ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.7977132

www.ijirid.in

Prediction of Disease in Smart Healthcare System

Aditya Thakare, Karan Kankatav, Rishika Sadhwani, Swikruti Shende, Vinita Chandel

Undergraduate Students, Sipna College of Engineering and Technology, Amravati, Maharashtra, India

Abstract: *It may happen frequently for you or a loved one needs immediate medical attention, but they're not available due to unforeseen circumstances, or we can't find them. The best physician for your needs. We will therefore attempt to address this issue by integrating an online intelligent Smart Healthcare System into this project. Using a web-based programme, it enables patients to receive immediate guidance on their medical issues. The smart healthcare system's goal is to develop a web application that can analyze a user's symptoms to identify diseases and act as an online consultant for different conditions. We developed the Smart Health Care System, an expert system that facilitates the work of doctors. A machine performs a basic examination of the patient and suggests any potential diseases. It starts by asking the patient about their symptoms; if the device can identify the relevant condition, it then suggests a doctor who is close by. Based on the information that has been gathered, the system will display the outcome. Here, we're going to employ some cunning data mining strategies. We employ several forms of intelligent data mining. Ways to make the most accurate disease predictions based on a database of many patients' medical records, we use an algorithm (Naive Bayes) to map the symptoms with potential diseases that could be related to a patient's symptoms. This system helps patients by getting them the care they require as soon as possible while also making doctors' jobs easier.*

Keywords: *Disease Prediction, Naïve Bayes, Machine Learning Algorithm, Smart Healthcare System.*

I. INTRODUCTION

\Everyone who has a disease already must visit a doctor, which is both time-consuming and expensive. The user may also experience difficulties if they are unable to get in touch with a physician or a hospital because the illness cannot be identified. The process might be made simpler for the patient if the aforementioned procedure can be completed using automated software that saves time and money. Other systems for predicting heart disease analyze the patient's risk using data mining techniques. Using the user's symptoms, a web-based programme known as a smart healthcare system can predict an individual's illness. The intelligent healthcare system has collected data sets from various websites that deal with health. By using the symptoms provided, this method would allow the consumer to assess the likelihood of a disease.

Particularly given how popular the internet is becoming, people are constantly curious to learn new things. People frequently want to search for information online when a problem arises. Compared to the general public, hospitals and doctors have fewer online resources. The options available to people who are ill are limited. This device may therefore be useful. The advanced medical system is a project that offers online consultation and end-user assistance. This paper examines propose a system that enables users to access online health advice from a smart healthcare system. The system was fed a variety of ailments and symptoms. Users can communicate with the system about their symptoms and issues. It then examines the user's symptoms to determine whether any diseases may be connected to them.

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.7977132

www.ijirid.in

The algorithm (Naive Bayes) is used in this paper to map the symptoms to potential diseases and to determine the most likely suspected disease that could be connected to the patient's symptoms. This approach benefits patients by getting them the care they require as soon as possible while also making doctors' jobs easier. The paper is structured as follows; Section 2 emphasizes the fundamental goal, the function, and the algorithm used to create this system. Section 3 details the project's minimal hardware and software requirements and how they will be used. The proposed design relation between each class using ER is provided in Section 4. Section 5 discusses Diagrams and Conclusions.

II. LITERATURE REVIEW

The paper “Smart health prediction system using data mining” [1] the author has discussed many topics related to data mining techniques such as Naive Bayes, KDD (Knowledge discovery in Database). The Bayesian statistics can be applied to economic sociology and other fields. This checks the patients at initial level and automatically suggest the possible diseases. The system uses Naive Bayes classifier for the construction of the prediction system. The advantage of this system is that the initial consultation cost of doctor fees can be avoided. Eclipse IDE is used for creating the front-end Graphical User Interface and Navicat MySQL is used for backend database purpose. Here java is used as a programming language to connect the database and GUI purpose. The only disadvantage of the system the efficiency in detecting the symptoms or symptom mapping.

