JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# DISEASE PROGNOSIS USING MACHINE LEARNING

**Dr. k N S Lakshmi, Gopavarjula Sai Madhuri**
Professor & HOD, MCA 2nd year
[1]Master of Computer Application,
[1]Sanketika Vidya Parishad Engineering College, Visakhapatnam, India

## ABSTRACT

This project aims to provide a web platform to predict the occurrences of disease on the basis of various symptoms. The user can select various symptoms and can find the diseases and consult to the doctor online. The early-stage prediction of a disease based on the symptoms becomes difficult for the patient alone. The information available online may always not be correct and may lead to tension and unnecessary panic. To avoid this people should look for their health-related queries on the right place. Thus, to make it easier for people to predict the right disease, development of a machine learning based system has become important. The system collects the symptoms from the user and predicts the correct disease. This will help people to recognize the disease at an earlier stage and take the further decision based on it. We use QUT disease dataset for training as well as testing purpose. We use machine learning algorithms such as Naive Bayes theorem for binary classification and random forest for predicting the disease. We use Django framework to provide a dynamic interface.

**Keywords: Python, Machine Learning, Disease prediction by Symptoms, Django, PostgreSQL.**

## 1 INTRODUCTION

Disease prognosis by symptoms using machine learning is a rapidly growing field that can help doctors and healthcare professionals make more accurate diagnoses and provide personalized treatment plans for patients. With the help of advanced artificial intelligence algorithms and machine learning techniques, healthcare professionals can leverage vast amounts of data to improve disease prediction and early detection, which can lead to better outcomes for patients.

Machine learning models can analyze large datasets of patient symptoms, medical histories, and other clinical data to identify patterns and predict the likelihood of various disease or conditions. These models can also take into account individual patient characteristics such as age, gender, and medical history to provide personalized disease prognosis and treatment recommendations.

## 2.1 FEASIBILITY STUDY

Feasibility study for disease prognosis by symptoms using machine learning involves analyzing the features of a patient's symptoms and predicting the likelihood of specific diseases using machine learning algorithms. The study would involve collecting a large dataset of symptoms and corresponding disease diagnoses from medical records, clinical trials, and relevant literature. This data would then be curated to ensure that it is of high quality, relevant, and comprehensive.

### 2.1.1Economical Feasibility

The economical feasibility of using machine learning for disease prognosis by symptoms depends on several factors, such as the availability of high-quality data, the cost of collecting, storing and processing the data, and the scalability of the machine learning models. Here are some key considerations:

1. Data availability: One of the critical factors that influence the feasibility of using machine learning for disease prognosis is the availability of high-quality data. Without sufficient data that accurately captures the symptoms and outcomes of people with different diseases, machine learning models cannot learn to predict disease prognosis accurately. If data is not readily available, then the cost of data collection and aggregation can be a significant deterrent.
2. Cost of data processing: Machine learning models require vast amounts of data to train effectively. The cost of collecting, storing, and processing this data can be prohibitively high, particularly in low-resource settings. However, recent advancements in data storage, cloud computing, and machine learning algorithms have significantly reduced the cost of processing data, making it more accessible to healthcare providers.

### 2.1.2 Technical Feasibility

It is important to note that technical feasibility is just one aspect of the overall feasibility study. Factors such as ethical considerations, legal compliance, cost-effectiveness, and acceptance from medical professionals and patients are equally important in determining the overall feasibility of a disease prognosis system using machine learning.

### 2.1.3 Social Feasibility

Engaging stakeholders, conducting user studies, and soliciting feedback from patients and healthcare professionals during the development and implementation process can help identify and address potential social concerns. Collaborating with healthcare organizations, regulatory bodies, and patient advocacy groups can also provide valuable insights and help ensure the social feasibility of a disease prognosis system using machine learning.

### 2.2EXISTING SYSTEM

Machine learning can prove to be valuable in disease prognosis by symptoms by identifying patterns and relationships between symptoms and diseases. The existing system for disease prognosis by symptoms using machine learning involves four main steps: 1. Data collection: This involves gathering and organizing a vast amount of data on the symptoms of various diseases. 2. Data preparation: Once the data is collected, it must be cleaned and prepared for analysis. This can involve removing duplicates, correcting errors, and standardizing variables. 3. Model training: This step involves feeding the prepared data into a machine learning model and training it to recognize patterns and make accurate predictions. 4. Prediction and evaluation: After the model is trained and optimized, it can be used to predict the prognosis of a patient based on their reported symptoms.

### 2.3PROPOSED SYSTEM

The proposed system for disease prognosis by symptoms using machine learning has the potential to revolutionize the healthcare industry by enabling early and accurate disease diagnosis and treatment. By leveraging machine learning algorithms, the system can provide personalized and accurate predictions based on the patient's symptoms, medical history, and other relevant data.

## 3SPECIFICATION

### 3.1HARDWARE REQUIREMENTS

| | |
|---|---|
| 1. System | : Pentium IV 2.4 GHz. |
| 2. Hard Disk | : 40 GB. |
| 3. Floppy Drive | : 1.44 Mb. |
| 4. Monitor | : 15 VGA Colour. |
| 5. Mouse | : Logitech. |
| 6. Ram | : 512 Mb. |
| 7.Intel i3 processor(or higher) | |

### 3.2SOFTWARE REQUIREMENTS

| | |
|---|---|
| 1. Operating system | : Windows 7. |
| 2. Frontend | : HTML, CSS, JAVASCRIPT, JQUERY. |
| 3. Backend | : Django(python based web framework). |
| 4. Database | : PostgreSQL. |
| 5. Tools | : PgMyadmin, Orange. |

### 4.1CODING LANGUAGE

#### 4.1.1PYTHON:

Python is an interpreted, high-level, general-purpose programming language. It was developed in the late 1980s by Guido van Rossum,and it was first released in 1991. Python is designed to be easy to read and write, with a syntax that emphasizes code readability. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming styles. Python isoften used in various fields like web development, scientific computing, data analysis, artificial intelligence, and machine learning, among others.

