



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Disease Prediction System Using Machine Learning

**Shubham Rameshwar Mali, Onkar Avinash Tepale, Shubham Ramrao Gilbile,
Tanmay Nitin Kulkarni.**

Department of Computer Science Engineering, SBERCT's BIT Barshi, India

ABSTRACT: Rapid proliferation of Internet technology and handheld devices has opened up new avenues for online healthcare system. There are instances where online medical help or healthcare advice is easier or faster to grasp than real world help. People often feel reluctant to go to hospital or physician on minor symptoms. However, in many cases, these minor symptoms may trigger major health hazards. As online health advice is easily reachable, it can be a great head start for users. Moreover, existing online health care systems suffer from lack of reliability and accuracy. Herein, we propose an automated disease prediction system (ADPS) that relies on guided (to be described later) user input. The system takes input from the user and provides a list (topmost diseases have greater likelihood of occurrence) of probable diseases. The accuracy of ADPS has been evaluated extensively. It ensured an average of 14.35% higher accuracy in comparison with the existing solution.

I. INTRODUCTION

Use of computing in the field of medicine can be seen from the early 1950s. However, the first applications of AI in medicine can only be seen during the 1970s through expert systems such as INTERNIST-I, MYCIN, ONCOSIN. The application of artificial intelligence in medicine was mostly limited in the US before 1980. An international conference was organized on 13-14 September 1985 in Italy. With the aim of creating an active research community, in 1986 "Society for Artificial Intelligence in Medicine" was established which biennially organizes international conferences.

One of the problems associated with using artificial intelligence in medicine is unavailability of data, which can be solved using Electronic Medical Records (EMR). The concept of EMR was introduced by Larry Weed during the late 1960s. The US government started using EMR in the 1970s with the Department of Veteran Affairs.

"Knowledge Engineering" was the major research theme during the 1980s, but after 1990 the research shifted to various topics. After 2000, "Machine Learning and Data Mining" has been the major research theme.

Types of Diabetes are following:

1) Type one diabetes outcomes due to the failure of the pancreas to supply enough hypoglycemic agent. This type was spoken as "insulin-dependent polygenic disease mellitus" (IDDM) or "juvenile diabetes". The reason is unidentified. The type one polygenic disease found in children beneath twenty years old. People suffer throughout their life because of the type one diabetic and rest on insulin vaccinations. The diabetic patients must often follow workouts and fit regime which are recommended by doctors.

2) The type two diabetes starts with hypoglycemic agent resistance, a situation in which cells fail to respond the hypoglycemic agents efficiently. The sickness develops due to the absence of hypoglycemic agent that additionally built. This type was spoken as "non-insulin dependent polygenic disease mellitus". The usual cause is extreme weight. The quantity of people affected by type two will be enlarged by 2025. The existence of diabetes mellitus is condensed by 3% in rural zone as compared to the urban zone. The prehypertension is joined with bulkiness, obesity and diabetes mellitus. The study found that an individual United Nations agency has traditional vital sign.

A. Detailed Problem Definition

The primary goal is to develop a prediction engine which will allow the users to check he has diabetes or heart disease, for further treatment. The prediction engine requires a large dataset and efficient machine learning algorithms to predict the whether they have diabetes or heart disease sitting at home. The user need not visit the doctor unless presence of the disease. Pre-processing the dataset to train the machine learning models, removing redundant, null, or invalid data for optimal performance of the prediction engine.

Doctors rely on common knowledge for treatment. When common knowledge is lacking, studies are summarized after some number of cases have been studied. But this process takes time, whereas if machine learning is used, the patterns can be identified earlier. For using machine learning, a huge amount of data is required. There is very limited amount

of data available depending on the disease. Also, the number of samples having no diseases is very high compared to the number of samples having the disease.

This project is about performing two case studies to compare the performance of various machine learning algorithms to help identify such patterns in (i) and to create a platform for easier data sharing and collaboration.

B. Justification of the Problem

Clinical decisions are often made based on doctor's insight and experience rather than on the knowledge rich data hidden in the dataset. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The proposed system will integrate clinical decision support with computer-based patient records (Data Sets). This will reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge rich environment which can help to significantly improve the quality of clinical decisions.

There are voluminous records in medical data domain and because of this, it has become necessary to use data mining techniques to help in decision support and prediction in the field of healthcare. Therefore, medical data mining contributes to business intelligence which is useful for diagnosing of disease.

