Grenze International Journal of Engineering and Technology, June Issue

Healthcare Application with Disease Prediction using Probabilistic Neural Networks

Dr. Jagadish S Kallimani[1], Mrs. Chandrika Prasad[2], Abhay[3], Ankit Bargotra[4] and Bhuvan Raj M S[5]
[1]Professor and Head, Department of AIML, Ramaiah Institute of Technology, Bengaluru, India
[2]Assistant Professor, Department of CSE, Ramaiah Institute of Technology, Bengaluru, India
[3-5]Student, Department of CSE, Ramaiah Institute of Technology, Bengaluru, India
Email: jagadish.k@msrit.edu,chandrika@msrit.edu, 1ms19cs001@gmail.com, 1ms19cs019@gmail.com, 1ms19cs036@gmail.com

***Abstract*— Healthcare is a major concern in the current world. Various types of mental as well as physical concerns are emerging day by day. Healthcare Infrastructure needs to be improved specially in developing countries. Medical Science is advancing day by day, but development of proper healthcare infrastructure is also important. Nobody should be deprived of services. Due to the present advancement in medical science, the patients nowadays have a lot of options for proper health care and diagnosis. Sometimes, the patients have to change the doctor or shift from one treatment to another altogether due to various reasons. In such case, there is a need of consolidated central storage where all the records of a patient are stored. There are various health related issues people face which are left unattended due to various personal or professional reasons. The project involves a healthcare application where people can examine their health issues and check whether they are symptoms of some serious health issue or not. The application involves symptom to disease mapping. The classification is done using the Probabilistic Neural Networks. The project brings together the application of web technologies and deep learning. The application involves a user-friendly interface, a chatbot, storage system and chatting options with the assigned healthcare professional. In this project, we focus on creating a web application which has a complete healthcare system involving the initial stage of finding a doctor and later on getting a proper treatment from the healthcare professional. It is a web application so it can be used on various kinds of devices. It focuses on improving the interaction between patients and doctors and work towards improving the healthcare infrastructure.**

***Index Terms*— Probabilistic Neural Network, PNN, classification, disease prediction, symptoms tracking, medical records storage and healthcare.**

I. INTRODUCTION

A. General introduction

Due to the present advancement in medical science, the patients nowadays have a lot of options for proper health care and diagnosis. Sometimes, the patients have to change the doctor or shift from one treatment to another altogether due to various reasons. In such case, there is a need of consolidated central storage where all the records of a patient are stored. Doctors can then access all the previous history of a patient. Both patients and doctors may then also face the problem of recognizing the previous handwritten health records. The project aims

Grenze ID: 01.GIJET.9.2.366

to deliver a web application that is designed for the purpose of creating a platform for the patients and doctors communities It will help people who are concerned about their health and are facing some symptoms to find a suitable doctor online. It will also help doctors to improve their networks and improve their availability to the patients. It is a web application implementing a modular system architecture.
The application will have multiple features. It enables the user to input the symptoms they are facing and get the probable health issue [1] they might have. It will then recommend doctors to the user for booking appointment for consultation [16]. The user can also interact with the doctors [17] using the chatting interface provided by the application. The user is also given the facility to upload his/her medical records [13] on the application for safe storage. The doctors can also view the medical records and history of his/her patients using the application.

B. Objectives of the Project

- To analyze the symptoms of the users and predict the disease using Probabilistic Neural Network.
- To create a consolidated secure storage system for medical records.
- Easy to use interface for patients and doctors.
- Enabling patients find a suitable doctor online.
- Enabling patients to book appointments with the doctors and interact with the doctors.
- To enable record sharing and interaction between patients and doctors.

II. LITERATURE SURVEY

A. Introduction

People today deal with a variety of diseases because of their environmental surroundings their living habits. There are various health related issues people face which are left unattended due to various personal or professional reasons. Therefore, it is very important to predict the disease at earlier stage before it becomes life threatening. But it may be difficult for a doctor to predict the disease on the basis of symptoms. The correct prediction of disease [2] based on the symptoms experienced by the patient is a challenging task.
As a part of literature review, we tried to analyze the work done by various researchers related to disease predictions based on the symptoms. We referred to a number of papers and journals and examined the various methods adopted for the same [5].
A comparison of accuracy achieved on disease prediction based on symptoms using these methods is presented in the next sub section.