Python has a large standard library that provides support for various tasks such as string processing, operating system interfaces, internet protocols, and more. Additionally, Python has a large and active community that contributes to its development and provides libraries and frameworks for various applications.

Python is also known for its simplicity, elegance, and readability, and it is considered an excellent choice for beginners in programming. Its ease of use and readability make it popular among data scientists, web developers, and other professionals who need to work with large amounts of data.

Overall, Python's versatility, simplicity, and extensive library support make it a powerful and widely used programming language.

**Django:** Django is a Python-based web framework that allows you to quickly create efficient web applications. It is also called batteries included framework because Django provides built-in features for everything including Django Admin Interface, default database – SQLlite3, etc. When you're building a website, you always need a similar set of components: a way to handle user authentication (signing up, signing in, signing out), a management panel for your website, forms, a way to upload files, etc. Django gives you ready-made components to use and that too for rapid development.

## 4.2DEVELOPMENT TOOLS

### 4.2.1PgMyadmin:

PgMyAdmin is the leading opensource management tool for PostgreSQL, the world's most advanced open source database. pgMyAdmin 4 is designed to meet the needs of both novice and experienced Postgres users alike, providing a powerful graphical interface that simplifies the creation, maintenance, and use of database objects.

## 4.3FRONTEND TOOLS

### 4.3.1HTML

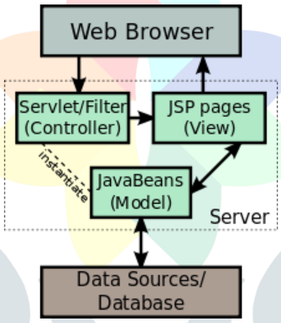
The Hyper Text Markup Language or HTML[11] is the standard markup language for documents designed to be displayed in a web browser. It is often assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as Java Script .Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for its appearance.

In 1980, physicist Tim Berners-Lee, a contractor at CERN, proposed and prototyped ENQUIRE, a system for CERN researchers to use and share documents. In 1989, Berners-Lee wrote a memo proposing an Internet-based hypertext system. Berners-Lee specified HTML and wrote the browser and server software in late 1990. That year, Berners-Lee and CERN data systems engineer Robert Cailliau collaborated on a joint request for funding, but the project was not formally adopted by CERN. In his personal notes of 1990, Berners-Lee listed "some of the many areas in which hypertext is used"; an encyclopaedia is the first entry

### 4.3.2JSP

Jakarta Server Pages[12] (JSP; formerly Java Server Pages) is a collection of technologies that helps software developers create dynamically generated web pages based on HTML, XML, SOAP, or other document types. Released in 1999 by Sun Microsystems, JSP is similar to PHP and ASP, but uses the Java programming language.

JSP allows Java code and certain predefined actions to be interleaved with static web mark-up content, such as HTML. The resulting page is compiled and executed on the server to deliver a document. The compiled pages, as well as any dependent Java libraries, contain Java byte code rather than machine code. Like any other .jar or Java program, code must be executed within a Java virtual machine (JVM) that interacts with the server's host operating system to provide an abstract, platform-neutral environment.



### 4.3.3CSS

Cascading Style Sheets (CSS)[13] is a style sheet language used for describing the presentation of a document written in a markup language such as HTML or XML (including XML dialects such as SVG, MathML or XHTML). CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

CSS is designed to enable the separation of content and presentation, including layout, colours, and fonts. This separation can improve content accessibility; provide more flexibility and control in the specification of presentation characteristics; enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, which reduces complexity and repetition in the structural content; and enable the .css file to be cached to improve the page load speed between the pages that share the file and its formatting.

Separation of formatting and content also makes it feasible to present the same markup page in different styles for different rendering methods, such as on-screen, in print, by voice (via speech-based browser or screen reader), and on Braille-based tactile devices. CSS also has rules for alternate formatting if the content is accessed on a mobile device.

## 5.1MODULE DESCRIPTION

### 5.1.1 Gathering the Data:

Data preparation is the primary step for any machine learning problem. We will be using a dataset from Kaggle for this problem. This dataset consists of two CSV files one for training and one for testing.

### 5.1.2 Cleaning the Data:

Cleaning is the most important step in a machine learning project. The quality of our data determines the quality of our machine learning model. So it is always necessary to clean the data before feeding it to the model for training. In our dataset all the columns are numerical, the target column
i.e. prognosis is a string type and is encoded to numerical form using a label encoder.

### 5.1.3 Model Building:

After gathering and cleaning the data, the data is ready and can be used to train a machine learning model. We will be using this cleaned data to train the Naive Bayes Classifier, and Random Forest Classifier.

