VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JNANA SANGAMA", MACHHE, BELAGAVI – 590018

Mini Project Report
on

Doctor Bot

Submitted in partial fulfillment of the requirements for the V semester
Bachelor of Engineering
in
Computer Science and Engineering
of
Visvesvaraya Technological University, Belagavi
By

Mr. C N V C Sri Harsha(1CD21CS034)
Mr. Darshan B Pattar (1CD21CS040)
Mr. Nandheeswaran M (1CD21CS095)
Mr. Chandra Mouli B N (1CD21CS037)

Under the Guidance of

Ms. Maria Kiran L
Assistant Professor,
Dept of CSE,CITech

CAMBRIDGE
INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering
CAMBRIDGE INSTITUTE OF TECHNOLOGY, BENGALURU – 560036

2023-2024

CAMBRIDGE INSTITUTE OF TECHNOLOGY

K.R. Puram, Bengaluru – 560036

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CAMBRIDGE
INSTITUTE OF TECHNOLOGY

CERTIFICATE

Certified that **Mr. C N V C Sri Harsha, Mr. Darshan B Pattar, Mr. Nandheeswaran M** and **Mr. Chandra Mouli B N** bearing USN **1CD21CS034, 1CD21CS040, 1CD21CS095, 1CD21CS0037** , respectively are bonafide students of **Cambridge Institute of Technology,** has successfully completed Mini Project entitled **"Doctor Bot"** in partial fulfillment of the requirements for VI semester **Bachelor of Engineering** in **Computer Science and Engineering** of **Visvesvaraya Technological University, Belagavi** during academic year 2023 - 2024. It is certified that all Corrections / Suggestions indicated for Internal Assessment have been incorporated in the report.

Internal Guide
Prof. Maria Kiran L
Dept. of CSE, CITech

Head of the Dept.
Dr. Shreekanth M Prabhu
Dept. of CSE, CITech

DECLARATION

We, **Mr. C N V C Sri Harsha, Mr. Darshan B Pattar, Mr. Nandheeswaran M and Mr. Chandra Mouli B N** bearing USN **1CD21CS034, 1CD21CS040, 1CD21CS095, 1CD21CS037** respectively, are the students of VI semester, Computer Science and Engineering, Cambridge Institute of Technology, hereby declare that the mini-project entitled **"DOCTOR BOT"** has been carried out by us and submitted in partial fulfillment of the course requirements of VI semester **Bachelor of Engineering in Computer Science and Engineering** as prescribed by **Visvesvaraya Technological University, Belagavi**, during the academic year 2023-2024.

We also declare that, to the best of our knowledge and belief, the work reported here does not form part of any other report on the basis of which a degree or award was conferred on an earlier occasion on this by any other student.

Date :

Place : Bangalore

NAME	USN	SIGNATURE
C N V C SRI HARSHA	1CD21CS034	Sriharsha
DARSHAN B PATTAR	1CD21CS040	Darshan B. Pattar
NANDHEESWARAN M	1CD21CS095	M. [illegible]
CHANDRA MOULI B N	1CD21CS037	Chandra Mouli

ACKNOWLEDGEMENT

We are extremely thankful to **Dr. Indumathi G,** Principal, CITech., Bengaluru, for providing us the academic ambience and everlasting motivation to carry out this work and shaping our careers.

We express our sincere gratitude to **Dr.Shreekanth M Prabhu,** HOD, Dept. of Computer Science and Engineering, CITech., Bengaluru, for his stimulating guidance, continuous encouragement and motivation throughout the course of present work.

We also wish to extend our thanks to **Prof. Maria Kiran L,** Internal Guide, Dept. of Computer Science and Engineering, CITech., Bengaluru, for their expert guidance and constructive suggestions to improve the quality of this work.

We would also like to thank all other teaching and technical staffs of Department of Computer Science and Engineering, who have directly or indirectly helped us in the completion of this Project Work.

And lastly we would hereby acknowledge and thank our parents who have been a source of inspiration and also instrumental in the successful completion of this project.