The paper “A Smart Health Prediction Using Data Mining” [2] is explaining the similar topics to the paper [1]. But there is detailed explanation of the internal algorithms used in the system. The Naive Bayes algorithm can be used for developing models that are used to assign class labels of different format. Naive Bayes algorithm is not a single, but a group of algorithms based on common principle. The steps involved in the Naive Bayes algorithm include.

I. Division of segments.
II. Comparing the first character of pattern until match occurs.
III. Comparing the last character of pattern.
IV. Perform each character comparison.

Also, the hardware requirements used are processor of 2.0 GHZ and Ram of 2GB. The software requirements are JAVA programming language, MySQL 5.0 database and Tomcat server.

In the paper “Smart E-Health Prediction System Using Data Mining” [3] most of the topics covered are on the system architecture. In this paper the design aspects of the system are primarily focused. In this paper the author has given a detailed framework to beat the downside of existing system. The smart health framework is used to implement the design aspects of the project. This framework asks for uses input and gather the symptoms to predict the disease based on data mining techniques. There are various modules such as login- used for authentication of patient and doctor, Diseases prediction, Doctor Searching, Feedback and Chatting with doctor for clearing patient doubts. There are some advantages such as finding the nearest doctor option to find doctor near to our location. These features can be used for better implementation of the system to help patients.

The paper “Analysis of Heart Disease Prediction Using Data Mining Techniques” [4] various data mining techniques of heart disease prediction are discussed. The proposed of this paper gives more accuracy than the present machine learning algorithms. Generally, Naive Bayes classifier is used for the prediction of heart diseases. The main advantage of Bayes classifier is the short training models is used to predict large datasets.

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.7977132

www.ijirid.in

III. PROPOSED SYSTEM

Data collection is the first step in the system's operation. The required data is then preprocessed into the format needed. After that, the data is split into training and testing data. The algorithms are used, and the training data is used to train the model. By testing the system with test data, the accuracy of the system is determined.

The modules listed below are used to implement this system.

- Dataset collection;
- Data pre-processing;
- Data balancing;
- Disease prediction

3.1 DATASET COLLECTION

A group of datasets in order to build our disease prediction system, we first gather a dataset. We divided the dataset into training and testing data after it was collected. The learning and testing of prediction models uses the training dataset. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this research project, 70% of training data is used and 30% of data is used for testing.

3.2 DATA PRE-PROCESSING

The pre-processing of data is a critical stage in the development of a machine learning model. Data that isn't initially clean or in the model's required format can lead to inaccurate results. Pre-processing involves transforming data into the format we need. It is used to handle the dataset's noise, duplicates, and missing values. Activities like importing datasets, splitting datasets, attribute scaling, etc. are all part of data pre-processing. Preprocessing the data is necessary to increase the model's accuracy.[2]

3.3 DATA BALANCING

There are two ways to balance unbalanced datasets. They are both under- and over-sampling.

1. **Under Sampling:**
   Under Sampling reduces the size of the large class to balance the dataset. When there is enough data, this process is taken into account.

2. **Over Sampling:**
   In Over Sampling, the size of the few, scarce samples is increased in order to balance the dataset. When there is not enough data, this process is taken into account.

3.4 PREDICTION OF DISEASE

Machine learning algorithms like Naive Bayes is used for classification.

