Disease Prediction Based On Image Processing Using Chatbot - MEDIKNOW

Mrs.K.Papithasri[1], Bakavath.K[2], Dhayanandham.D.E[3], Jabarson Richard.G.S[4], Jeeva.R[5]

Author 1 : Assistant Professor, Dept of CSE, SNS College of Technology, Coimbatore - 641035
Author 2, Student, Dept of CSE, SNS College of Technology, Coimbatore- 641035
Author 3 : Student, Dept of CSE, SNS College of Technology, Coimbatore - 641035
Author 4 : Student, Dept of CSE, SNS College of Technology, Coimbatore - 641035
Author 5 : Student, Dept of CSE, SNS College of Technology, Coimbatore – 641035

Abstract: Hospitals are the most common way for patients to receive physical examinations, diagnoses, and medical advice. This has become the practice of almost all people in the world. It is considered the most reliable method of health examination. The system aims to create an alternative to today's method of going to the hospital and making an appointment with the doctor for diagnosis. The proposed system uses natural language processing and machine learning concepts to create chatbot applications. People can interact with the chatbot as well as with other people, and the chatbot will analyze the user's symptoms through questions to predict the disease and recommend treatment pain. This system can help people perform daily health checks, enable people to understand their health, and encourage people to take appropriate measures to stay healthy. According to research, such systems are not widely used and little is known about them. Following this principle can help people avoid unnecessary visits to the hospital by using this free app wherever they are.

Key Words : Disease prediction, Treatment recommendations, Chatbot, Symptom analysis, Machine learning, Natural Language Processing.

1. INTRODUCTION

Today, the medical and healthcare industries are increasingly recognizing the imperative for enhanced data mining capabilities. Strategic utilization of data holds the potential to empower healthcare professionals, enabling them to make early decisions and enhance treatment outcomes by distilling crucial information from vast datasets. Spirit, our innovative system, employs advanced distribution techniques to assist doctors in this data-driven endeavor.

Beyond the general utility of data mining, Spirit specializes in addressing specific diseases and health issues such as malaria, dengue fever, impetigo, diabetes, migraine, jaundice, measles, and more. Acknowledging the profound impact of these conditions on individuals' eating habits and, in extreme cases, their lives, Spirit strives to extract pertinent insights for proactive healthcare interventions.

Recognizing the intricate relationship between occupational health and extensive data repositories, Spirit employs a probing approach to unveil patterns and relationships concealed within these voluminous datasets. Data mining algorithms, including decision trees, random forests, and naive Bayes, emerge as formidable tools in deciphering complex health dynamics.

In response to the pressing need for comprehensive virus-related information, we have developed a cutting-edge system. This system not only detects but also extracts latent virus-related data from historical records, encompassing symptoms and adhering to established procedural rules. This forward-looking approach ensures that healthcare

professionals can glean valuable insights, ultimately contributing to more effective diagnosis and treatment strategies.

2. OVERVIEW

There are 132 symptoms, combinations, or changes in the data we considered that cause 41 diseases. Based on the records of 4920 patients, we aim to create a prediction model that uses the user's symptoms and predicts the diseases he will have more.

The considered symptoms are:

TABLE I: SYMPTOMS

Symptoms		
Back pain	Bloody stool	scurrying
Constipation	depression	Passage of gases
Abdominal pain	Irritation in anus	Weakness in limbs
Diarrhea	Neck pain	Fast heart rate
Mild fever	dizziness	Internal itching
Yellow urine	cramps	Toxic look
Yellowing of eyes	bruising	palpitations
Acute liver failure	obesity	Painful walking
Fluid overload	Swollen legs	Prominent veins on calf
Swelling of stomach	irritability	Fluid overload
Swelled lymph nodes	Swollen blood vessels	Excessive hunger
malaise	Muscle pain	Black heads
Blurred and distorted vision	Pain in anal region	Pain during bowel movements
phlegm	Brittle nails	Rusty sputum
Throat irritation	Belly pain	Mucoid sputum
Redness of eyes	Enlarged thyroid	Puffy face and eyes

The diseases considered are:

TABLE II: DISEASES

Diseases		
Fungal Infection	Malaria	Varicose veins
Allergy	Chickenpox	Hypothyroidism
Gerd	Dengue	Vertigo
Chronic cholestasis	Peptic ulcer disease	Acne
Drug reaction	Hepatitis A	Urinary tract infection
Piles	Hepatitis B	Psoriasis
AIDS	Hepatitis C	Impetigo
Diabetes	Hepatitis D	Hyperthyroidism
Gastroenteritis	Hepatitis E	Hypoglycemia
Bronchial Asthma	Alcoholic hepatitis	Cervical Spondylosis
Hypertension	Tuberculosis	Arthritis
Migraine	Common cold	Osteoarthritis
Paralysis	Pneumonia	Typhoid