Mr. CNVC Sri Harsha (1CD21CS034)

Mr. Darshan B Pattar (1CD21CS040)

Mr. Nandheeswaran M (1CD21CS095)

Mr. Chandra Mouli B N (1CD21CS037)

ABSTRACT

In the realm of medical diagnosis and treatment, machine learning models are trained using a vast array of patient data, encompassing demographic information, clinical histories, diagnostic results, and treatment outcomes. These models leverage algorithms to identify patterns and correlations within the data, which may not be immediately evident to human practitioners. For instance, by analyzing large datasets of medical images, machine learning algorithms can learn to detect anomalies and classify conditions with high accuracy. The process begins with data preprocessing, where the raw data is cleaned, normalized, and transformed into a suitable format. Following this, feature extraction is performed to identify and select the most relevant attributes that contribute to the prediction task. The model is then trained using supervised learning, where it learns from labelled examples, or unsupervised learning, where it seeks to find inherent structures in the data without predefined labels. Once trained, the model is validated and tested to ensure its reliability and accuracy in real-world scenarios. The ultimate goal is to augment medical decision-making, enhance diagnostic precision, and personalize treatment plans, thereby improving patient outcomes and the efficiency of healthcare delivery.

CONTENTS

List of Figures

CHAPTER 1

INTRODUCTION

Training a machine learning model involves a systematic approach to teaching a computer algorithm to recognize patterns and make decisions based on data. The process begins with data collection, where relevant information is gathered from diverse sources such as electronic health records, medical imaging scans, or wearable devices. Once collected, the data undergoes preprocessing steps to ensure quality and uniformity, including cleaning to remove errors and inconsistencies, and normalization to standardize numerical values. Feature engineering follows, where meaningful features are extracted or selected to represent the data effectively for the model.

Next, a suitable machine learning algorithm is chosen based on the nature of the problem and the characteristics of the data. Algorithms range from traditional methods like linear regression and decision trees to more advanced techniques such as support vector machines and neural networks. The selected algorithm is trained using a portion of the data, known as the training set, where it learns to recognize patterns and relationships. During training, the model adjusts its internal parameters iteratively to minimize errors and improve performance.

Validation is crucial to assess how well the trained model generalizes to new, unseen data. This involves evaluating its performance on a separate validation set, tuning hyperparameters to optimize performance without overfitting to the training data. Ethical considerations, such as ensuring patient privacy and data security, are integral throughout the process, with adherence to regulatory guidelines like HIPAA in healthcare settings.

Continuous evaluation and refinement are essential as models are deployed in real-world applications, monitoring performance metrics and updating models with new data to maintain accuracy and relevance. In conclusion, training a machine learning model is a dynamic process that integrates data science expertise, domain knowledge, and ethical principles to harness the power of data for predictive insights and decision support across various fields, including healthcare, finance, and beyond.

1.1 Background

Machine learning (ML) in healthcare involves leveraging computational algorithms to analyse complex medical data, aiming to extract meaningful insights for clinical decision-making and patient care. The process of training a machine learning model in healthcare begins with defining the problem statement. This could range from predicting patient outcomes based on electronic health records (EHRs) to diagnosing diseases from medical imaging data.

Next, data acquisition and preprocessing play crucial roles. Medical data often come from heterogeneous sources such as hospitals, clinics, and research studies, necessitating careful curation to ensure quality and consistency. Preprocessing steps involve cleaning the data, handling missing values, and standardizing formats to make it suitable for modelling. Feature selection or extraction follows, where relevant information from the data is identified and transformed into numerical or categorical features that can be processed by machine learning algorithms.

Choosing an appropriate algorithm depends on the specific problem and data characteristics. Commonly used algorithms in healthcare include logistic regression for binary classification tasks, decision trees for interpretable insights, and deep learning models like convolutional neural networks (CNNs) for image analysis tasks. The selected algorithm is trained using labelled data, typically split into training and validation sets to evaluate performance and prevent overfitting. Model performance metrics such as accuracy, precision, recall, and F1-score are used to assess the model's effectiveness in meeting the defined objectives.

Ethical considerations, such as patient privacy and data security, are paramount in healthcare ML. Adhering to regulatory guidelines like HIPAA ensures that patient information is handled responsibly and ethically throughout the data lifecycle. Continuous evaluation and refinement of the model are essential to account for changes in patient demographics, evolving medical practices, and technological advancements.