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.7977132

www.ijirid.in

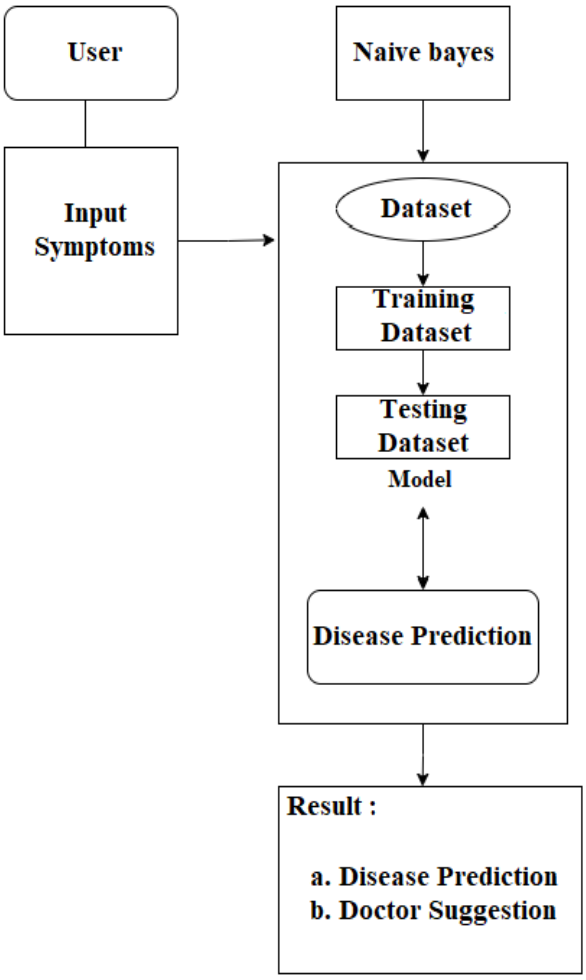

Figure 1: Proposed system

IV. IMPLEMENTATION PROCESS

1. **Data Collection:**

One of the most crucial steps in machine learning workflows is data collection. The quality of the data you gather during data collection determines the potential usefulness and accuracy of your project. To gather data, you must choose your sources and combine the information from each source into a single dataset. This could entail downloading open-source data sets, streaming data from Internet of Things sensors, or building a data lake out of various files, logs, and media.[1]

2. **Data Preprocessing:**

Following data collection, preprocessing is required. Data must first be cleaned, checked, and formatted to create a usable dataset. This may be a fairly simple process if you are gathering data from a single source. However, if you are combining data from various sources, you must ensure that data formats are compatible, that the data is equally trustworthy, and that any potential duplicates are eliminated. [4]

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.7977132

www.ijirid.in

3. **Creating a Dataset:** In this stage, processed data is divided into three datasets for training, validating, and testing:
   - **Training Set:** used to instruct the algorithm in the initial stages of information processing. This set uses parameters to specify model classifications.
   - **Validation Set:** used to gauge the model's degree of accuracy. The model's parameters are adjusted using this dataset.
   - **Test Set:** utilised to evaluate the performance and accuracy of the models. This set aims to highlight any flaws or improper training in the model.

4. **Training and Improvement:** You are ready to train your model once you have the datasets. In order for your algorithm to learn the appropriate parameters and features used in classification, you must feed it your training set. Following the conclusion of training, the model can be improved using the validation dataset. This includes adjusting model-specific settings (hyperparameters) until an acceptable accuracy level is attained, which may entail changing or eliminating variables.[3]

5. **Evaluation of Machine Learning:** After your model accuracy has been optimised and a suitable set of hyperparameters has been identified, you can test your model. To confirm that your models are using accurate features, testing makes use of your test dataset. Based on the feedback you receive you may return to training the model to improve accuracy, adjust output settings, or deploy the model as needed.[3]