The generalized prediction model can be given as:

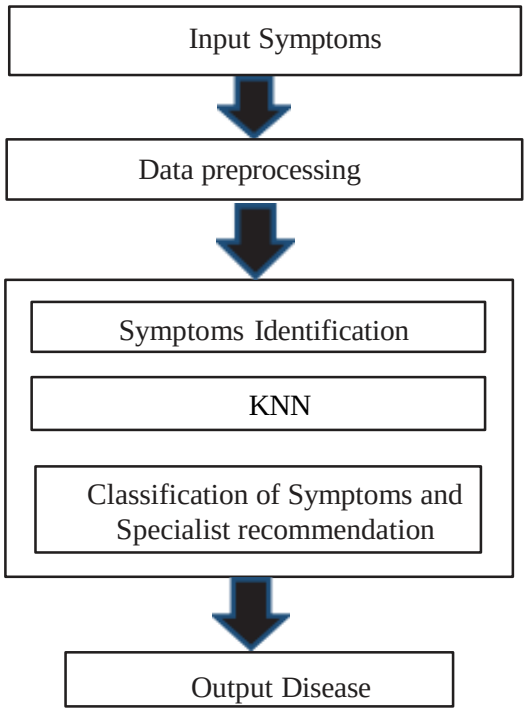

FIG 1: PREDICTION MODEL

A. Input (Symptoms)

When creating the model, we think that the user clearly understands the symptoms he is experiencing. The generated predictions take into account 95 symptoms for which users can follow the instructions.

B. Data Preprocessing

The data search process that transforms raw data or encodes data into a form that can be easily interpreted by algorithms, called data preprocessing. The main techniques used in this study are:

• Data cleaning: Clean the data by methods such as filling in missing values to eliminate inconsistencies in the data.
• Data reduction: Analysis becomes difficult when processing big data. Therefore, we eliminate independent variables (symptoms) that have little or no effect on the target variable (disease). In this study, 95 of 132 symptoms associated with the disease were selected.

C. Output (Diseases)

Once the system is trained using the training method using the above method, the system will be created. The rule is defined so that the resulting disease can be classified and predicted.

3. LITERATURE SURVEY

Developments in technology have also deeply affected healthcare services. Machines not only assist doctors but also provide first-hand tests to patients.

[1] Chatbot uses machine learning for disease prediction and treatment recommendations – Published in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI): Recommended process, hospital Create an alternative way to go and make an appointment with the doctor for diagnosis. People can interact with the chatbot as well as with other people by asking questions from the chatbot. The chatbot will analyze the user's symptoms to predict the disease and recommend treatment. [2] Medbot: Interactive AI-powered chatbot for telemedicine delivery post-COVID-19 - 2020 5th International Conference on Communications and Electronic Systems (ICCES) report: During the recent Coronavirus, Doctors can use telemedicine to connect with patients COVID-19 between patients and doctors While trying to reduce the spread of -19. Telemedicine can be helpful during the pandemic, using AI-based communication for treatment, allowing patients to receive support without having to visit the hospital. Therefore, telemedicine will quickly turn personal care into a conversation between patients. That's why he created a multilingual interactive program based on natural language processing (NLP) to provide free health education, information and advice to patients with chronic diseases. This study introduces a new computer application that mimics the doctor's personality, carefully designed and further learned to interact with patients like humans. Based on serverless architecture, the application combines the services of doctors by providing prevention, home treatment, consultation, health tips and symptoms covering the most common diseases in rural India.
This article introduces "Aapka" "Chikitsak" chatbot on Google Cloud Platform (GCP) for providing telemedicine services in India so that patients can access and benefit from health information. Artificial intelligence has the ability to connect products and different demands from human providers. This communication app reduces hospital treatment interruptions, enables remote consultation, provides timely care and quality treatment, and helps people well. [3] Medical Chatbots - International Journal of Computer Trends and Technology, June 2018: Many users do not know all the treatments or symptoms of a particular disease. For minor problems, users need to go to the hospital for self-diagnosis, which takes more time. Also, dealing with complaint calls is quite intense. Such problems can be solved by using medical chatbots to provide appropriate health advice. Medical chatbot functionality relies on language processing to help users submit health-