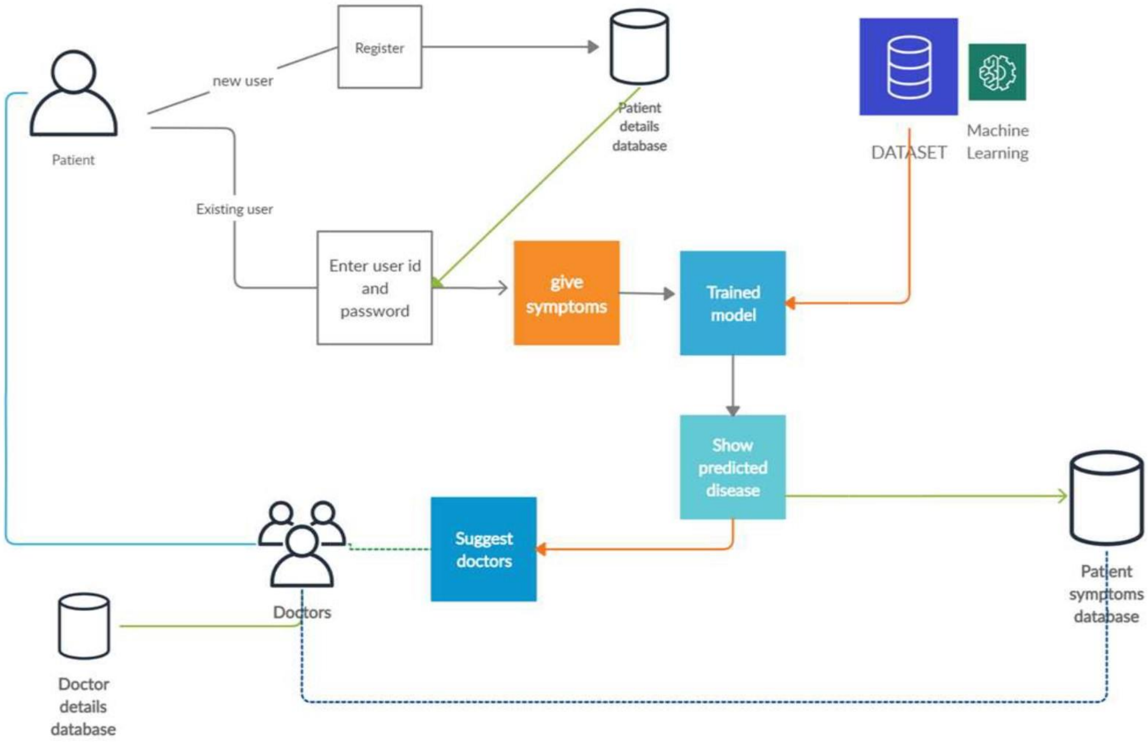
**5.1.4 Inference:** After training the three models we will be predicting the disease for the input symptoms by combining the predictions of the models. This will makes our overall prediction more robust and accurate. The interface will be created using the frontend stack tools.

### 5.1.5 Data Storage:

We use PostgreSQL as a database which is a free and open-source relational database management system. This will be maintained by the admin.

## 6.1ARCHITECTURE

### 6.1.1System Architecture



## 7.1UML DIAGRAMS

Unified Modelling Language is known as UML . A general-purpose modelling language with standards, UML is used in the field of object-oriented software engineering. The Object Management Group oversees and developed the standard.
The objective is for UML to establish itself as a standard language for modelling object-oriented computer programmes. UML now consists of a meta-model and a notation as its two main parts. In the future, UML might also be coupled with or added to in the form of a method or process. The Unified Modelling Language is a standard language for business modelling, non-software systems, and describing, visualising, building, and documenting the artefacts of software systems.

There are several types of UML diagrams that can be used to model disease prognosis by symptoms:

**Use Case diagram:** This diagram can be usedto represent the different actors, such as doctors or patients, and how they interact with the system. The primary use case would be disease prognosis, and the primary actor would be the doctor. The diagram would illustrate the different steps involved in the prognosis process2. Activity diagram: This diagram can be used to model the different activities or steps involved in the disease prognosis process. The activities could include symptoms assessment, analysis, diagnosis, treatment and follow up.

**Class diagram:** This diagram can be used to represent the different classes involved in the disease prognosis process, including symptoms, diseases, medical tests, and treatments. It can help to depict the relationships between the different classes, such as the hierarchical structure of diseases or the association of specific symptoms with certain diseases.

**Sequence diagram:** This diagram can be used to illustrate the interactions between different system components or classes during the disease prognosis process. It can show the different steps involved in the diagnosis and treatment process, including the role of medical tests and the decisions made by doctors based on the symptoms.

**State chart diagram:** This diagram can be used to model the different states and transitions involved in the disease prognosis process. For example, a state chart diagram could illustrate the different stages of a disease and how the treatment approach would change over time.

### 7.2Relationships in UML

There are four kinds of relationships in the UML:

•Dependency
•Association
•Generalization
•Realization

A dependency is a semantic relationship between two things in which a change to one thing may affect the semantics of the otherthing (the dependent thing)

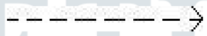


**Figure7.2:** Dependencies

An association is a structural relationship that describes a set links, a link being a connection among objects. Aggregation is a special kind of association, representing astructural relationship[14] between a whole and its parts.

**Figure7.2.1:** Association

A generalization is a specialization/ generalization relationship in which objects of the specialized element (the child) are substitutable for objects of the generalized element (the parent).

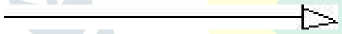


**Figure7.2.2:** Generalization

A realization is a semantic relationship between classifiers, where in one classifier specifies a contract that another classifier guarantees to carry out.