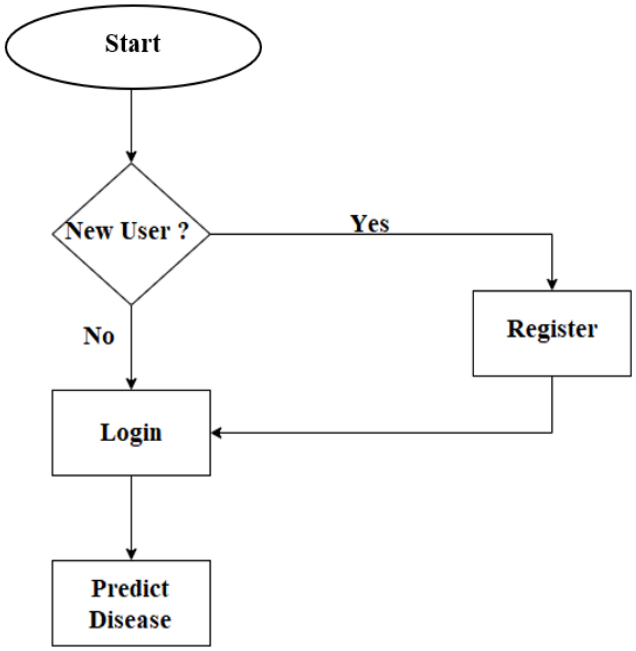

Figure 2: Dataflow Diagram

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.7977132

www.ijirid.in

V. RESULT

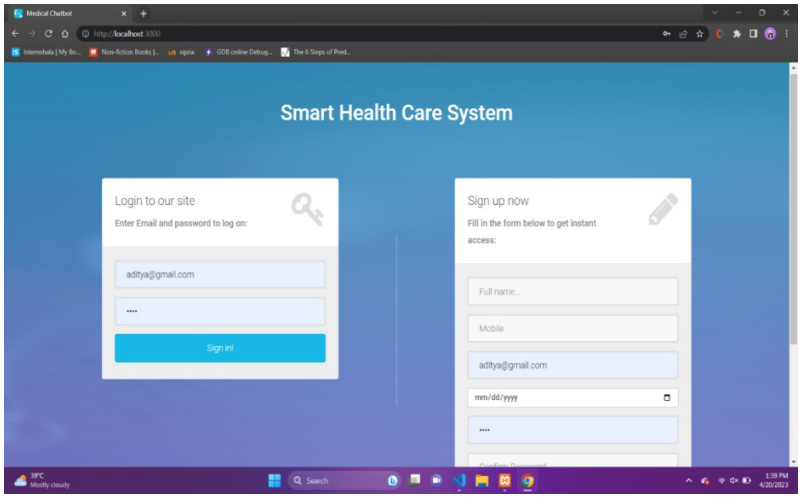

Figure 3: User authentication using Login and Sign Up.

