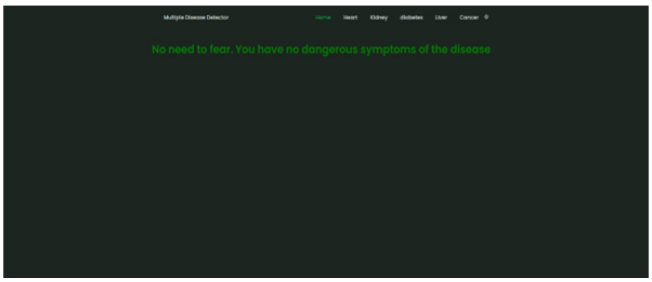

Fig 15. Liver Disease Detection Safe Predicted Result

Fig 16. Liver Disease Detection Unsafe Input page

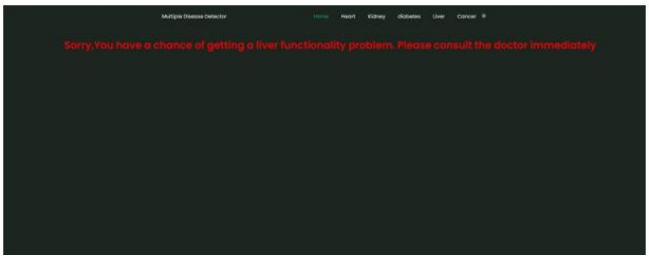

Fig 17. Liver Disease Detection Unsafe Predicted Result

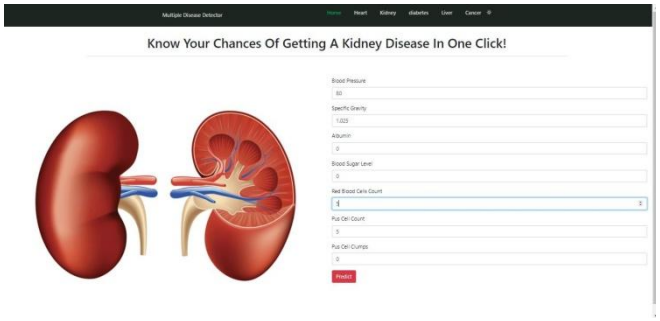

Fig 18. Kidney Disease Detection Safe Input page

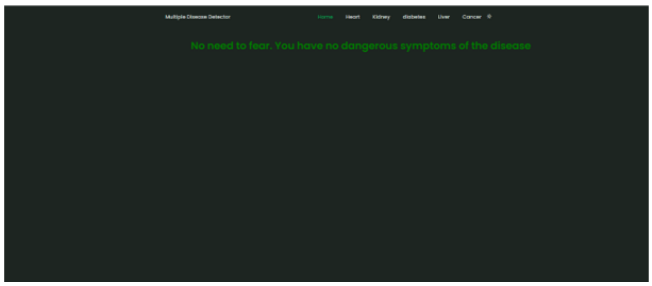

Fig 19. Kidney Disease Detection Safe Predicted Result

Fig 20 .Kidney Disease Detection Unsafe Input page

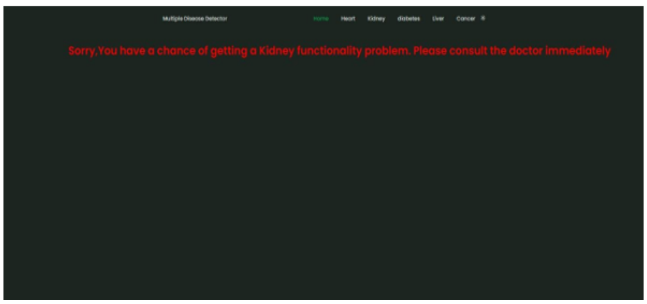

Fig 21. Kidney Disease Detection Unsafe Predicted Result

CONCLUSION

We created a web application that has all the characteristics and functions we had in mind and can identify a number of conditions, including liver problems, breast cancer, heart disease, diabetes, and renal disease.

The issue with the actual environment is fixed by this system. It accurately foretells if a patient will have a certain disease. The total system may address the issue of the absence of quick diagnostic techniques and provide a fresh tool for illness prediction in hospitals and for regular users. By doing a literature study as part of our domain and technical research, we were able to explore or investigate our subject. In our domain study, we identified key characteristics of each illness as well as a suitable method to provide high accuracy for the dataset that was made accessible.

However, the Kaggle datasets we have access to are study outcomes from patients throughout the world, and for Indian patients, we need additional indigenous research datasets.

FUTURE SCOPE

By including several other illness detections in one location and a feature to preserve the anticipated findings, this project may be further developed (reports). We may include a function like a live meeting with any of the available physicians if there is ever a case when we need to consult a doctor. This enables patients to communicate with the doctor more effectively.

A gateway for ordering prescription medications that individuals need to take for their health conditions may also be included. These medications may be purchased online by clicking one touch to navigate to the accessible drug online store if folks are at a point where they can't even get to a medical store. The web application as a whole is more effective thanks to this capability.

We may incorporate a CNN-based module for various skin conditions that will be able to identify the condition by uploading a photograph of your skin; the same can be done for hair conditions.

REFERENCES

[1] Soni J, Ansari U, Sharma D &Soni S (2011). Predictive data mining formedical diagnosis: an overview of heart disease prediction. InternationalJournal of Computer Applications, 17(8), 43-8

[2] Dangare C S &Apte S S (2012). Improved study of heart disease predictionsystem using data mining classification techniques. International Journal ofComputer Applications, 47(10), 44-8.

[3] Ordonez C (2006). Association rule discovery with the train and testapproach for heart disease prediction. IEEE Transactions on InformationTechnology in Biomedicine, 10(2), 334-43.

[4] Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heartdisease prediction system using k-means clustering and Naïve Bayesalgorithm. International Journal of Computer Science and InformationTechnologies, 6(1), 637-9.

[5] Bashir S, Qamar U &Javed M Y (2014,November). An ensemble-baseddecision support framework for intelligent heart disease diagnosis. InInternational Conference on Information Society(i-Society 2014) (pp. 259-64). IEEE. ICCRDA 2020 IOP Conf. Series: Materials Science andEngineering 1022 (2021) 012072 IOP Publishing doi:10.1088/1757-899X/1022/1/012072 9

[6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014).A coronary heart disease prediction model: the Korean Heart Study. BMJopen, 4(5), e005025.

[7] Ganna A, Magnusson P K, Pedersen N L, deFaire U, Reilly M, Ärnlöv J&Ingelsson E (2013). Multilocus genetic risk scores for coronary heart diseaseprediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267-72.

[8] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart diseaseprediction using lazy associative classification. In 2013 InternationalMutliConference on Automation,Computing,Communication, Control andCompressed Sensing (iMac4s) (pp. 40- 6). IEEE.