IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

HealthMitra: Digital Diagnose Suggestion and Disease Prediction using Machine Learning and Microsoft Azure

Prof. Vaishali Rajmane[1],Chinmay Halsikar[2], Bhushan Deshmukh[3], Shubham Sakhare[4], Pranoj Gonjari[5], Shreyas Patil[6]
Department of Computer Science and Engineering[1-6]
SVERI's College of Engineering, Pandharpur, India
Punyashlok Ahilyadevi Holkar Solapur University, Solapur

Abstract: *In the IT era, technology has greatly changed the medical field. The goal of this research is to create a diagnosis model for a variety of diseases based on their medical information. To create such a model, this system used Random Forest. The intelligent agent is trained using datasets containing copious data regarding patient diseases that have been gathered, refined, categorized, and utilized. After classifying the dataset into training and testing we built a model using a random forest classifier. Model can predict disease based on the medical information of the patient. The patient might then contact the doctor for further therapy based on the results by using AI Chatbot. It is an example of how technology and medical expertise are flawlessly woven together with the goal of achieving "prediction is better than cure.".*

Keywords: Random forest classifier, medical data, classification, and data mining, Microsoft azure, Microsoft cognitive service, knowledge base, AI Bot

I. INTRODUCTION

In the age of the internet and machine learning, we still treat disease the same way. For serious illnesses, traditional diagnostic methods may be inadequate. In response, we are developing a complete AI-based healthcare system called HealthMitra. It's a web application with his amazingly user-friendly UI built using Streamlit, an open-source Python app framework, and core concepts that leverage Microsoft Azure and machine learning. The methodology proposed here provides a better and more efficient alternative to randomly Googling for diagnoses and predicts disease more accurately than conventional methods. Registering with the network allows you to diagnose yourself, get a diagnosis and find contact information for your doctor. HealthMitra uses machine learning algorithms to predict and aid diagnosis in multiple diseases. Advances in machine learning may create methods to improve disease prediction accuracy. Utilization of existing medical information and random forests. If further treatment is needed, the patient can contact the nearest or preferred disease specialist with the help of her AI bot. This method enables free disease prediction and medical consultation.

1.1 Motivation

- Our innovation takes inspiration from ancient Indian civilization; the development of the ancient Indian medicine system can be traced right from the Indus Valley Civilization. We are living in a country where the medical field is far ahead at that time.
- Need of the user is an accurate diagnosis for a disease, cheaper cost, and ease of availability so as per the user requirement.
- Our innovation strictly follows the principle - high accuracy and cheap medical facility for everyone and also it can easily be accessible to everyone at no cost.
- With the help of this innovation we can make some positive changes in the medical field.
- This innovation is built to make ease access to medical prediction and diagnosis systems in rural areas and women.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

- Our final agenda is an accurate and on-time analysis of the health-related problem and free consultation for diagnosis.

II. METHODOLOGY

The proposed system automatically predicts diseases using models trained on medical datasets. This method also displays the confidence value of the prediction. After diagnosing the expected disease, we introduce a specialist who specializes in that disease and provide an online consultation system. The proposed technology will act as a decision support system to assist doctors in diagnosing.