related questions. Users can ask personal health-related questions via the chatbot without having to go to the clinic in person. through use Google API for text-to-speech and text-to-speech. Questions are sent to the Chatbot and relevant answers are received and displayed in the Android application. The basic principle in the design of this website is the determination of customer needs.[4] In recent years, healthcare services have become increasingly accessible to people thanks to technology. The concepts of artificial intelligence, machine learning and neural networks have brought many benefits in medicine. In the rapidly developing world, people tend to neglect their health, and this can lead to serious problems. Such problems can be prevented by using symptom-focused predictive applications. Our program is dedicated to providing users with instant and accurate disease predictions based on their symptoms and detailed analysis of their disease reports. The virus prediction chatbot is built using natural language processing and machine learning algorithms. We use two classification algorithms for disease prediction: decision tree and KNN (k nearest neighbors). Compare the performance of these methods and choose the best model based on its accuracy. According to our results, the accuracy of the decision tree and KNN is 92.6% and 95.74%, respectively. The program also follows medical recommendations for prescribed diseases. Pathology report analysis is done using the concept of character recognition (OCR). Tesseract is an open- source recognition engine for OCR processing. The printed text of the report is used to provide more convenient interpretation of results and graphical analysis of test results.

4. PROBLEM STATEMENT

Hospitals are the most common way for patients to receive physical examinations, diagnoses, and medical advice. This has become the practice of almost everyone in the world. It is considered the most reliable method of health examination. Often users do not know all the treatments or symptoms of a particular disease. For minor problems, users have to go to the hospital for self-diagnosis, which takes a long time. That's why predictive machine learning has emerged recently as an effort to reduce the physician's work and improve the overall capabilities of machine learning for healthcare.

5. PROPOSED SYSTEM

In our plan, users can talk to the bot about questions via voice or text. The system uses expert techniques to answer questions. Users can also view doctors who are available for the disease in question. The system can be used by many users to conduct online meetings. Chatbot data is stored in files in the form of schema templates. The robot will give you antibiotic and food recommendations, meaning what foods you should eat based on the disease.

5.1 Advantages of the proposed system

- Save on medical costs
- Save time
- Do not go to hospital for minor problems
- Show to specialist

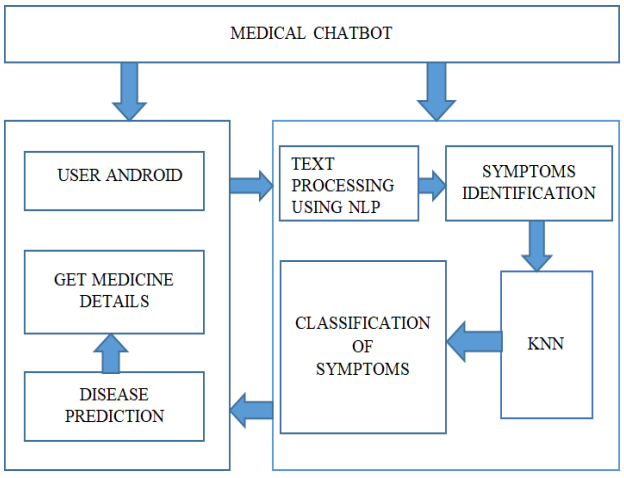

FIG 2: WORKING MODEL OF THE CHATBOT

6. IMPLEMENTATION METHODOLOGY

6.1 DATASET

Data on various diseases were retrieved from Kaggle. Some disorders of the liver and malaria can be seen in figures 3 and 4. The model also treats diabetes, lung disease, cancer, cold, flu, etc. It also includes many other diseases.

	A	B	C	D	E	F	G	H	I	J	K
1	Age	Gender	Total_Bilirı	Direct_Bil	Alkaline_I	Alamine_A	Aspartate	Total_Pro	Albumin	Albumin_	Dataset
2	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	1
3	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
4	62	Male	7.3	4.1	490	60	68	7	3.3	0.89	1
5	58	Male	1	0.4	182	14	20	6.8	3.4	1	1

FIG 3: DATASET FOR LIVER PATIENTS

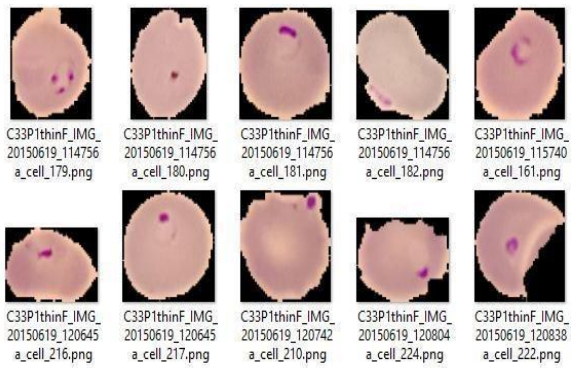

FIG 4: DATASET FOR MALARIA PATIENTS

6.2 BLOCK DIAGRAM

Fig-5 shows how the attributes will function using a diagrammatic representation.