Training machine learning models for medical applications requires a multidisciplinary approach integrating clinical expertise, data science methodologies, and ethical considerations. The ultimate goal is to harness the power of data-driven insights to improve healthcare delivery, enhance diagnostic accuracy, and personalize treatment strategies for better patient outcomes.

1.2 Brief history of Technology and Concept

1.2.1 Decision Tree

The concept of decision trees in machine learning traces back to the early 1960s, evolving from research in artificial intelligence and decision theory. Initially conceptualized as a graphical representation of decision-making processes, decision trees gained traction as a powerful predictive modeling technique. The seminal work of J. Ross Quinlan in the 1980s led to the development of practical algorithms such as ID3 (Iterative Dichotomies 3) and its successors like C4.5, which automated the construction of decision trees from data. Decision trees operate by recursively partitioning the dataset into subsets based on the most significant attributes, optimizing for purity or homogeneity in each partition.

This hierarchical structure of nodes and branches allows for interpretable rules that can be easily understood and validated by domain experts. Decision trees are versatile, capable of handling both classification and regression tasks, and are robust to outliers and missing values. However, they can be prone to overfitting, particularly with complex datasets, which has spurred research into ensemble methods like Random Forests and Gradient Boosting Machines (GBMs) to enhance predictive accuracy and generalizability. Despite advancements in deep learning and neural networks, decision trees remain foundational in machine learning due to their transparency, ease of implementation, and ability to provide actionable insights in diverse domains including healthcare, finance, and marketing.

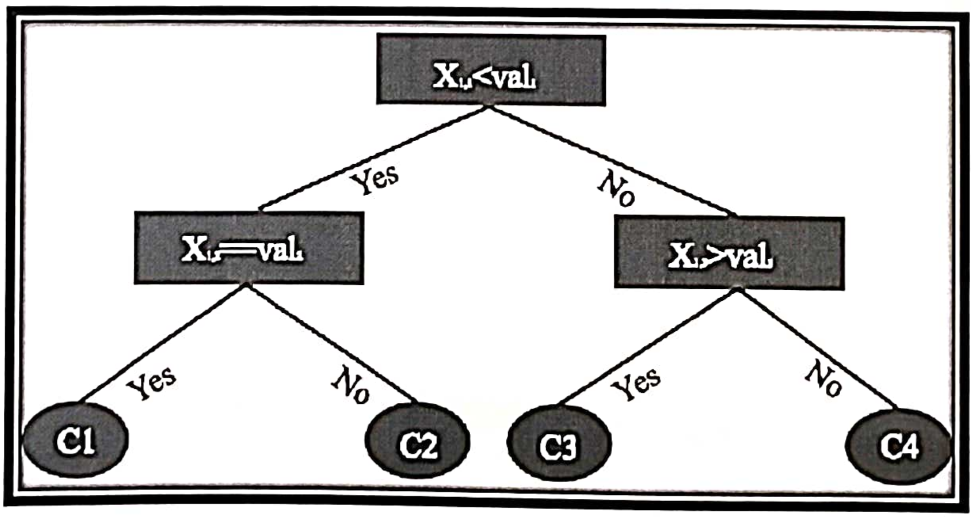


Figure 1.1: Decision Tree Architecture

1.2.2 KNN (K-Nearest Neighbors)

Machine learning has evolved significantly since its inception, rooted in the early development of artificial intelligence (AI) in the 1950s. Initially focused on symbolic approaches, which involved manually programming rules for decision-making, machine learning shifted towards statistical methods in the 1980s. This marked the beginning of a more data-centric approach, where algorithms could autonomously learn patterns from data rather than relying solely on explicit instructions.

One fundamental concept in machine learning is the K-Nearest Neighbors (KNN) algorithm, dating back to the 1950s but gaining popularity in the 1960s and 1970s. KNN is a simple yet effective supervised learning algorithm used for classification and regression tasks. Its principle revolves around the idea of similarity: it assumes that similar things exist in close proximity. In classification tasks, when given a new, unlabeled observation, KNN looks at the labeled observations nearest to it (in terms of distance metrics like Euclidean distance) and assigns the majority label among its K nearest neighbors. For regression tasks, KNN predicts the value of the new data point based on the average of its K nearest neighbors.