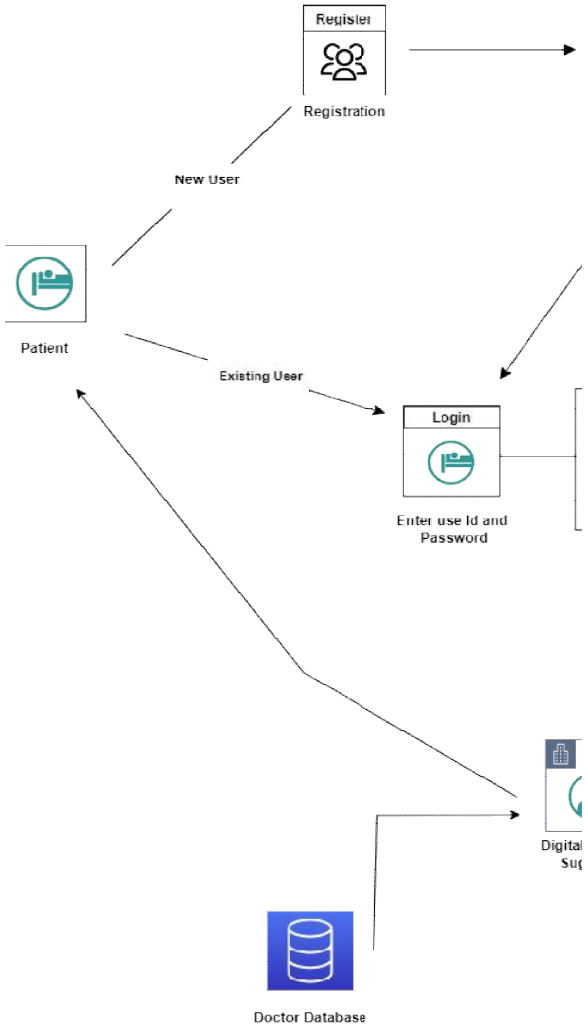

Fig1. Flow Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

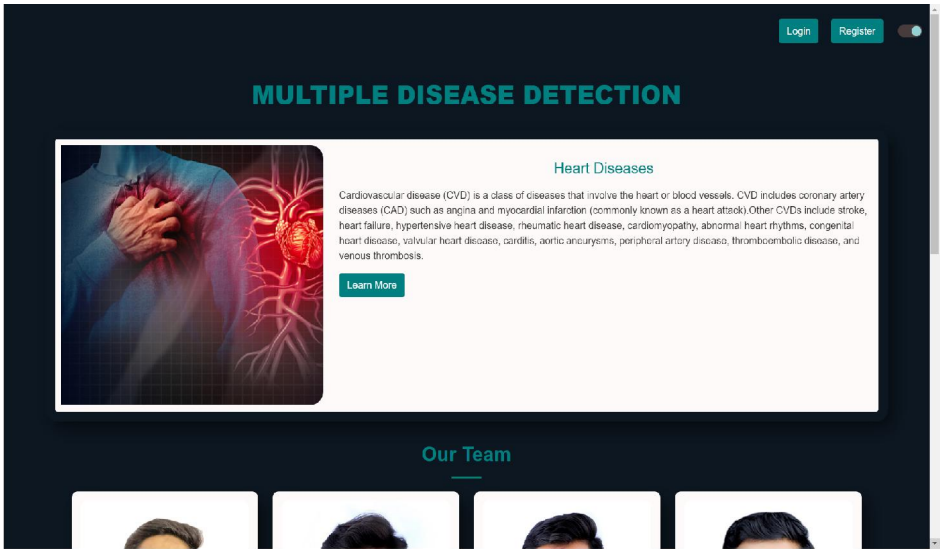

Fig 2. Home Page[6]

When the user visits the application they can register as a patient on the network after the user has successfully registered on the network. They can login in to the web application as a patient.