B. Related Works

TABLE I. COMPARISON OF VARIOUS EXISTING METHODS FOR DISEASE PREDICTION BASED ON SYMPTOMS.

S. No.	Existing Methods	Accuracy
[1]	Naïve Bayes (NB) Algorithm	94%
[1]	K- Nearest Neighbors (KNN) Algorithm	67%
[3]	LSTM (Long Short-Term Memory)	89%
[4]	Convolutional Neural Network (CNN)	81.13%
[6]	Multilayer Perceptron Neural Network (MLPNN) with Backpropagation algorithm (BP)	90%
[22]	Support Vector Machine (SVM)	94.60%

C. Conclusion of the survey

After examining and doing literature survey on various research papers, it is observed that disease prediction on the basis of symptoms [3] is a challenging task. There are various existing systems that can predict the disease based on the symptoms [4] experienced by the patients. There are various techniques and algorithms that can be used. Some of them are Naïve Bayes algorithm, random forest, k-nearest-neighbor (KNN), decision tree algorithm, neural network models like CNN and RNN, etc.
Most of the existing systems are used for predicting one particular disease, given the symptoms that a patient is experiencing. Among all of the algorithms used for predicting the disease, the KNN algorithm is perhaps the simplest one. In addition, the ability of neural networks to work with insufficient knowledge and their good fault tolerance makes them a good choice. However, they are computationally expensive.

III. DESIGN

We designed a web application for the proposed project. The Figure. 1 gives the basic architectural design of the web application developed.

IV. METHODOLOGY

The web application to be developed for the project can be divided into three parts as discussed in the below three sub sections.

A. Front-end of the application

The application provides a very user friendly and easy to use interface to the users. The users can very easily interact with the application. For designing the user interface, the following technologies are used.

1) *HTML:* Hyper Text Markup Language (HTML). It is the standard markup language which is used for creating web pages. The structure of a web page is described using HTML.

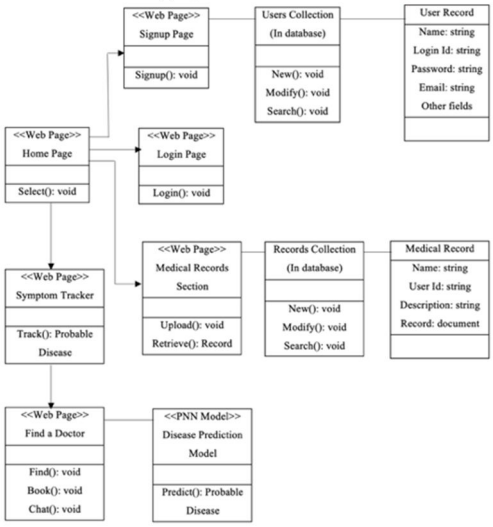

Figure 1. Basic Architectural Design.

2) *CSS:* CSS, or Cascading Style Sheets, specifies how HTML elements should be rendered on a screen. For styling the web pages, CSS is used.
3) *JavaScript:* JavaScript (JS) is a lightweight, interpreted, or just-in-time compiled programming language. For defining the behavior of the web pages and for creating dynamic web pages, JavaScript is used.
4) *Bootstrap:* The most widely used HTML, CSS, and JavaScript framework for creating responsive, mobile-first websites is Bootstrap. For styling and making the web pages responsive, Bootstrap is used.
5) *React:* React is a JavaScript library for building user interfaces. React is used to build single-page applications. React allows us to create reusable UI components. It is an open-source, component-based front end library which is responsible only for the view layer of the application.

B. Back-end of the application

The back-end of the application processes the user's requests and sends back appropriate responses to the user. It also manages the database which organizes the information. For developing the back-end of the application, the following technologies are used in the proposed project.

1. *Node.js:* Node.js is an open source, cross-platform runtime environment for developing server-side and networking applications.
2. *Express:* Express is a minimal and flexible Node.js web application framework that provides mechanisms to write handlers for requests with different HTTP verbs at different URL paths.
3. *MongoDB Database:* MongoDB is an open-source document database and leading NoSQL database. MongoDB is a cross-platform, document-oriented database that provides, high performance, high availability, and easy scalability.
4. *Mongoose:* It manages relationships between data, provides schema validation, and is used to translate between objects in code and the representation of those objects in MongoDB.