C. Objectives.

Main Objectives.

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set Heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases.

Specific Objectives.

- Provides new approach to concealed patterns in the data.
- Helps avoid human biasness.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests.

Provides new approach to concealed patterns in the data.

- Helps avoid human biasness.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests.

Provides new approach to concealed patterns in the data.

- Helps avoid human biasness.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests.

Provides new approach to concealed patterns in the data.

- Helps avoid human biasness.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests.

Provides new approach to concealed patterns in the data.

- Helps avoid human biasness.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests.

Provides new approach to concealed patterns in the data.

Provides new approach to concealed patterns in the data.

Provides new approach to concealed patterns in the data.

- Provides new approach to concealed patterns in the data.
- Helps avoid human biasness.
- Helps avoid human biasness.
- Helps avoid human biasness.
- Helps avoid human biasness.
- Helps avoid human biasness.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests.

II. ANALYSYS

A. Project plan

Project planning is part of project management, which relates to the use of schedules such as Gantt charts to plan and subsequently report progress within the project environment. Initially, the project scope is defined and the appropriate methods for completing the project are determined. Following this step, the durations for the various tasks necessary to complete the work are listed and grouped into a work breakdown structure. The logical dependencies between tasks are defined using an activity network diagram that enables identification of the critical path.

Project Plan:

Once we examine that the project is feasible, I undertake project planning. The table below describes how we planned my project.

III. DESIGN

System Architecture

The general disease prediction system predicts chance of presence of a disease present in a patient on the basis of their symptoms. It will also recommend necessary precautionary measures required to treat the predicted disease. The system will initially be fed data from different sources i.e. patients, the data will then be pre-processed before further process is carried out, this is done so as to get clean data from the raw initial data, as the raw data would be noisy, or flawed. This data will be processed using Data mining algorithms, the system, will be trained so as to predict the disease based on the input data given by the user.

The system is implemented into two parts, admin part and the user part. The duty of the admin is training the system for creation of the disease prediction model. The user uses the services provided by the model after logging in as the user, entering the symptoms into the model, which in turn returns the predicted results and necessary precautionary measures.

Sr.No	Task Name	Duration	Start	Finish
1	Planning and Design	28 days	11/01/22	09/02/2022
2	ML Model	30 days	10/02/2022	13/03/2022
4	UI, Flask, Coding	32 days	14/04/2022	14/05/2022
5	Deployment	10 days	15/05/2022	25/05/2022