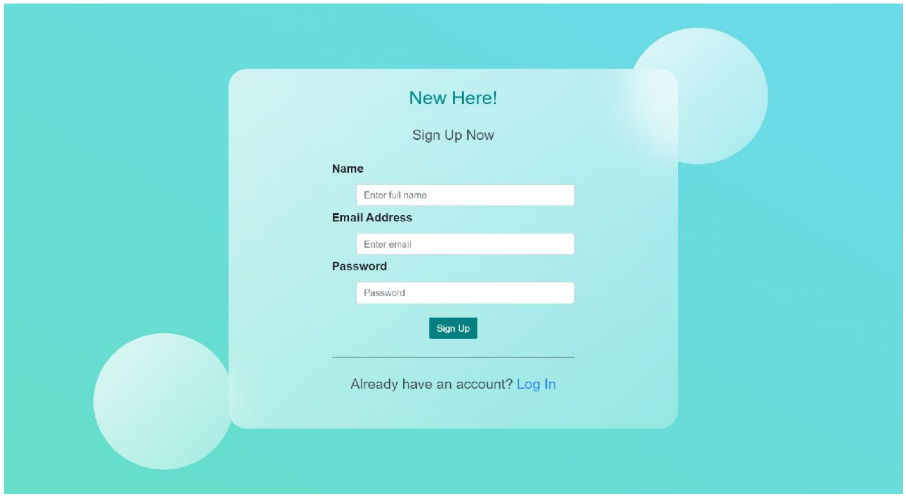

Fig 3. Registration page

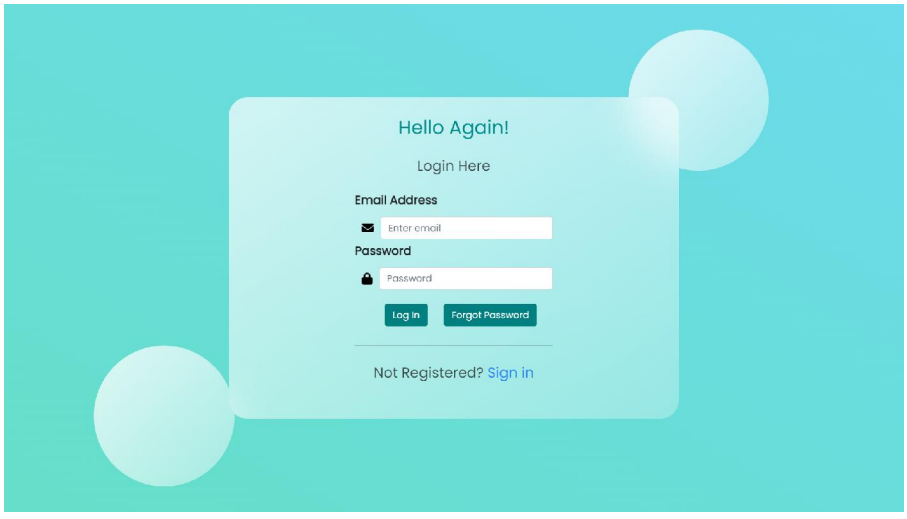

Fig 4. Login page

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

After login as patient –

A) The user is directed to their profile page where they can see their patient ID, name, and email along with the option to edit their information.

B) Alongside that, they have three options- to predict the disease and consultation from AI Bot.

C) When they click on predict disease they will get the option to choose multiple disease from sidebar.

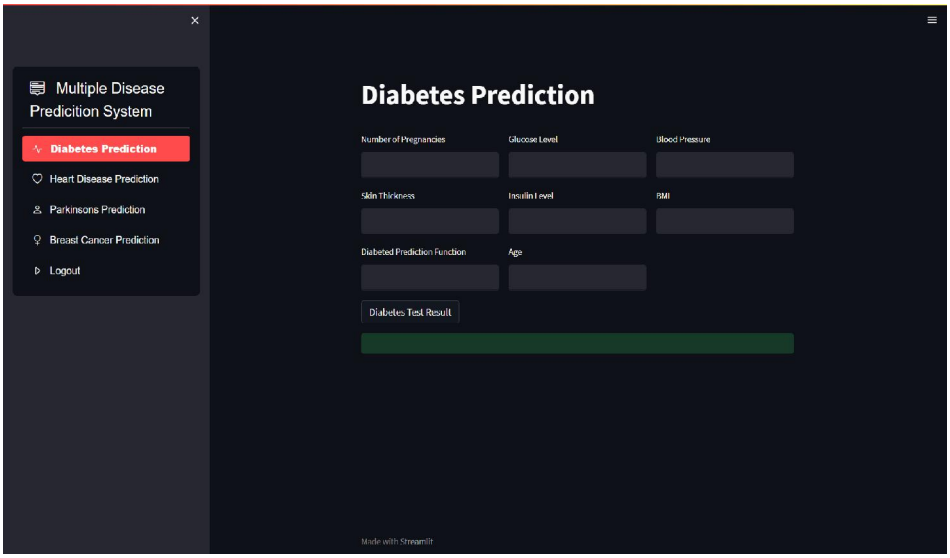

Fig 5. Sidebar option

D) Depending upon the medical information of the patient model will predict the disease with high accuracy.

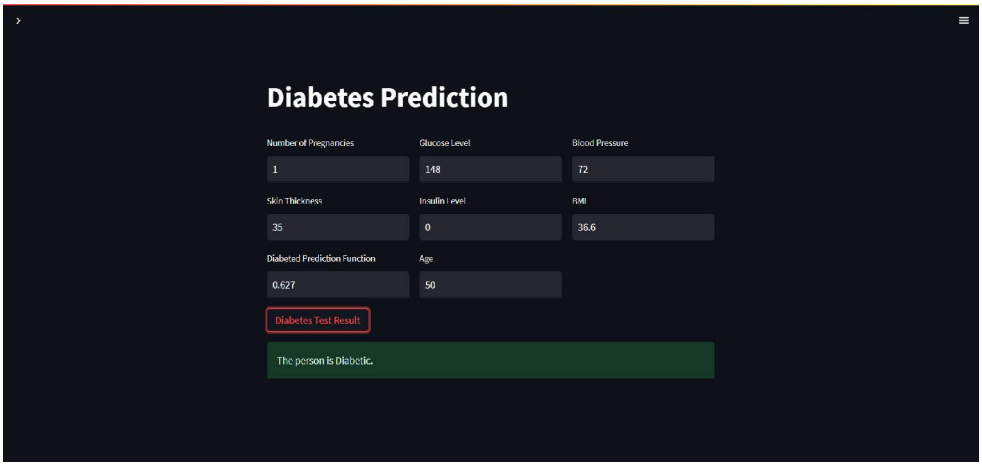

Fig 6. Disease test result

E) Additionally, the application provides links that guide the user to better understand the predicted disease.