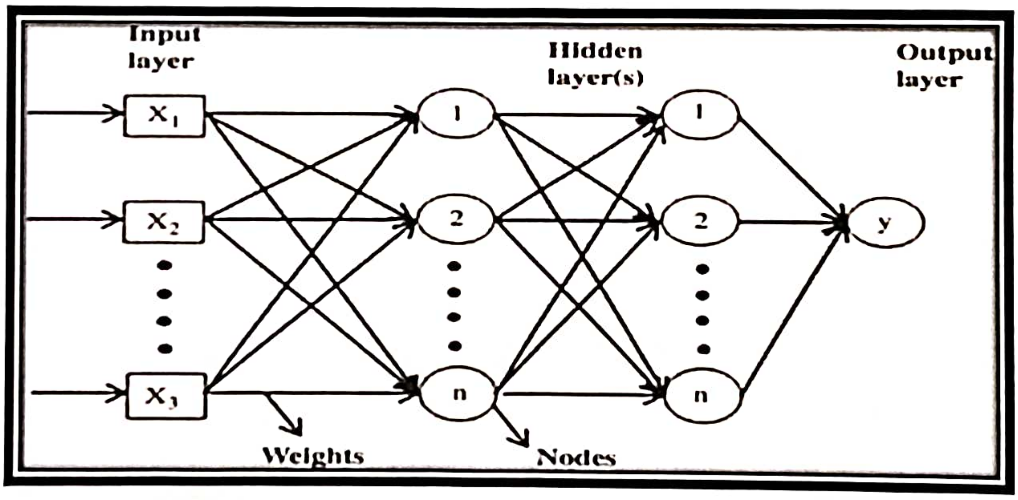


Figure 1.2 KNN Architecture

The KNN model's simplicity and intuitive nature make it a popular choice for initial exploration of data and as a baseline for comparison with more complex algorithms. However, its computational efficiency can be a limitation when dealing with large datasets or high-dimensional data due to the need to compute distances between all pairs of data points.

1.2.3 Random Forest

Training and testing a machine learning model using the Random Forest algorithm in a doctor bot system involves several critical steps aimed at enhancing diagnostic accuracy and personalized treatment recommendations. The process begins with the collection of extensive patient data, which includes demographic information, medical histories, symptoms, diagnostic results, and treatment outcomes. This data undergoes preprocessing, where it is cleaned to remove inconsistencies, normalized to ensure uniformity, and transformed into a format suitable for analysis. Feature extraction is then performed to identify and select the most relevant attributes that influence the diagnosis and treatment process.

Once the data is prepared, the Random Forest algorithm is selected for model training due to its robustness and ability to handle high-dimensional data. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks or the mean prediction for regression tasks. The training phase involves splitting the dataset into training and validation sets, where the model learns from the training set by building numerous decision trees based on different subsets of data and features. Each tree in the forest votes on the output, and the most common output is selected as the final prediction.