System Architecture

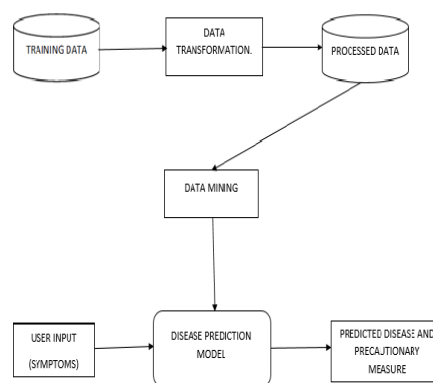


Fig3.1 : System Architecture

State Chart Diagram

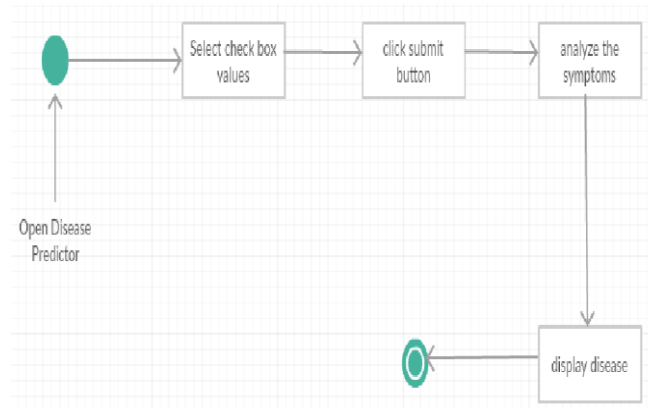


Fig 3.2: State Chart Diagram

Sequence diagram

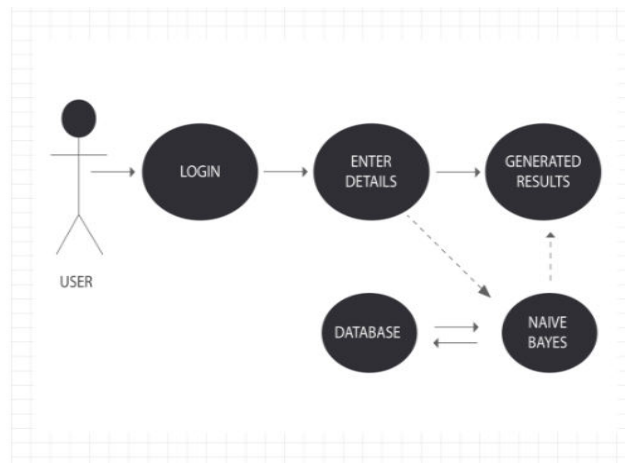


Fig 3.3: Sequence diagram

Activity Diagram

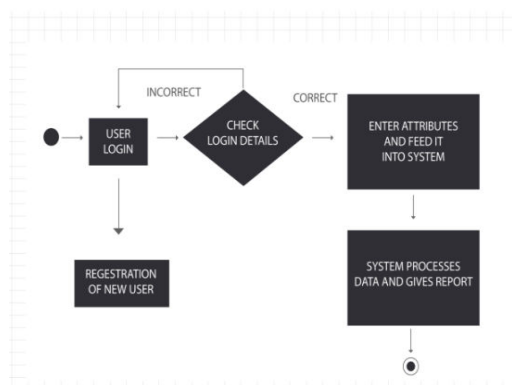


Fig 3.4: Activity Diagram

IV. CODING

Machine Learning Algorithms

In our system we have used immense power of Machine learning algorithms to predict diseases for the end user on the basis of symptoms which they provide with the confidence score with which it is predicting. We have implemented six Machine Learning algorithms and trained them in so that we can use them in our future endeavors. The Algorithms implemented are as follows :-

Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable.

K - Nearest Neighbour

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.

Kernel - Support Vector Machine

SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example, linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. Introduce Kernel functions for sequence data, graphs, text, images, as well as vectors. The most used type of kernel function is RBF. Because it has localized and finite response along the entire x-axis.

Naive Bayes

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e., every pair of features being classified is independent of each other. Naive Bayes has three different models i.e., Gaussian, Multinomial and Bernoulli Naive. Gaussian and Multinomial Naive Bayes are the Two types used in our system for prediction.