C. Training the PNN model for disease prediction based on symptoms

1) Explanation of PNN Algorithm: The algorithm used for disease prediction based on symptoms is PNN [8]. The PNN (Probabilistic Neural Network) algorithm is a non-parametric, supervised classification algorithm that is used to classify data points into different classes based on their features.

Probabilistic Neural Network (PNN) has a specific structure composed of four layers: Input Layer, Pattern Layer, Summation Layer, and Output Layer. The operation of the PNN model on your dataset can be explained as follows in terms of these layers:

i. *Input Layer:* This layer accepts the feature values. In our case, each neuron in the input layer corresponds to a feature in your dataset, such as "itching", "skin_rash", "nodal_skin_eruptions", etc. When a new patient's data is inputted, these features (symptoms) are passed to the neurons in this layer.

ii. *Pattern Layer:* Each neuron in this layer represents a vector of a training example. When a new patient's data comes in, it is compared to every training example. The comparison is performed using a Gaussian kernel function (a radial basis function), which calculates the Euclidean distance between the new input vector and the stored vector. The computed value is then passed through the Gaussian function, resulting in a similarity measure (the output of the pattern layer neuron). The width of the Gaussian function is determined by the sigma parameter defined in the PNN model.

iii. *Summation Layer:* The neurons in this layer sum up the outputs of the Gaussian functions of the pattern layer for each class. This sum is weighted by the prior probability of each class, which is computed during the training phase (using the fit function). The output of each neuron in this layer is a score indicating the likelihood of the new input belonging to the corresponding class.

iv. *Output Layer:* The output layer consists of a single neuron for each class in the dataset. In your case, each neuron would correspond to a unique prognosis like "Peptic ulcer disease", "Chicken pox", etc. Each neuron in this layer corresponds to a neuron in the summation layer and simply passes on the value it receives. The class corresponding to the neuron with the highest score in the output layer is selected as the predicted class for the new input.

2) Information about the implementation of Modules: Figure 2 shows the steps involved in training the PNN model for disease prediction from symptoms.

i. *Data pre-processing:* This module includes steps taken to prepare the medical dataset before feeding it into the PNN model. The initial step involves data cleaning to ensure that the dataset is free from errors and inconsistencies. For ethical reasons and to maintain the integrity of the dataset, certain rows such as specific prognosis were removed. Additionally, irrelevant symptoms that did not contribute significantly to the prognosis were also omitted from the dataset. This careful pruning of the data helps to focus the PNN model on the most relevant features and improves its predictive capabilities.

Following data cleaning, the target variable, which is the diagnosis or prognosis in this case, is encoded into a numerical format to allow the PNN model to process it. After these pre-processing steps, the data is ready to be fed into the PNN model. These tasks are achieved using the pandas library for data manipulation, and scikit-learn for feature selection and normalization. As our dataset is Binary encoded where features are of 0 or 1 there is no need of normalizing.

ii. *PNN class:* The PNN class implemented is the TraditionalPNN class, a customized model that doesn't use scikit-learn directly but follows a similar structure. It accepts a sigma parameter, which is used to adjust the width of the Gaussian kernel. The class has a fit method that calculates class priors based on the target variable, 'prognosis'. The predict method computes probabilities for each class using the Gaussian kernel and returns the top_n probabilities and their corresponding classes.

Training- The PNN model doesn't perform traditional training as it is a type of instance-based learning or non-parametric method. In the fit function, it basically stores the training data (X) and the corresponding labels (y).

Prediction with the PNN [7] model involves using the stored training data and the Gaussian kernel function to calculate the likelihood of each test instance belonging to each class. This is done in the predict function.

In this process, the unique class labels are identified and stored, along with the training data and labels. The priors for each class are also calculated, which represent the proportion of training instances that belong to each class.

In this process, for each test instance x, the Euclidean distances to all training instances are computed. These distances are fed into the Gaussian kernel function to compute the kernel values. The kernel values are then averaged and multiplied by the class priors to calculate the probability of x belonging to each class.

The function then identifies the top n classes with the highest probabilities and stores them in y_pred. It also calculates the relative probabilities of these top n classes and stores them in y_pred_proba.
Therefore, for each test instance, the predict function returns the top n most likely classes, along with their relative probabilities, each instance x would represent a patient, with binary features indicating the presence or absence of certain symptoms. The PNN model would predict the top n most likely diagnoses for each patient, along with the relative probabilities of each diagnosis.