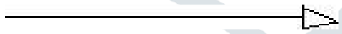
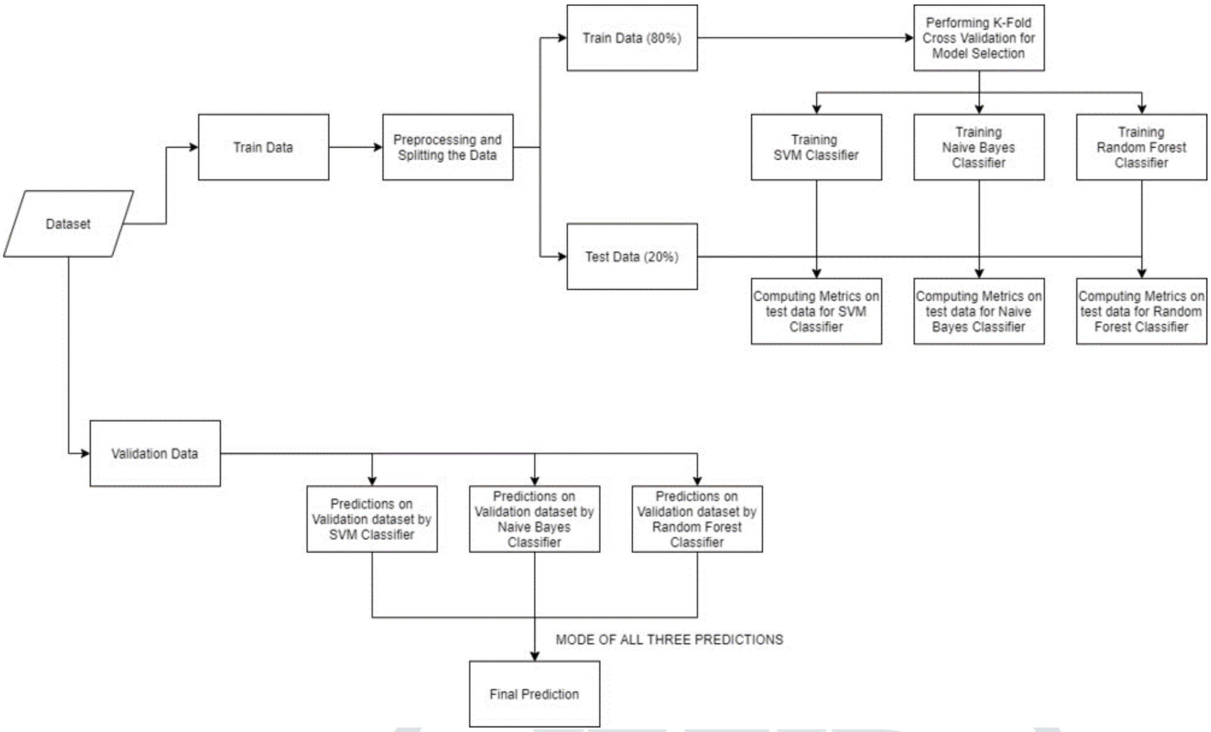


**Figure7.2.3: Realization**

### 8.1SYSTEM IMPLEMENTATION

Systems implementation is the process of:


Dataset
Train Data
Preprocessing and Splitting the Data
Train Data (80%)
Test Data (20%)
Performing K-Fold Cross Validation for Model Selection
Training SVM Classifier
Training Naive Bayes Classifier
Training Random Forest Classifier
Computing Metrics on test data for SVM Classifier
Computing Metrics on test data for Naive Bayes Classifier
Computing Metrics on test data for Random Forest Classifier
Validation Data
Predictions on Validation dataset by SVM Classifier
Predictions on Validation dataset by Naive Bayes Classifier
Predictions on Validation dataset by Random Forest Classifier
MODE OF ALL THREE PREDICTIONS
Final Prediction

### 9.1 TESTING

Testing is done to look for mistakes. Testing is the process of looking for any flaws or weaknesses in a piece of work. It offers a means of examining the operation of parts, subassemblies, assemblies, and/or a finished product. It is the process of testing software [18] to make sure that it satisfies user expectations and meets requirements without failing in an unacceptable way. Different test types exist. Every test type responds to a certain testing requirement.

**9.1.1 Classification:** This is a type of testing technique that involves identifying and categorizing a patient's symptoms to help predict the likelihood of a particular illness. Machine learning algorithms can be trained on large datasets of symptoms and known diagnoses to accurately predict the probability of a certain disease based on a patient's symptoms.

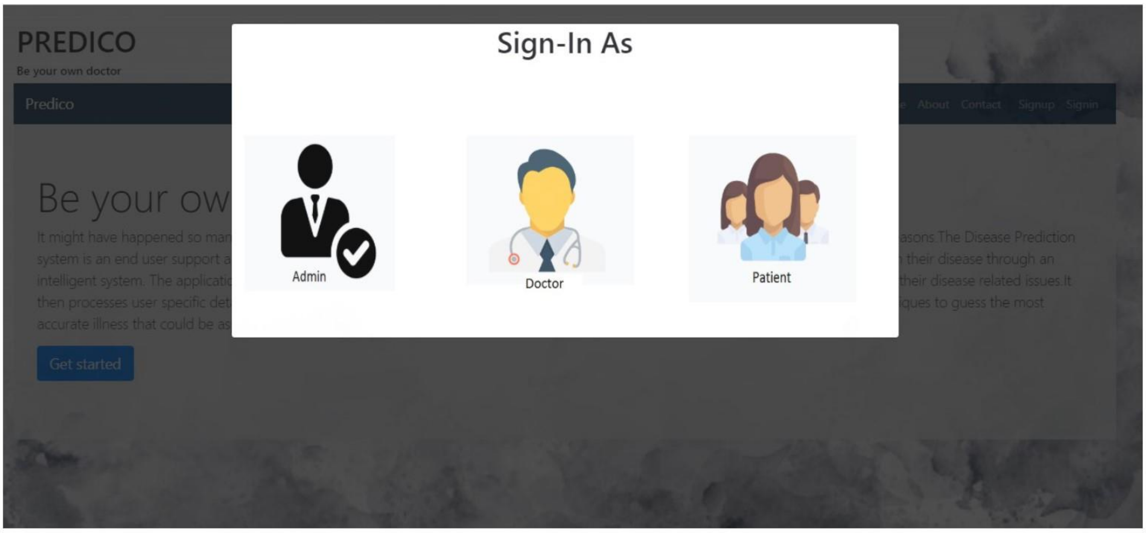
**Regression:** Another type of testing technique involves using regression analysis to correlate a patient's symptomswith disease prognosis. Regression models can be trained on a large dataset of patient information to identify patterns and relationships between symptoms and disease outcomes. The model can then use this information to predict disease prognosis based on a patient's symptoms**.**

**Clustering:** Clustering is a technique that involves grouping patients with similar symptoms and disease prognosis together. This type of testing can help identify trends and patterns in patient data. Machine learning algorithms can be trained to cluster patients based on factors such as age, gender, and symptoms, allowing doctors to more accurately predict disease prognosis and tailor treatment plans accordingly.