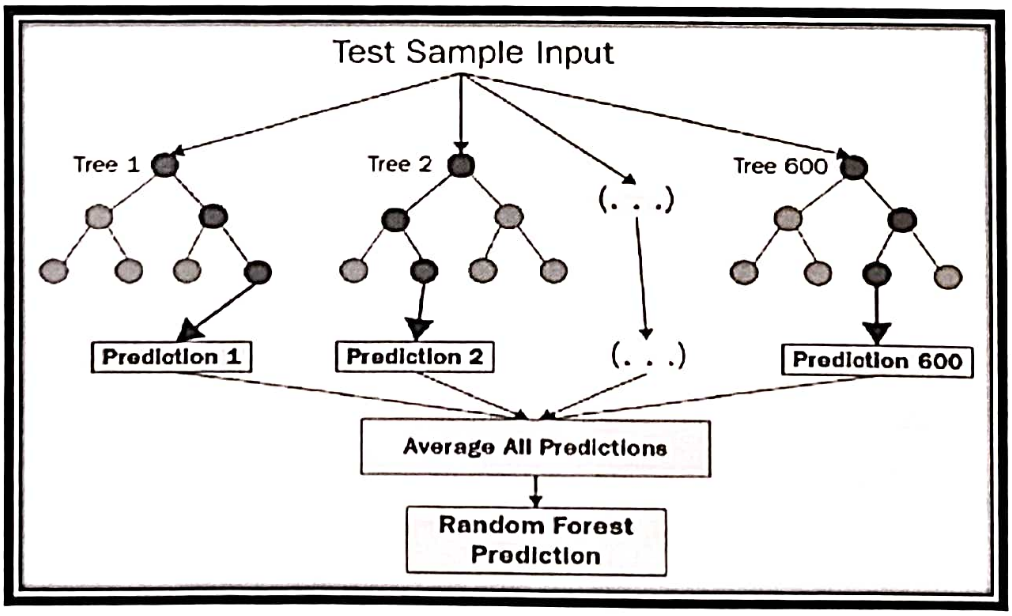


Figure 1.3 Random Forest Architecture

1.3 Applications

Doctor bots, powered by AI and machine learning, have a wide range of applications across healthcare settings. Some key applications include:

1. **Telemedicine and Remote Consultations:** Doctor bots enable remote consultations between patients and healthcare providers. They can assist in initial assessments, symptom checking, and recommending whether a patient needs further medical attention.
2. **Health Information and Education:** Bots provide valuable medical information to patients, caregivers, and the general public. They explain medical conditions, treatments, medications, and lifestyle recommendations based on individual health profiles.
3. **Appointment Scheduling and Reminders:** Automating appointment scheduling processes, doctor bots help patients book appointments, reschedule as needed, and send reminders to reduce no-shows.
4. **Symptom Checker and Triage**: Doctor bots can perform initial symptom assessment and triage, helping patients understand the severity of their symptoms and whether they should seek immediate medical care.
5. **Medication Management and Adherence:** Bots assist patients in managing their medications by providing reminders for doses, explaining medication instructions, and offering information on potential side effects.
6. **Health Monitoring and Feedback:** Integrated with wearable devices and health apps, doctor bots can monitor patients' health metrics (e.g., heart rate, blood pressure) and provide feedback or alerts for abnormal readings.
7. **Emergency Response and First Aid:** In emergency situations, doctor bots offer immediate first-aid advice, guide users through CPR procedures, and provide instructions for accessing emergency services.
8. **Support for Chronic Disease Management:** Bots aid in managing chronic conditions by offering personalized care plans, tracking disease progression, and providing guidance on lifestyle modifications.
9. **Mental Health Support:** Providing mental health assessments, coping strategies, and resources for emotional well-being, doctor bots contribute to mental health support initiatives.

1.4 Motivation

Building a doctor bot is motivated by the compelling potential to revolutionize healthcare delivery through artificial intelligence (AI) and machine learning technologies. This motivation stems from several key factors driving innovation in the healthcare sector.

Firstly, the primary goal of developing a doctor bot is to enhance access to healthcare services. Many regions face shortages of healthcare professionals, particularly in rural and underserved areas. By deploying doctor bots, individuals can access medical advice and information remotely, bridging geographical gaps and ensuring timely healthcare interventions. This accessibility is crucial for improving health outcomes, especially in emergencies or for individuals with limited mobility.

Secondly, doctor bots aim to improve efficiency within healthcare systems. Automating routine tasks such as appointment scheduling, medication reminders, and basic triage allows healthcare providers to focus more on direct patient care. This streamlining of administrative processes reduces wait times for patients, optimizes resource allocation, and enhances overall healthcare service delivery.

Moreover, the development of doctor bots is driven by the potential to provide personalized healthcare solutions. Through advanced algorithms and data analytics, these bots can analyze individual health data, preferences, and medical histories to offer tailored recommendations and treatment plans. This personalization not only improves patient satisfaction but also contributes to better treatment outcomes by aligning interventions with specific patient needs and characteristics.

Doctor bots contribute significantly to patient empowerment and health literacy. By providing comprehensive information on medical conditions, treatments, and preventive care measures, these bots enable individuals to make informed decisions about their health. Empowered patients are more likely to engage in proactive health management and adhere to prescribed treatments, ultimately leading to better long-term health outcomes.

1.5 Problem Statement

Healthcare delivery faces challenges in accessibility, efficiency, and patient engagement. Many patients struggle with timely access to medical advice, while healthcare providers contend with high patient volumes and administrative burdens.