D. Overall Proposed System and flow of user's interaction with the system

Figure 3 represents the sequence diagram of the proposed system. The overall flow of how the user interacts with the application is also depicted in the figure.

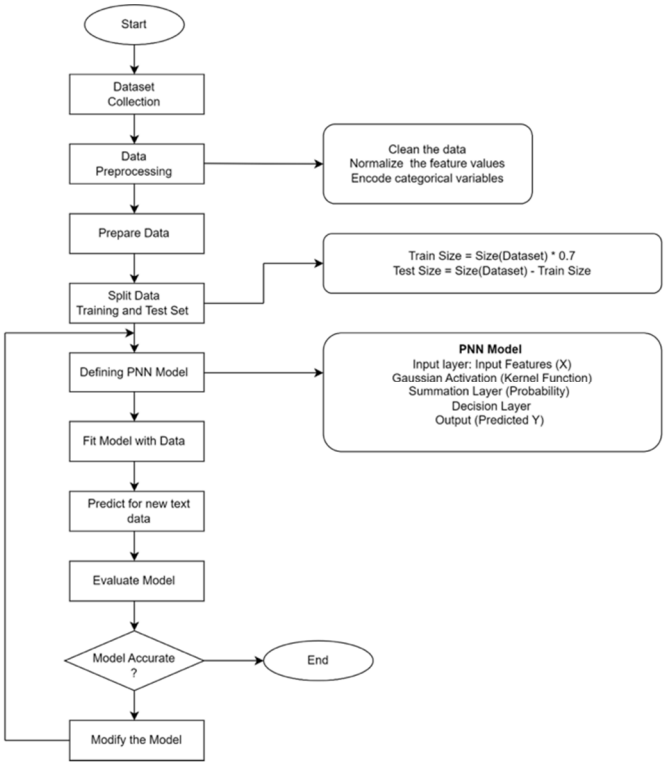

Figure 2. Flow Diagram of training the PNN model for Disease Prediction

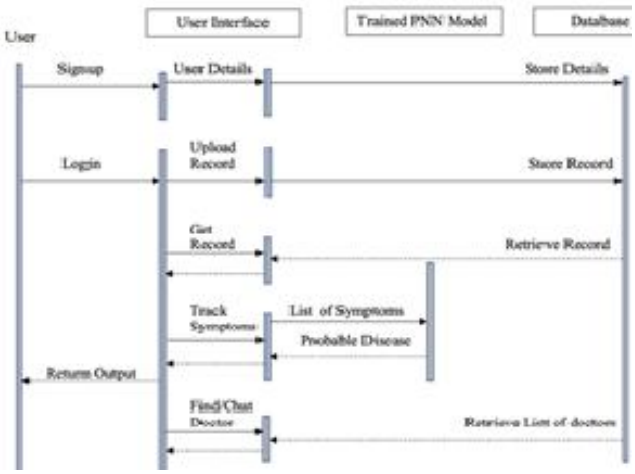

Figure 3. Basic Sequence Diagram of the Proposed System

The user can navigate to the signup page and enter his/her details like name, username, password, etc. and can register on the application. For doctors, they are required to enter their specialization and their clinic address.

Figure 3. Basic Sequence Diagram of the Proposed System. The user enters the login credentials which are validated against the data stored in the database. The authenticated user is then redirected to the home page of the application. The user can then access the features of the application. The user can find a doctor and book appointment. The user can start chatting with the appointed doctors. The user can also view his/her appointments. The user can go the track symptoms page and then enter the symptoms experienced by him/her. Based on the symptoms, the user is displayed the list of probable disease [24] along with the list of recommended doctors. The user may book appointment with the doctor. The user can also navigate to the medical records section. There, the user is asked to enter the OTP sent to the registered email id. Then, the user can upload the medical records. The user can also view the uploaded records. For doctors, they can view the appointments made by the patients with them. They can also access the records of their patients. The user is also provided with the option of updating or editing his/her profile.

V. Results And Snapshots

A. Performance Analysis

The proposed method for disease prediction from symptoms [25] is based on a PNN model. The performance metrics of the proposed method along with their values are shown in the table given below.