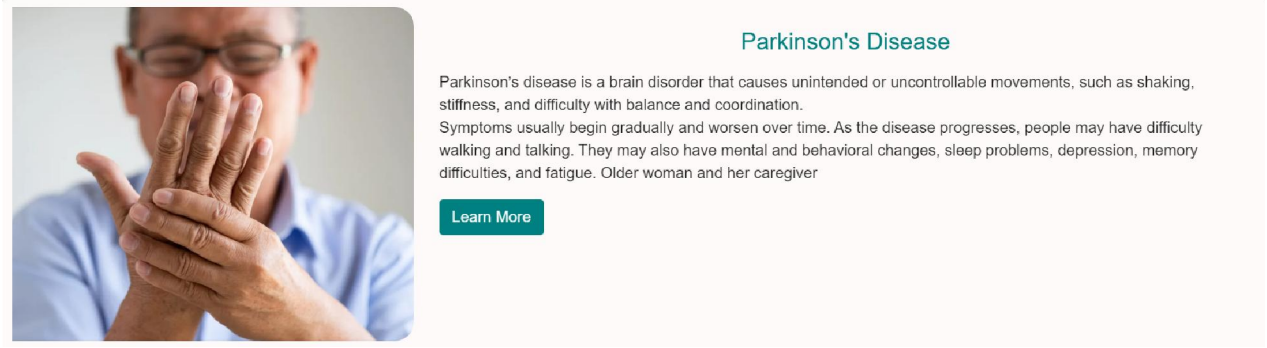

Fig 7. Information about disease[6]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

F) Furthermore, the main element of the proposed model is that alongside the visualization the framework offers the chance to associate with the specialist representing considerable authority in that specific field to the client who is enrolled in the organization alongside their contact subtleties.

G) Patients can get to a rundown of specialists who have practical experience in their field with assistance of Azure cognitive service based intelligence Chatbot, simulated intelligence Chatbot will assist patients to get specialist data with geographic area.

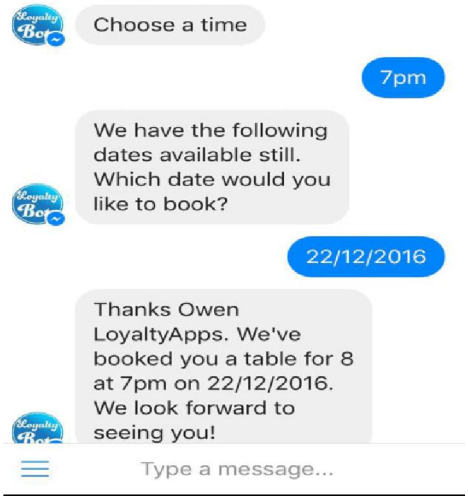

Fig 8. AI Chatbot

H) Users can contact us 24x7 for any kind of query.

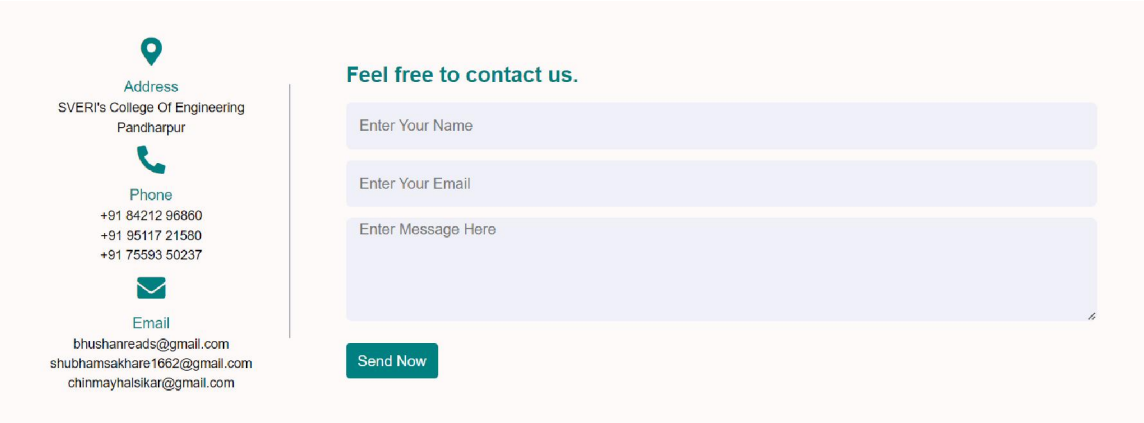

Fig 9. Contact form

Data Preparation: The dataset is publicly available on the Kaggle website. H. No dummy value is entered. The goal of classification is to predict whether a patient will have her 10-year risk of multiple diseases. Records provide patient information. It contains over 4,000 records and over 15 attributes for each disease.