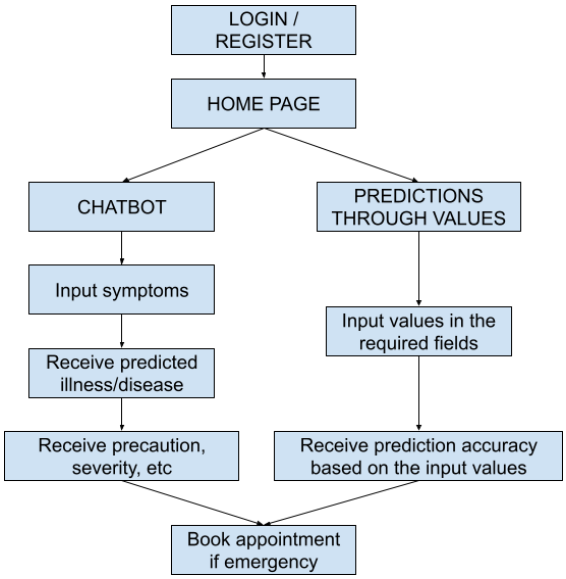

FIG 5: BLOCK DIAGRAM

6.3 MACHINE LEARNING TECHNIQUES

Decision Trees:

The decision tree algorithm is a tree-structured classifier that can use classification and iteration. It has a root node that defines the attributes of the given data and pages that define the output. The decision tree algorithm was chosen for this task because it simulates human reasoning ability and is easy to understand.

Bidding starts from the root of the decision tree and compares the weight of the value or underlying value. In comparison, it continues to the next node down the branch. The algorithm continues comparing the value of the feature with other child nodes and moves down the tree.The decision tree algorithm is a tree-structured classifier that can use classification and iteration. It has a root node that defines the attributes of the given data and pages that define the output. The decision tree algorithm was chosen for this task because it simulates human reasoning ability and is easy to understand.

Bidding starts from the root of the decision tree and compares the weight of the value or underlying

value. In comparison, it continues down the branch to the next node. The algorithm continues comparing the value of the feature with other child nodes and moves down the tree.

Logistic Regression:

The method of modeling the probability of a random event given different inputs is called logistic regression. Binary logistic regression models are true or false, yes or no, etc. It has dual consequences that can occur. Logistic regression is a simple analytical tool used to determine whether a new model best fits certain categories in a distribution.

Natural Language Processing:

Natural Language Processing (NLP) is a field focused on enabling computer algorithms to understand human language. Natural Language Tools (NLTK) is a Python package for NLP. The Natural Language Toolkit (NLTK) is a collection of Python packages specifically designed to detect and record spoken language (such as English) in text.

Neural Networks:

Neural networks are a form of artificial intelligence used to teach computers to analyze data inspired by the human brain. It is a form of deep learning that uses interconnected networks, or neurons, in a hierarchical structure similar to our brain. It creates an adaptable system that computers can use to learn from their mistakes and continue to improve.

7. CONCLUSION

Our project aims to be a medical chat bot that can be used to replace the traditional method of disease diagnosis and treatment recommendation. A chat bot can act as a doctor. A chatbot acts as a user application. The user of this app can mention their symptoms to the chat bot and in turn, the chat bot will mention the health measures to take. General information about symptoms and diseases is available in the dataset and thus a chat bot example can provide the user with information about the disease and treatment. We believe that this approach incorporated into existing strategies in the field of healthcare will provide support to healthcare professionals and patients.

As new variants of the Covid-19 virus emerge, fear for the safety of their lives is growing among people. Although various standards and regulations are being implemented, this fact should not be neglected. Even when appropriate precautions are followed, people are still afraid to go to hospitals and risk getting infected. This work will help create a healthy and safe environment without having to go through the fear and risk of contagion. People who find it difficult to afford hospital bill scanning will also benefit from this project.