1. **Access to Immediate Medical Advice:** Patients often encounter difficulties in accessing timely medical advice, leading to delayed treatment and unnecessary emergency visits.
2. **Efficiency in Healthcare Delivery:** Healthcare providers face challenges in efficiently managing patient inquiries, scheduling appointments, and maintaining accurate patient records.
3. **Patient Engagement and Education:** There is a need to enhance patient engagement through personalized health information, medication adherence support, and proactive health monitoring.
4. **Symptom Assessment and Triage:** Implement a robust symptom checker module that accurately assesses user-reported symptoms and provides triage recommendations based on severity.
5. **Personalized Health Recommendations:** Offer personalized health recommendations, including medication reminders, lifestyle advice, and preventive care measures tailored to individual health profiles.
6. **Appointment Scheduling and Management:** Enable patients to schedule appointments seamlessly through the bot, with integration capabilities into existing healthcare provider systems for real-time updates.
7. **Emergency Response Guidance:** Provide immediate first-aid instructions and emergency response guidance to users facing critical medical situations.
8. **Integration with Wearable Devices:** Integrate with wearable health devices to monitor real-time health metrics and provide actionable insights to users and healthcare providers.
9. **Medical Information and Education:** Offer comprehensive medical information and educational resources on diseases, treatments, and health conditions to empower patients in managing their health.
10. **Data Security and Compliance:** Ensure robust data security measures and compliance with healthcare regulations (e.g., HIPAA) to protect patient confidentiality and maintain trust.

CHAPTER 2

LITERATURE SURVEY

2.1 Introduction

A doctor bot that provides references integrates a feature where it can cite and refer to authoritative sources of medical information during interactions with users. This capability enhances the bot's credibility and reliability by ensuring that the information provided is backed by evidence-based sources, such as medical literature, clinical guidelines, and reputable healthcare institutions.

2.2 Related Work

[1]. M. Athish, E. Solomon, R. Indumathi, T. Manoj Prasath and R. Vinodhini, "Online Doctor Consultation and Automatic Medicine Dispensers in Using Machine Learning gap. 1-5, Doi: 10.1109/ADICS58448.2024.10533555. 2024

Health care is crucial for living a pleasant life. However, getting a doctor's appointment for a checkup is exceedingly challenging. Before contacting a doctor, it is suggested that healthcare chatbots be developed utilizing artificial intelligence that can identify the illness and provide basic information about it. The purpose of the medical chatbot is to lower healthcare expenditures and provide access to health-related information. A few chatbots serve as virtual medicinal reference materials, educating patients on their conditions and promoting better health. Chatbots can only truly serve the user when they can diagnose all diseases and deliver the correct information. An interactive message diagnostics bot converses with individuals about their health problems and offers a customized diagnosis based on their symptoms. Individuals will therefore be conscious of their health and be properly protected. Here, the patient who has arrived for therapy will immediately receive the appropriate medications based on the surgeon's digital prescription. This developed dispenser provides medicines based on the patient's symptoms with the help of healthcare professionals.

[2]. A. Goel, Satyam and S. Sharma, "Artificial Intelligence based Healthcare Chat Bot System," 2023 8th International Conference on Communication and Electronics Systems (ICCES)Coimbatore,India,2023,pp.1362-1366, doi:10.1109/ICCES57224.2023.10192727.

Machines are becoming capable of performing tasks like humans in this age of Artificial Intelligence. The most vital as pact of living a healthy life is health care, and the most difficult aspect is finding a doctor's consultant. Everyone cannot afford to visit a doctor for every health issue. The objective of the proposed research work is to design or build a Healthcare chat bot in AI to assist in determining the patient's health and providing basic information before contacting a doctor, but only for minor issues. Using-a medical chat bot will save healthcare costs while also increasing medical knowledge. Chat bots are computer programs that employ AI and ML to connect with people. The chat bot system retrieves the query from the database that the user has requested and makes a judgement based on it before presenting the responses.

[3]. T. A. H. Shaikh and M. Mhetre, "Autonomous AI Chat Bot Therapy for Patient with Insomnia," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-5, Doi: 10.1109/I2CT54291.2022.9825008.

This paper's endeavor is concentrated primarily aiding insomniacs. The importance of accomplishing the work is critical since insomnia is a common sleep condition that stops people from sleeping and therefore is common in today's society. Because of its chronic nature, insomnia is connected to a considerable deterioration in an individual's quality of life. The system is powered via Artificial Intelligence and Deep Learning. The goal of this system is to increase the number of encounters with these individuals as they get sadder and more anxious. So, essentially, a friendly Chat Bot is created to better understand and treat insomniacs. The algorithm, how it works, and why it has to be implemented are briefly discussed in the paper.