TABLE II. Proposed Method's Performance Metrics

S. No.	Performance Metric	Value
1.	Accuracy	97.70 %
2.	Precision	96.61 %
3.	Recall	95.57 %
4.	F1 score	94.16%

B. Snapshots of the web application

1) Home page: The Home page of the web application shows a simple Chatbot asking the user if they are feeling well or not. If the user is not feeling well, the Chatbot provides the redirect link to the user for symptom tracking. Figure 4.
2) Login page: The user can log in to his/her account using the login page by entering his/her username and password. Figure 5.
3) Signup page: The user can create a new account using the signup page by entering the details. Figure 6.
4) Symptom Tracker page: The user enters the symptoms experienced which are then analyzed using the deep learning model [21] and the user is then displayed the list of probable diseases along with the recommended doctors. Figure 7 and Figure 8.
5) Chatting page: The user can interact with the doctors. Figure 9.
6) Find a Doctor page: The user is displayed the list of suitable doctors. The user can filter the doctors based on specialization, city and state. Figure 10.
7) Book an Appointment page: The user can book an appointment with the selected doctor. Figure 11.
8) Upload and View Medical records page: The user can upload and view their medical records. Figure 12 and Figure 13.

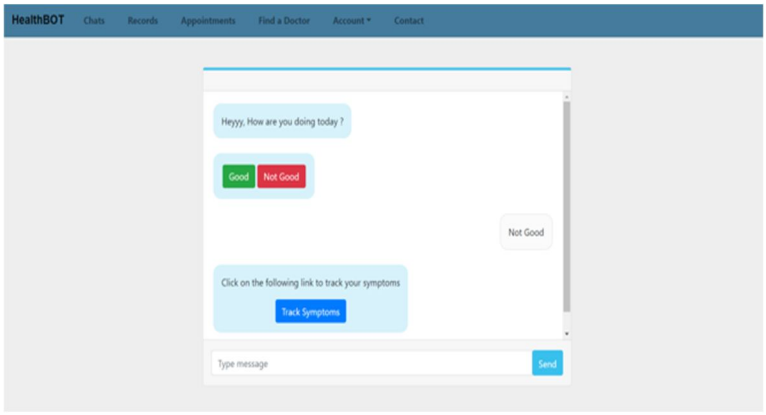

Figure 4. Home Page

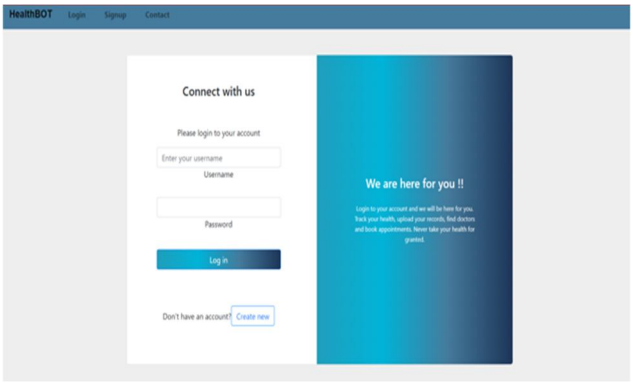