**Decision trees:** Decision trees are a type of machine learning algorithm that can be used to predict disease prognosis based on a patient's symptoms. Decision trees work by creating a series of yes/no questions based on a patient's symptoms, with each answer leading to a new question until a definitive diagnosis is reached. This type of testing can be particularly useful for complex diseases with multiple potential diagnoses.

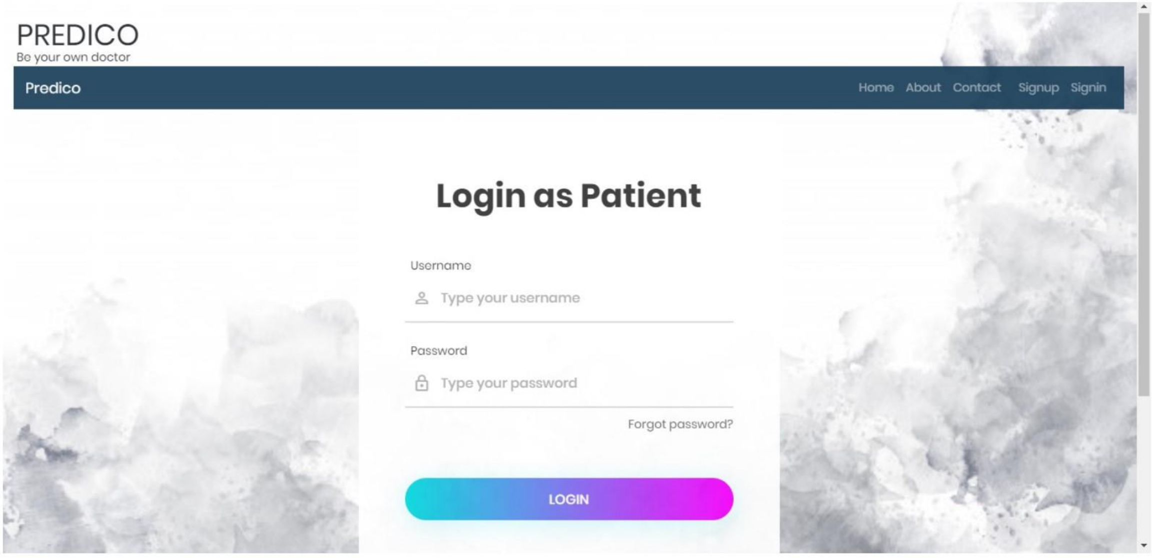
Overall, machine learning has the potential to revolutionize disease prognosis by allowing doctors to analyze large amounts of patient data and quickly identify patterns and trends. By combining machine learning with traditional diagnostic techniques, doctors can provide more accurate diagnoses and better patient outcomes.

**Login Modal-**



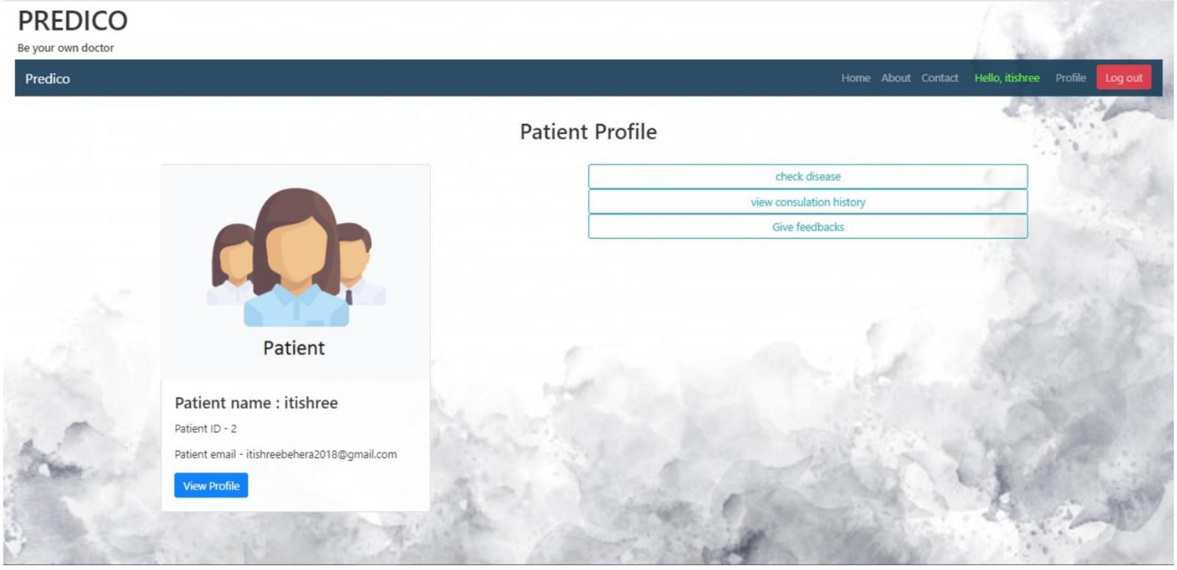
The disease prediction method has three users: a doctor, a patient, and an administrator. The system authenticates each of the system's users

**Login as Patient -**



If the patient and a doctor is a new user, they have to register using signup form.