Decision Tree

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Framework Used

1. Flask

IDE Used

1. Visual Studio

2. Jupyter Notebook Programming Languages

1. Python

Python Libraries Used

1. Scikit-learn
2. NumPy
3. Pandas
4. Matplotlib
5. SQLAlchemy

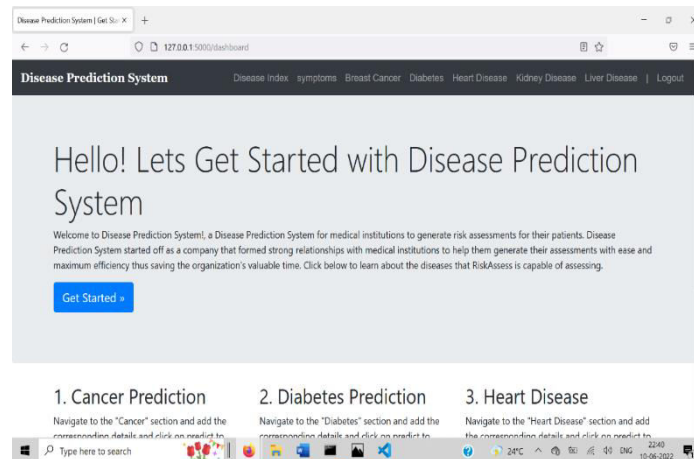
2. HTML

3. CSS

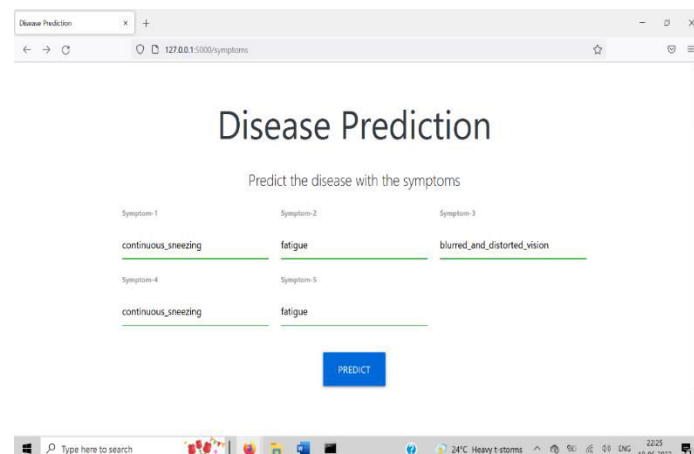
4. JavaScript

V. RESULT SETS

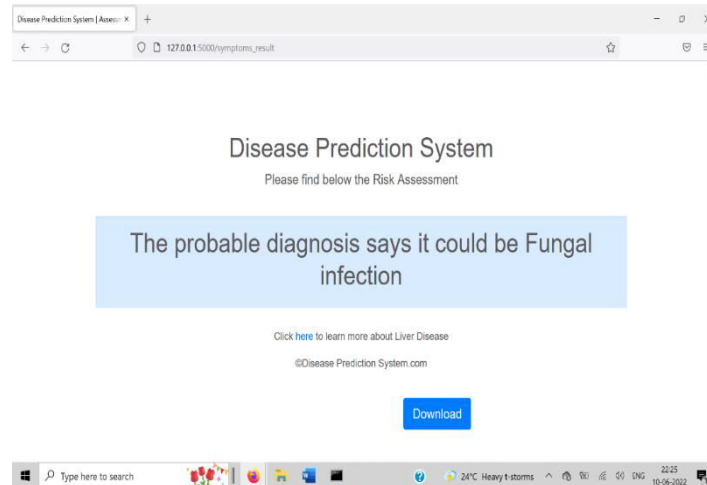
After user login dashboard



Prediction page



Predicted result page



VI. CONCLUSION

As we have seen the real importance and value of Health and lives in the current ongoing pandemic. There is no wealth more than a Healthy life. As a part of serving the common people and solving their health-related problems was our motto throughout the Development of this project. We developed this Webapp for predicting and analyzing the diseases using various Machine Learning Algorithms for early cure and ailment of diseases based on the symptoms provided by the user which indeed became very fruitful and successful by providing highest accuracy and accurate prediction. We also facilitated our users with live consultation with Doctors in ongoing pandemic so that they can get proper medical care from their home. This project is a well-planned Combo for Generic Health care with best optimal solutions be it in predicting a disease and analyzing it. Also getting the information of the risks associated to other organs of body from predicted disease. This project also recommended Specialist Doctor for communication and getting cure as early as possible. This project involved our immense and rigorous efforts for development and resolving the hurdles in day-to-day health problems of people. We would like to make more Projects for the help of needy people and also for the upliftment of society in future.

VII. FUTURE SCOPE

We wish to keep working on our project to make it a marketable and leading product. Some of the additions that we feel are needed for it to happen are:

- Making it deployable on Cloud for global accessibility.
- Increasing the dataset by using our Disease Addition Program for predicting more diseases.
- Making User Interface more User Friendly.
- Adding more functionalities for Real Time Health Care using various parameters like BP, Sugar Level, Body Temperature and RBC WBC Readings, etc.
- Customizing this project with live Health Blogs using data scrapping.

REFERENCES

- [1] B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," 2018
- [2] A. Aldallal and A. Al-Moosa, "Using Data Mining Techniques to Predict Diabetes and Heart Diseases," 2018 4th International Conference on Frontiers of Signal Processing (ICFSP), Poitiers, 2018
- [3] Vijaya Shetty, Kathik GA, M Ashwin, "Symptom Based Health Prediction using Data Mining," 2019
- [4] Dhiraj Dahiade, Gajanan Patle, Ektaa Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," 2019
- [5] K. Lakshmi Priya, Mourya Sai Charan Reddy Kypa, Muchumarri Madhu Sudhan Reddy, G. Ram Mohan Reddy, "A Novel Approach to Predict Diabetes by Using Naive Bayes Classifier," 2020



- [6] R.Jane Preetha Princy, Saravanan Parthasarathy, P. Subha Hency Jose, " Prediction of Cardiac Disease using Supervised Machine Learning Algorithms," 2020
- [7] Aravindkumar.S, Ajith.A, "Prediction of Chronic Disease by Machine Learning," 2019
- [8] Sneha Grampurohit, Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020
- [9] Priyanka Sonar, K. JayaMalini, "DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES," 2019



INNO  SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN  INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details