[4]. B. Dinesh, P. Chilukuri, G. P. Sree, K. Venkatesh, M. Delli and K. R. Nandish, "Chat and Voice Bot Implementation for Cardio and ENT Queries Using NLP," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 124-130, Doi: 10.1109/ICIDCA56705.2023.10099942.

Human to machine conversational UI's, commonly referred to as chatbots, helps users with a new method to talk with computers. Users can ask queries to chatbots just like they would to a human. But chatbots are increasingly often utilized in online conversation services. Artificial intelligence is used by these bots to comprehend human input and react appropriately. The three most important areas to focus on are medicine/health, agriculture, and education. Nowadays, chatbots can be utilized anyplace a person can talk to a computer at any time. The fastest expanding sectors for hiring are customer service, sales, marketing, and branding, as well as human resources. This prediction tasks are now carried out by other chatbots (especially in the medical industry), made possible by developments in AI and data mining methods.

[5]. M. G, "Mega Bot – The Healthcare Chatbot," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1386-1395, Doi: 10.1109/ICESC51422.2021.9533025.

This project aims to help people to check their health condition before going to a hospital this saves a lot of time and value and also allows you to know your symptoms in advance. It can perform tasks related to symptoms same like a doctor who takes a checkup but in a more accurate way. Machine Learning (ML), Artificial Intelligence (AI), Deep-Learning (DL) and IOT are used to detect the symptoms and checks the temperature. This algorithm using dependencies such as pysklearn, nltk, Tensor Flow, OpenCV, AI, and Raspberry Pi. It usually enters by entering the login details and then it asks you to click the start button and then you can say "yes" or "no" and then it finally uses to suggest the online doctors. There are also way more features like face checkup, fever checkup and third-party intermedia API.

[6]. R. Kaladevi, S. Saidineesha, P. Keerthi Priya, K. M. Nithiya and S. Sai Gayatri, "Chatbot For Healthcare Using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-4, Doi: 10.1109/ICCCI56745.2023.10128261.

Healthcare has become an important part of living a healthy life in today's society, but consulting with a doctor for every health problem is very difficult for common people. The Medical chat bot is created to diagnose diseases and provide basic information before consulting a physician. By doing so, healthcare costs will be reduced and medical knowledge will improve. A chat bot interacts with users through natural language. The chat bot then saves the information in a database, chooses sentence keywords, decides whether to answer an inquiry-, and provides information. With this approach, people can spend less time in hospitals while performing repetitive tasks like providing solutions, sending emails, marketing, and analyzing results. Similarly, calculations are done using n-grams, TF-IDFs, and cosine. So, this gives the best result for detecting the disease through a medical chat bot.

[7]. A. K. Tripathy, R. Carvalho, K. Pawaskar, S. Yadav and V. Yadav, "Mobile based healthcare management using artificial intelligence," 2015 International Conference on Technologies for Sustainable Development (ICTSD), Mumbai, India, 2015, pp. 1-6, Doi: 10.1109/ICTSD.2015.7095895.

In this growing age of technology, it is necessary to have a proper health care management system which should be sent percent accurate but also should be portable so that every person carry with it as personalized health care system. The health care management system which will consist of mobile based Heart Rate Measurement so that the data can be transferred and diagnosis based on heart rate can be provided quickly with a click of button. The system will consist of video conferencing to connect remotely with the Doctor. The Doc-Bot which was developed earlier is now being transferred to mobile platform and will be further advanced for diagnosis of common diseases. The system will also consist of Online Blood Bank which will provide up-to-date details about availability of blood in different hospitals.

[8]. P. Dohare, S. Johri, S. Priya, S. Singh and S. Upadhyay, "Good Fellow : A Healthcare Chatbot System," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, Doi:10.1109/ICCCNT56998.2023.10308009.