**Patient UI -**



PREDICO

Be your own doctor

Patient Profile

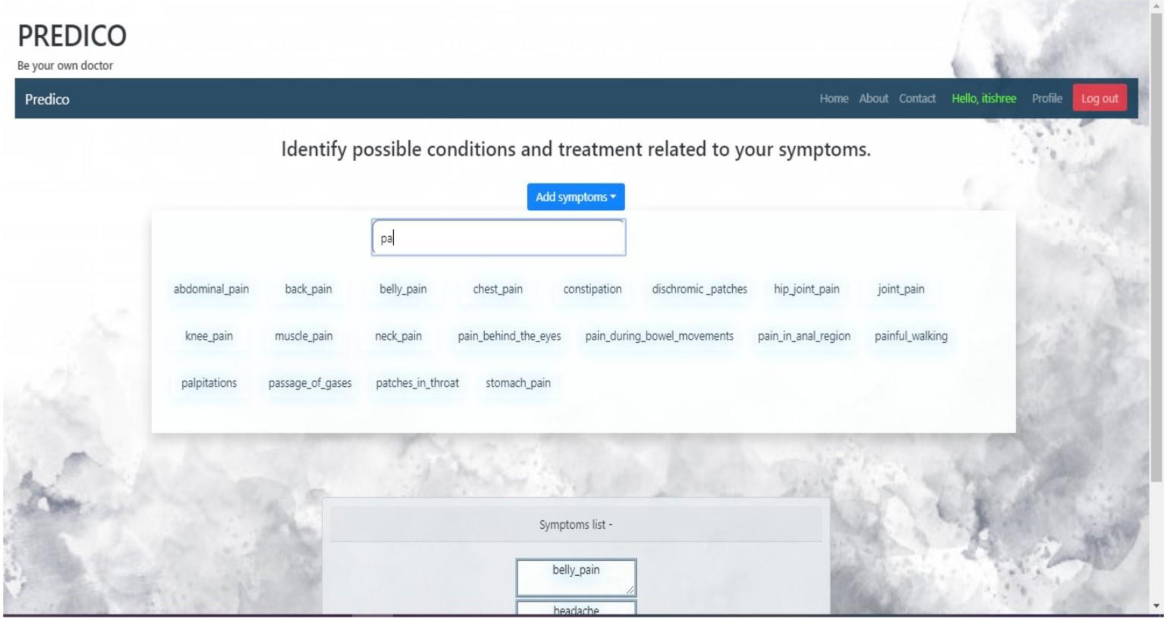
Patient

Patient name : itishree

Patient ID - 2

Patient email - itishreebehera2018@gmail.com

**Check Disease Entering Symptoms -**



PREDICO

Be your own doctor

Identify possible conditions and treatment related to your symptoms.

pa

abdominal_pain back_pain belly_pain chest_pain constipation dischromic _patches hip_joint_pain joint_pain

knee_pain muscle_pain neck_pain pain_behind_the_eyes pain_during_bowel_movements pain_in_anal_region painful_walking

palpitations passage_of_gases patches_in_throat stomach_pain

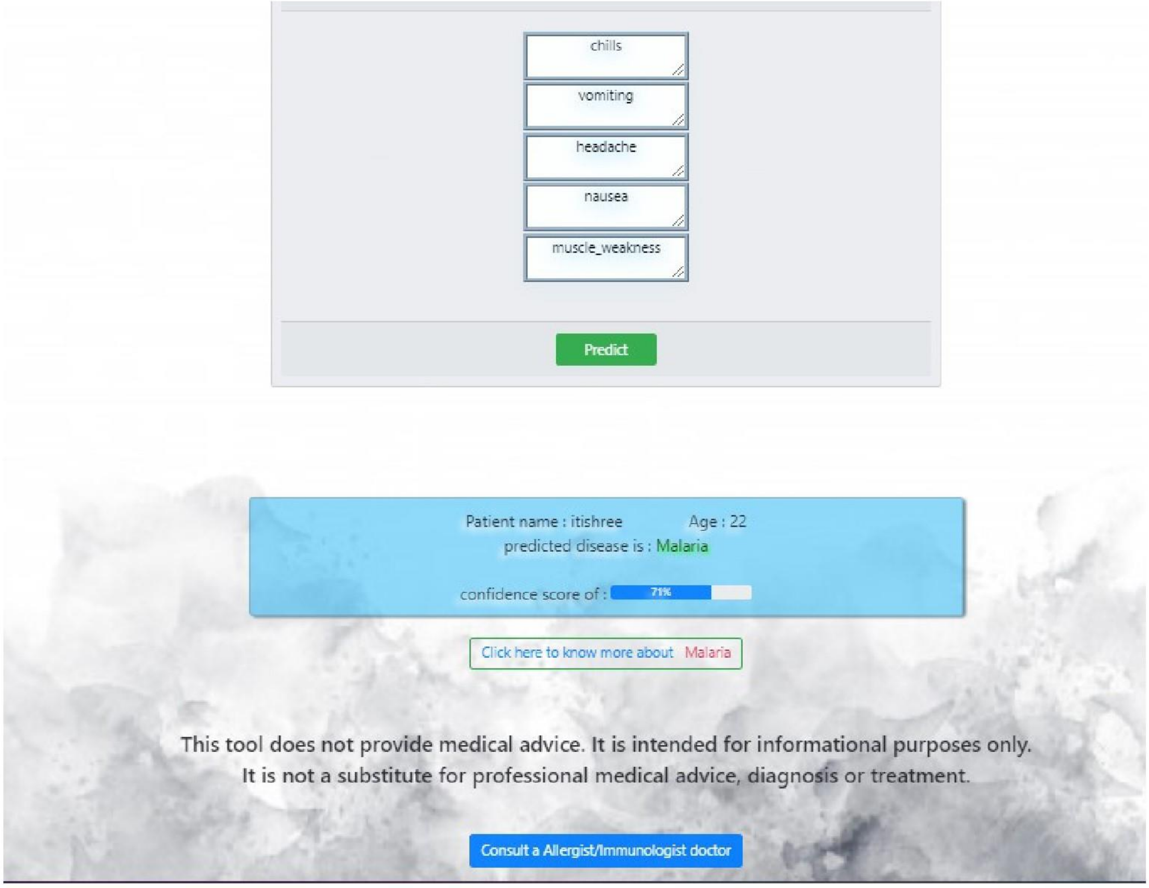
Symptoms list -

belly_pain

headache

Patients after signing in will be prompted to a screen where they can see their profile, they can enter the symptoms faced and can get the disease from the trained model and can consult a specialist doctor and can chat with them.