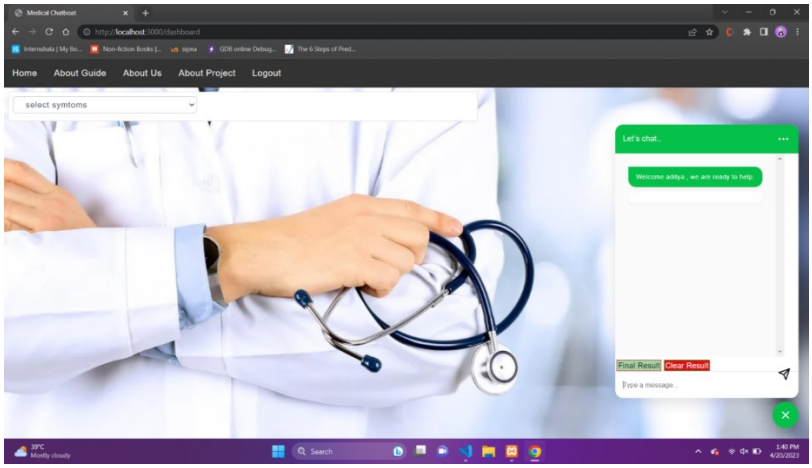

Figure 4: Disease Prediction Chatbot

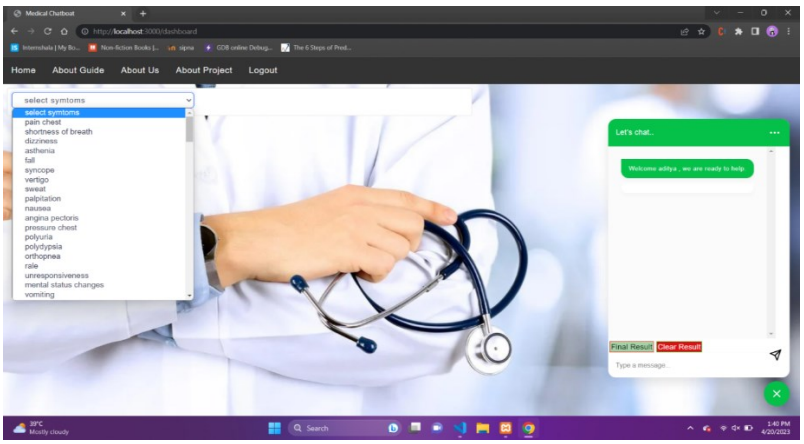

Figure 5: Selection of Symptoms

ISSN (Online): 2583 648X

I J I R I D

International Journal of Ingenious Research, Invention and Development

Volume 1 | Issue 3 | April 2023

Scientific Journal Impact Factor (SJIF 2023): 3.647

DOI: 10.5281/zenodo.7977132

www.ijirid.in

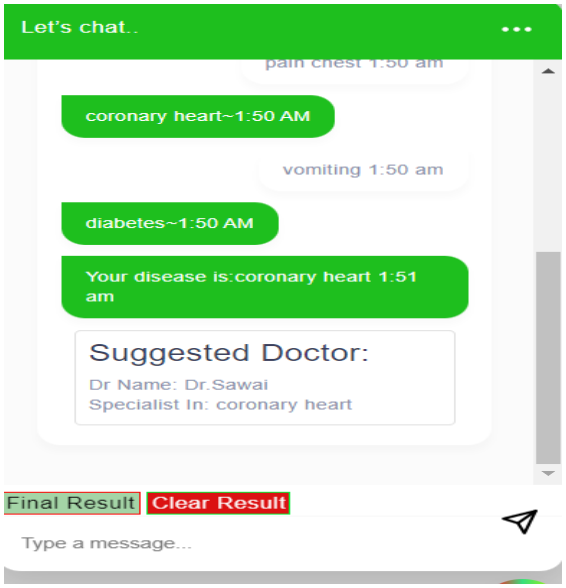

Figure 6: Predicted Disease with Suggested Doctor

VI. CONCLUSION

In conclusion, by giving healthcare professionals cutting-edge tools to analyze patient health data and anticipate potential health issues, smart healthcare prediction systems have the potential to revolutionize healthcare. Early health issue detection, individualized treatment plans, better patient outcomes, cost savings, and improved patient experience are just a few advantages that these systems provide. They may also have some disadvantages, such as issues with data privacy, limitations on accuracy, reliance on technology, the possibility of bias, and moral and legal issues. Despite these difficulties, the creation and application of intelligent healthcare prediction systems are probably going to gain popularity and significance in the healthcare sector. To make sure that these systems are efficient, accurate, secure, and respectful of patients' privacy and rights, healthcare providers and developers must collaborate. By doing this, they can decrease healthcare costs and boost the effectiveness of the healthcare system while also enhancing the quality of care and outcomes for patients.

REFERENCES

[1] Nikita Kamble, International Journal of Scientific Research in Computer Science Engineering and Information Technology, Vol. 2, Issue 5, 2017, “Smart Health Prediction System Using Data Mining”.

[2] Prof. Krishna Kumar Tripathi, International Research Journal of Engineering and Technology (IRJET), Vol.5 Issue:4, Apr-2018, “A Smart Health Prediction Using Data Mining”.

[3] G. Pooja Reddy, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol-8 Issue-6, April 2019, “Smart E-Health Prediction System Using Data Mining”.

[4] S. SHARMILA, International Journal of Advanced Networking & Applications (IJANA), Vol: 08, Issue: 05, 2017, “Analysis of Heart Disease Prediction Using Data Mining Techniques”.