Figure 5. Login Page

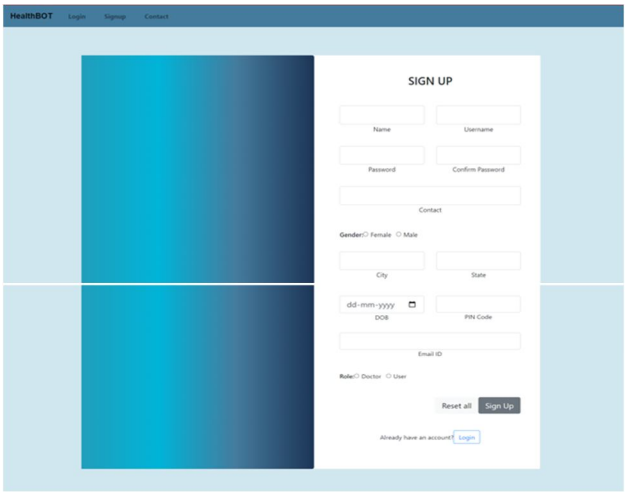

Figure 6. Signup Page

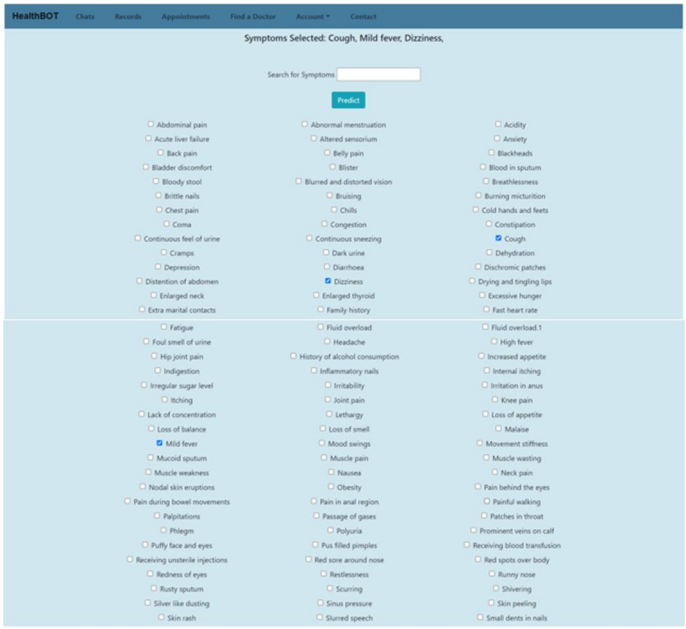

Figure 7. Symptoms Tracker Page

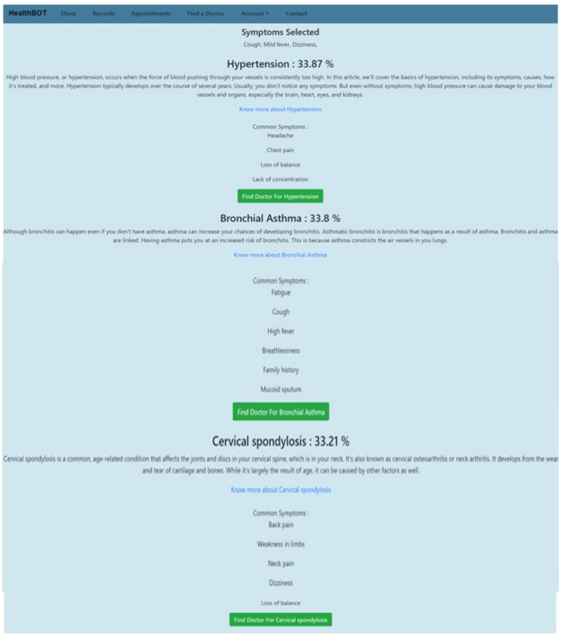

Figure 8. Predicted Diseases Page

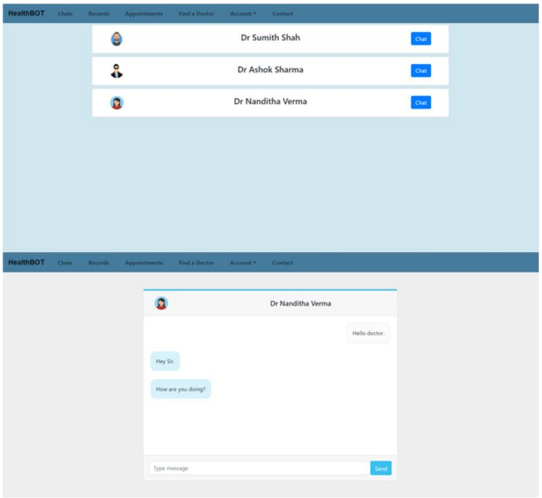

Figure 9. Chatting Page

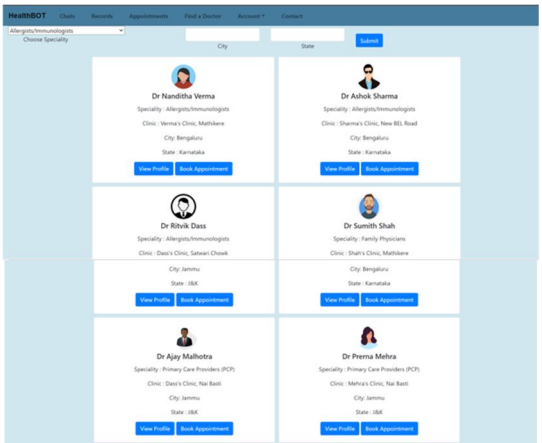

Figure 10. Find a Doctor Page

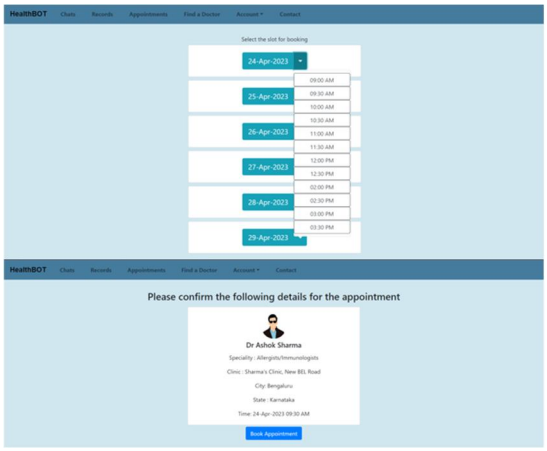

Figure 11. Book an appointment Page

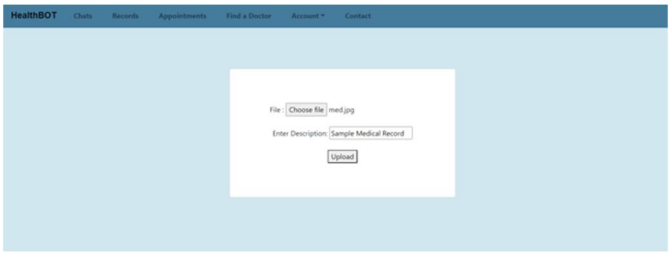

Figure 12. Upload Medical Records Page

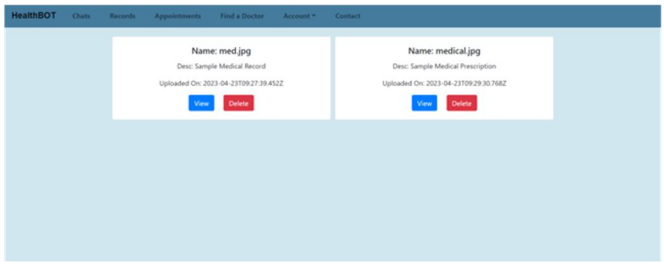

Figure 13. View Medical Records Page

III. Conclusion and Future Work

The web application can be used by people having health issues ranging from mild to serious. This is a multipurpose application. Users can come to learn about various health issues, they can track their health status [22]. This application will improve doctor patient interaction. It will help the doctors improve their availability and will also help the patients to connect with doctors easily sitting at home.

The work plays a significant role in the lives of the people who are having some health issues. Such people can use this web application to interact with the doctors and share their medical records. Also, the web application can also be used by various doctors for checking the health status of a person, their previous medical history which will help them to analyse the patient's status more properly.