**Predictions -**



chills

vomiting

headache

nausea

muscle_weakness

Patient name : itishree Age : 22

predicted disease is : Malaria

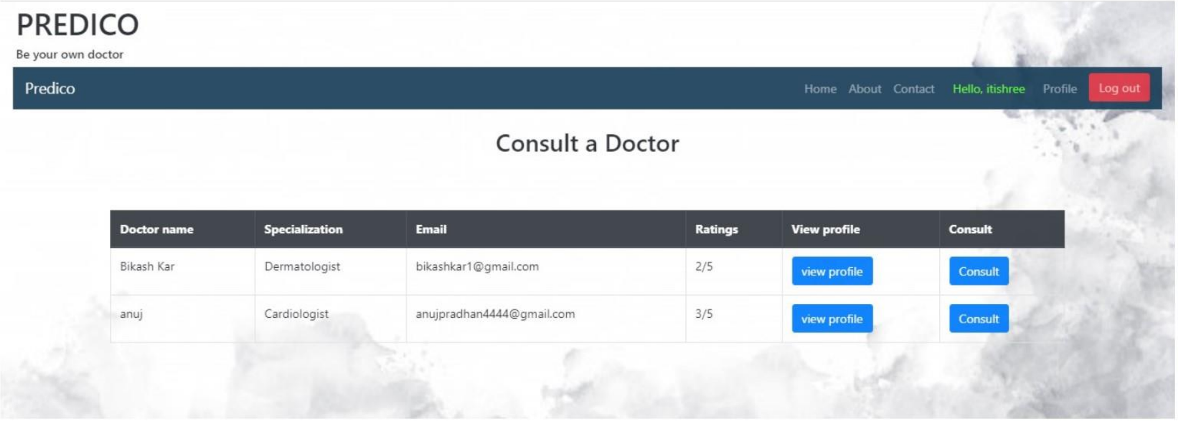
confidence score of : 71%

Click here to know more about Malaria

This tool does not provide medical advice. It is intended for informational purposes only. It is not a substitute for professional medical advice, diagnosis or treatment.

Consult a Allergist/Immunologist doctor

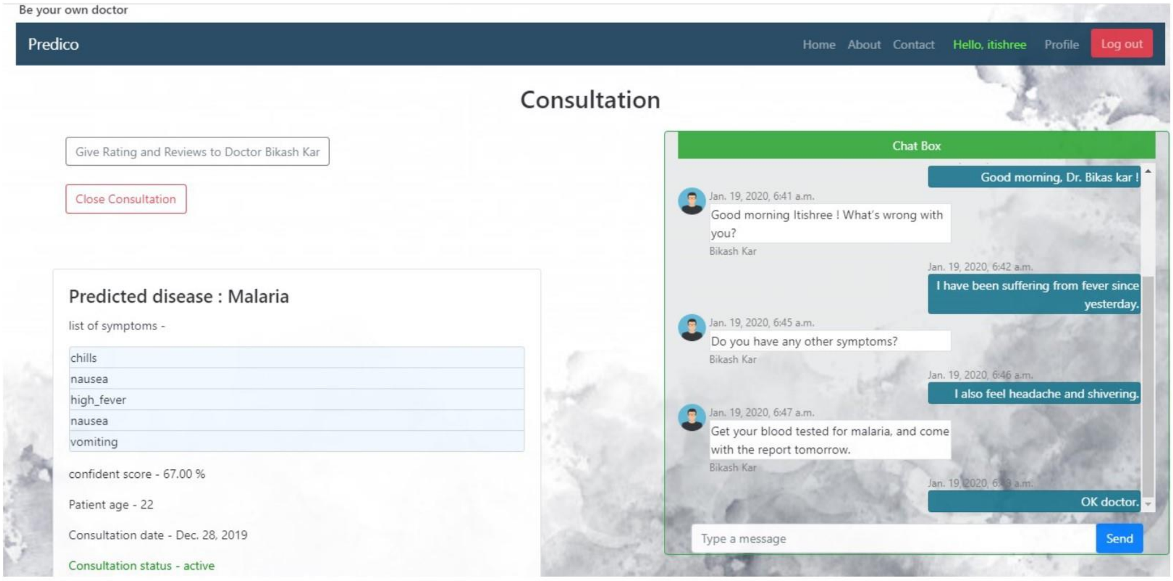
**Consult a Doctor -**



Consult a Doctor

| Doctor name | Specialization | Email | Ratings | View profile | Consult |
|---|---|---|---|---|---|
| Bikash Kar | Dermatologist | bikashkar1@gmail.com | 2/5 | view profile | Consult |
| anuj | Cardiologist | anujpradhan4444@gmail.com | 3/5 | view profile | Consult |

**Consultation UI-**



Consultation

Give Rating and Reviews to Doctor Bikash Kar

Close Consultation

Predicted disease : Malaria

list of symptoms -

- chills
- nausea
- high_fever
- nausea
- vomiting

confident score - 67.00 %

Patient age - 22

Consultation date - Dec. 28, 2019

Consultation status - active

Chat Box

Good morning, Dr. Bikas kar !

Jan. 19, 2020, 6:41 a.m.
Good morning Itishree ! What's wrong with you?
Bikash Kar

Jan. 19, 2020, 6:42 a.m.
I have been suffering from fever since yesterday.