Out[ ]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Fig 10. Dataset

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

Cleaning the Data:

The most important steps of any machine learning project is cleaning. The quality of machine learning models depends on the quality of data. Therefore, we need to clean the data before feeding it to the model for training. All columns in the dataset are numeric except Forecast, which is the target column. Forecast is a text type encoded into a numeric format using a label encoder.

Dataset Splitting: When training a machine learning model, the dataset is split into two parts:

A) The training dataset

B) The testing dataset.

Data is divided into an 80:20 structure, which means that 80% of the information is utilized to train the model and 20% is used to evaluate the model's performance.

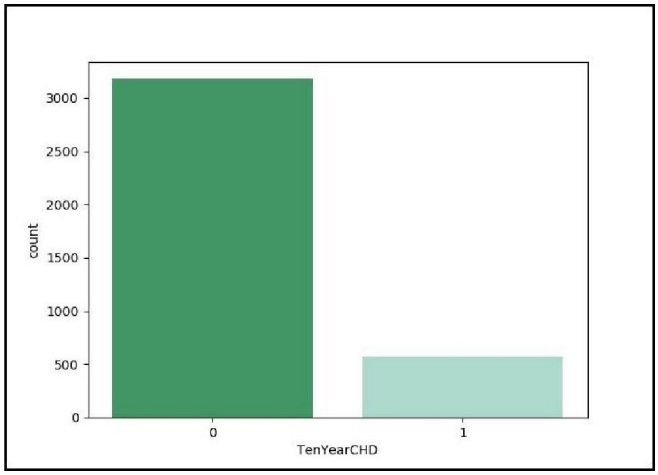

Fig 11. Dataset splitting[6]

Random Forest Classifier: A random forest consists of many individual decision trees acting as an ensemble.Each individual tree in the random forest produces a hypothesis, and the class with the most votes is the model's prediction. In order to build a comprehensive model two different factors are combined..

III. RESULT

If the patient is logged in, they can access disease predictions. This gives you accurate forecasts in a seamless, one-click solution.

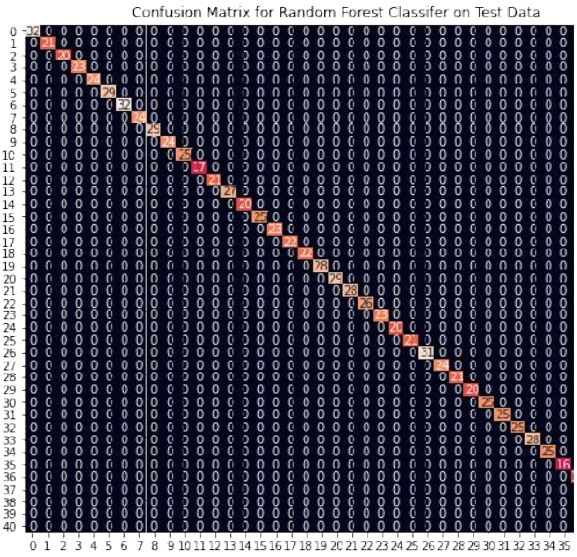

Fig 12. Accuracy on train data by Random Forest Classifier: 100. [6]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

Accuracy on test data by Random Forest Classifier: 100.0
The result is shown to the patient with not only a high accuracy but this ensures that not one medical information entity is affiliated to disease prediction and the result is not biased.

IV. FUTURE SCOPE

- A premium account option is available for patients.
- Video call function.
- This Website's Account Linking feature allows users to connect their accounts to other online services such as Gmail and social media.
- Add map functionality to your site, including adding APIs. Partner with pharmacies and offer discounts on medicines to your patients.

V. AREA OF IMPROVEMENT

In future work, the creation of more complex Machine Learning algorithms will be much needed to increase the efficiency of disease prediction. In addition, learning models should be calibrated more often after the training phase for potentially better performance. Moreover, datasets should be expanded on different demographics to avoid overfitting and also increase the accuracy of the deployed models. Finally, more relevant feature selection methods should be used to improve the performance of the learning model.

VI. CONCLUSION

The machine learning model we have built is around 90% to 97% accurate. The diseases for which there are no diagnostics methods. Machine learning models are able to predict whether the person has disease or not. This is the power of machine learning technology by using which many of the real-world problems can be solved.

REFERENCES

[1] Visit our project at http://healthmitra.live
[2]Github