As voice-based interactions are more convenient than the text-based interactions, this project can be extended to include the voice-based interactions in the future. Also, as it is AI based disease prediction, it can be improved with improved datasets and algorithms.

Acknowledgement

We would like to express our deepest gratitude and appreciation to MSRIT Alumni Association for their support and generous grant of Rs. 7000 (rupees seven thousand) towards our project.

REFERENCES

[1] Dahiwade, D., Patle, G., & Meshram, E. (2019, March). Designing disease prediction model using machine learning approach. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1211-1215). IEEE.

[2] Wang, T., Tian, Y., & Qiu, R. G. (2019). Long short-term memory recurrent neural networks for multiple diseases risk prediction by leveraging longitudinal medical records. IEEE journal of biomedical and health informatics, 24(8), 2337-2346.

[3] Men, L., Ilk, N., Tang, X., & Liu, Y. (2021). Multi-disease prediction using LSTM recurrent neural networks. Expert Systems with Applications, 177, 114905.

[4] Zhang, X., Zhao, H., Zhang, S., & Li, R. (2019). A novel deep neural network model for multi-label chronic disease prediction. Frontiers in genetics, 10, 351.

[5] Sharmila, S. L., Dharuman, C., & Venkatesan, P. (2017). Disease classification using machine learning algorithms-a comparative study. International Journal of Pure and Applied Mathematics, 114(6), 1-10.