Our model was designed to work seamlessly on the aforementioned features such as website traffic, input prediction and chatbot implementation. The accuracy achieved by the chatbot model is about 95.52%. For the future scope, we want to improve the functionality of our chatbots by adding features like translation so that non-English speaking people can easily use the chatbot. We would also like to add a chatbot for mental health care.

8. REFERENCES

[1] Murali Krishna Ogirala, Rahul Tallapaneni, Sai Meghana Chalamcharla, Anuradha Chinta, “A Medical Diagnosis and Treatment Recommendation Chatbot using MLP” Published in 2nd International Conference 4 May 2023.

[2] Nazlı Tokatli, Muhammed Tayyip Kocak, Seda Kirtay, Gurkan Goztepeli, İbrahim Serhat Aktas, Halis Altun, “Healthcare Service Accessibility Path Planner: Unveiling a New Era of Intelligent Appointment Management Systems Based on Outpatient Prioritizing” Published in Innovations in Intelligent 11 October 2023,

[3] Sanjay Chakraborty, Hrithik Paul, Sayani Ghatak, S. Pandey, Ankit Kumar, K. Singh, M. Shah, “An AI-

Based Medical Chatbot Model for Infectious Disease Prediction," Published in IEEE Access 2022, Computer Science.

[4] M. Polignano, F. Narducci, A. Iovine, C. Musto, M. DeGemmis and G. Semeraro, "Health Assistant Bot: A Personal Health Assistant for the Italian Language" in IEEE Access, vol. 8, pp. 107479-107497,2020

[5] Mohnal Gupta, Arunkumar.B, Tejaswani.V.B, Naveen.K.P, "Design of Smart Chatbot for Medical Application using Natural Language Processing", Published at [IRJET] Volume: 08 Issue: 06 on June 2021

[6] Lee, H, Kang, J, & Yeo, J (2021). Medical Specialty Recommendations by an Artificial Intelligence Chatbot on a Smartphone: Development and Deployment. Journal of medical Internet research,

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989. X. Ren, G. Spina, S. DeVries, A. Bijkerk, B. Faber and A. Geraedts, "Understanding Physician's Experience With Conversational Interfaces During Occupational Health Consultation" in IEEE Access, vol. 8, pp. 119158-119169

[8] R. Oruche et al. "Evidence -Based Recommender Systemfor a COVID-19 Publication Analytics Service" in IEEE Access, vol. 9, pp. 79400-79415, 2021

[9] Dr. V Sheeja Kumari, Beksy S George, Nisha Varghese, Sherin Mary Mathew, Telma Elza Ninan, "HUMICARE-A CHATBOT FOR HEALTHCARE SYSTEM", Published at International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 05 on May 2021

[10] T.M. M. Hayir, M.A. Magan, L.M. Mohamed, M.A. Mohamud and A.A. Muse, "Barriers for full immunization coverage among under 5 years children in Mogadishu, Somalia." Journal Family Med Prim Care, vol. 10, no. 11, pp. 2664-2673, 2020

[11] G. Battineni, N. Chintalapudi, and F. Amenta, "AI chatbot design during an Epidemic like the Novel Coronavirus," Healthcare, vol. 8, no. 2,p. 154, 2020.

[12] Lokman, A.S.; Ameedeen, M.A. Modern Chatbot Systems: A Technical Review. In Proceedings of the Future Technologies Conference (FTC) 2018; Arai, K., Bhatia, R.,Kapoor, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 1012 – 1023

[13] Kavitha, B. R., and Chethana R. Murthy.' Chatbot for healthcare system using Artificial Intelligence' Int J Adv Res Ideas Innov Technol 5 (2019): 1304-1307

[14] Linder, Courtney. The Effects of a Healthcare Chatbots; Language and Persona on User Trust, Satisfaction, and Chatbot Effectiveness. Diss. Clemson University, 2020.

[15] Yu, Hong. (2019). Experimental Disease Prediction Research on Combining Natural Language Processing and Machine Learning .145 1500.1109/ICCSNT47585.2019.8962507.

[16] S. P. Erdeniz, I. Maglogiannis, A. Menychtas, A. Felfernig,T. Ngoc, and T. Tran, "Recommender systems for IoT enabled-health applications," in Proceedings of the AIAI 2018: Artificial Intelligence Applications and Innovations, pp. 227–237, Springer, Cham, May 2018

[17] Dimple veer Singh, Sumit Malhotra (2018). Intra News Category Classification using N-gram TFIDF Features and Decision Tree Classifier, IJSART