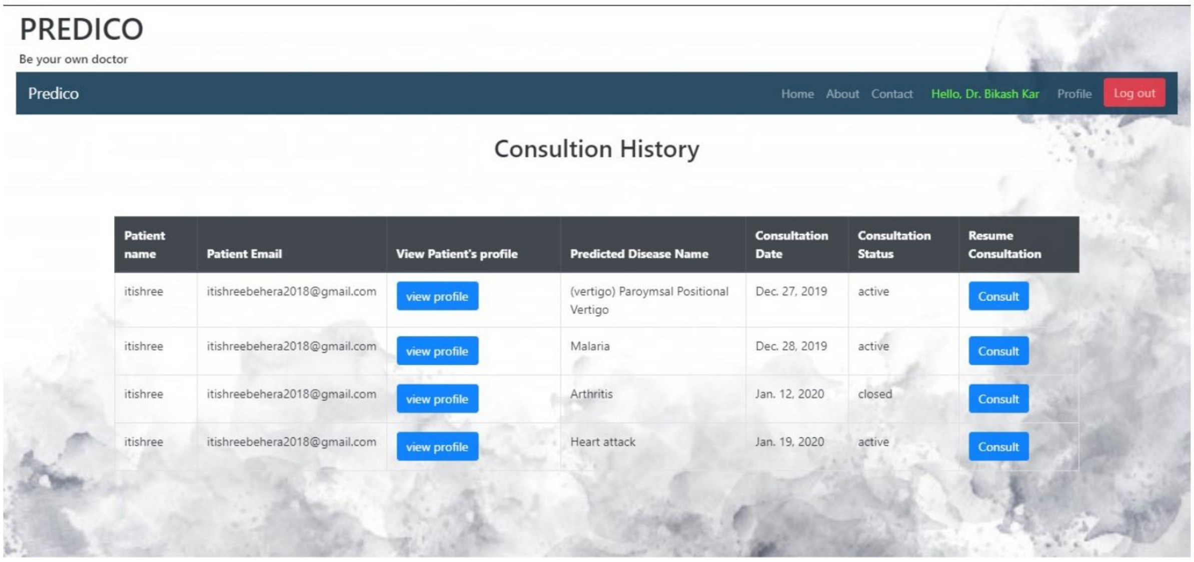
Jan. 19, 2020, 6:45 a.m.
Do you have any other symptoms?
Bikash Kar

Jan. 19, 2020, 6:46 a.m.
I also feel headache and shivering.

Jan. 19, 2020, 6:47 a.m.
Get your blood tested for malaria, and come with the report tomorrow.
Bikash Kar

Jan. 19, 2020, 6:48 a.m.
OK doctor.

**Consultation history – (Doctor)**



Consultion History

| Patient name | Patient Email | View Patient's profile | Predicted Disease Name | Consultation Date | Consultation Status | Resume Consultation |
|---|---|---|---|---|---|---|
| itishree | itishreebehera2018@gmail.com | view profile | (vertigo) Paroymsal Positional Vertigo | Dec. 27, 2019 | active | Consult |
| itishree | itishreebehera2018@gmail.com | view profile | Malaria | Dec. 28, 2019 | active | Consult |
| itishree | itishreebehera2018@gmail.com | view profile | Arthritis | Jan. 12, 2020 | closed | Consult |
| itishree | itishreebehera2018@gmail.com | view profile | Heart attack | Jan. 19, 2020 | active | Consult |

## 10.1. CONCLUSION

In conclusion, utilizing machine learning algorithms for disease prognosis by symptoms has shown promising results in improving accuracy, reducing error rates, and enhancing the efficiency of the diagnosis process. This approach enables healthcare professionals to make more accurate and effective decisions in treating patients. By analyzing large datasets and correlations between symptoms and diseases, machine learning can help identify disease patterns and predict potential risks.

## 10.2REFERENCES

1. A. Mir, S.N. Dhage, in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (IEEE, 2018), pp. 1–6

2. Y. Khourdifi, M. Bahaj, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization, Int. J. Intell. Eng. Syst. 12(1), 242 (2019)

3. S. Vijayarani, S. Dhayanand, Liver disease prediction using svm and na¨ıve bayes algorithms, International Journal of Science, Engineering and Technology Research (IJSETR)

4(4), 816 (2015) 4. S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7, 81542 (2019)

5. T.V. Sriram, M.V. Rao, G.S. Narayana, D. Kaladhar, T.P.R. Vital, Intelligent parkinson disease prediction using machine learning algorithms, International Journal of Engineering and Innovative Technology (IJEIT) 3(3), 1568 (2013)

6. A.S. Monto, S. Gravenstein, M. Elliott, M. Colopy, J. Schweinle, Clinical signs and symptoms predicting influenza infection, Archives of internal medicine 160(21), 3243 (2000)

7. R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm 8. D.R. Langbehn, R.R. Brinkman, D. Falush, J.S. Paulsen, M. Hayden, an International Huntington’s Disease Collaborative Group, A new model for prediction of the age of onset and penetrance for huntington’s disease based on cag length, Clinical genetics 65(4), 267 (2004)

## BIBILIOGRAPHY

Dr.K.N.S Lakshmi currently working as professor from Department of Computer Science and Engineering at Sankethika vidya parishad engineering college, affiliated to Andhra University, accredited by NAAC. Madam is currently working as Head of The Department, Published Papers in Various National & International journals. Her Subjects of interests are Machine Learning, Data Mining & Warehousing.

Gopavarjula Sai Madhuri is studying her 2nd year, Master of Computer Applications in Sanketika Vidya Parishad Engineering College, affiliated to Andhra University, accredited by NAAC.With her interest in java, cloud computing and as a part of academic project, she chooses Disease prognosis using Machine Learning. A completely developed project along with code has been submitted for Andhra University as an Academic Project. In completion of her MCA.