[6] Dangare, C., & Apte, S. (2012). A data mining approach for prediction of heart disease using neural networks. International Journal of Computer Engineering and Technology (IJCET), 3(3).

[7] Ford, W., Xiang, K., Land, W., Congdon, R., Li, Y., & Sadik, O. (2013). A multi-class probabilistic neural network for pathogen classification. Procedia Computer Science, 20, 348-353.

[8] Hasan, S., Yousif, M., & Al-Talib, T. M. (2018). Brain tumor classification using Probabilistic Neural Network. Journal of Fundamental and Applied Sciences, 10(4S), 667-70.

[9] Kshirsagar, P. R., Akojwar, S. G., & Bajaj, N. D. (2018). A hybridised neural network and optimisation algorithms for prediction and classification of neurological disorders. International Journal of Biomedical Engineering and Technology, 28(4), 307-321.

[10] Poddar, M. G., Birajdar, A. C., & Virmani, J. (2019). Automated classification of hypertension and coronary artery disease patients by PNN, KNN, and SVM classifiers using HRV analysis. In Machine learning in Bio-signal analysis and diagnostic imaging (pp. 99-125). Academic Press.

[11] Mantzaris, D., Anastassopoulos, G., & Adamopoulos, A. (2011). Genetic algorithm pruning of probabilistic neural networks in medical disease estimation. Neural Networks, 24(8), 831-835.

[12] Laskar, M. T. R., Hossain, M. T., Kamal, A. R. M., & Rashid, N. (2016). Automated disease prediction system (ADPS): a user input-based reliable architecture for disease prediction. International Journal of Computer Applications, 133(15), 24-29.

[13] Wong, R. E. X., & Bradley, E. H. (2009). Developing patient registration and medical records management system in Ethiopia. International journal for quality in health care, 21(4), 253-258.

[14] Gummerus, J., Liljander, V., Pura, M., & Van Riel, A. (2004). Customer loyalty to content-based Web sites: the case of an online health-care service. Journal of services Marketing, 18(3), 175-186.

[15] De Witte, N. A., Carlbring, P., Etzelmueller, A., Nordgreen, T., Karekla, M., Haddouk, L., ... & Van Daele, T. (2021). Online consultations in mental healthcare during the COVID-19 outbreak: An international survey study on professionals' motivations and perceived barriers. Internet Interventions, 25, 100405.

[16] Xing, W., Hsu, P. Y., Chang, Y. W., & Shiau, W. L. (2020). How does online doctor–patient interaction affect online consultation and offline medical treatment?. Industrial Management & Data Systems, 120(1), 196-214.

[17] Singh, A. P., Joshi, H. S., Singh, A., Agarwal, M., & Kaur, P. (2018). Online medical consultation: A review. International Journal of Community Medicine and Public Health, 5(4), 1230-1232.

[18] Noergaard, B., Sandvei, M., Rottmann, N., Johannessen, H., Wiil, U., Schmidt, T., & Pedersen, S. S. (2017). Development of a web-based health care intervention for patients with heart disease: lessons learned from a participatory design study. JMIR research protocols, 6(5), e7084.

[19] Powell, J. A., Darvell, M., & Gray, J. A. M. (2003). The doctor, the patient and the world-wide web: how the internet is changing healthcare. Journal of the royal society of medicine, 96(2), 74-76.

[20] Clark, E. J. (2002). Health care web sites: are they reliable?. Journal of medical systems, 26, 519-528.

[21] Pant, K., Bordoloi, D., & Pant, B. (2021). Developing a Machine Learning-Based Framework for Disease Prediction. Webology, 18(5), 3132-3139.

[22] Hamsagayathri, P., & Vigneshwaran, S. (2021, February). Symptoms based disease prediction using machine learning techniques. In 2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV) (pp. 747-752). IEEE.

[23] Mantzaris, D. H., Anastassopoulos, G. C., & Lymberopoulos, D. K. (2008, October). Medical disease prediction using artificial neural networks. In 2008 8th IEEE International Conference on BioInformatics and BioEngineering (pp. 1-6). IEEE.

[24] Dahiwade, D., Patle, G., & Meshram, E. (2019, March). Designing disease prediction model using machine learning approach. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1211-1215). IEEE.

[25] Bhanuteja, T., Kumar, K. V. N., Poornachand, K. S., Ashish, C., & Anudeep, P. Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN, 2278-3075.