Health is essential to our lives. Every aspect of life is dependent on overall health. Healthcare is critical to living a healthy lifestyle. It is difficult to get a consultation with a doctor in certain unsocialized places. With such a large population, it is also difficult for doctors to attend so many patients. The fundamental aim here is to create a healthcare chatbot that can identify diseases and offer necessary information about the individual ailment before consulting with or seeing a doctor. This will assist in lowering the cost of health care while also improving access to medical information via a medical chatbot. Purpose: Nowadays, individuals are less concerned with their health. People avoid coming to the hospital for issues that could eventually develop into serious illnesses. Instead of looking through the list of potentially pertinent papers from the web, creating question-and-answer forums is increasingly becoming a straightforward method of responding to those inquiries. The purpose is to develop a healthcare chatbot system that, before contacting a doctor, can diagnose the illness and provide basic information about the disease.

[9]. U. Bharti, D. Bajaj, H. Batra, S. Lalit, S. Lalit and A. Gangwani, "Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 870-875, Doi: 10.1109/ICCES48766.2020.9137944.

Telemedicine can be used by medical practitioners to connect with their patients during the recent Coronavirus outbreak, whilst attempting to reduce COVID-19 transmission among patients and clinicians. Amidst the pandemic, Telemedicine has the potential to help by permitting patients to receive supportive care without having to physically visit a hospital by using a conversational artificial intelligence-based application for their treatment. Thus, telehealth will rapidly and radically transform in-person care to remote consultation of patients. Because of this, it developed a Multilingual Conversational Bot based on Natural Language Processing (NLP) to provide free primary healthcare education, information, advice to chronic patients. The study introduces a novel computer application acting as a personal virtual doctor that has been opportunely designed and extensively trained to interact with patients like human beings.

[10]. Y. A. R, A. S. Selvaraj, S. N and H. R. T, "Chatbot for Healthcare System Using NLP and Python," 2022 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Delhi, India, 2022, pp. 1-7, Doi: 10.1109/SOLI57430.2022.10294817.

After the 20th century ended, people in the 21st century started to worry more about their own health and the health of their families. Realizing that "health is wealth," they decide that their health is the one asset they should safeguard throughout their lives. It is significant that medication is essentially necessary for most individuals to lead happy lives. However, it may be challenging to schedule an appointment with a doctor for every issue that arises in our bodies. The goal is to use Python and JSON to build a healthcare bot that, prior contacting a physician, could recognize the illness and give accurate, basic information about that as well. This will essentially contribute to lowering healthcare expenses and expanding access to health records through a healthcare chatbot, both of which are very important. These AI systems were software applications which utilise Python as a driver and will basically be kept in a CV file to basically communicate with users, proving that this will basically help to cut medical expenses and enhance accessibility to medical information through medical chatbots, or so they think.

CHAPTER 3

SYSTEM REQURIMENTS

3.1 Hardware Requirements

When considering the hardware requirements for a doctor bot system based on machine learning models like decision trees and K-Nearest Neighbors (KNN), it's important to assess the computational needs of both training and inference phases. Here's a breakdown focusing on your laptop as a baseline:

1. **Processor (CPU):** A multi-core processor (e.g., Intel Core i5) is sufficient for basic machine learning tasks. However, for more intensive computations, especially with large datasets or complex models, a higher-end processor (e.g., AMD Ryzen 7) or even a workstation-class CPU (e.g., Intel Xeon) may be beneficial.
2. **Memory (RAM):** At least 8 GB of RAM is recommended for running machine learning algorithms effectively. For larger datasets or more memory-intensive tasks, 16 GB or more can significantly improve performance.
3. **Storage (Hard Drive):** SSD storage is preferable over HDD due to faster read/write speeds, which can accelerate data loading and model training times. A minimum of 256 GB SSD is recommended, with more storage space beneficial for storing datasets and model checkpoints.
4. **Graphics Processing Unit (GPU):** While not essential for all machine learning tasks, having a dedicated GPU can drastically speed up training times, especially for deep learning models.

Setup for Doctor Bot Systems:

1. System: Intel Core i5 12th Gen 12500H.

2. Ram: 16 GB/512 GB.

3. Hard Disk: 10 GB or above

4. Input Device: Mic, Keyboard and Mouse.

5. Output Device: Monitor or PC

3.2 Software Requirements

To analyze the software requirements for a doctor bot system based on machine learning, considering your laptop specifications, we'll focus on the typical components and tools needed for developing and deploying such applications.

1. **Development Environment:**

   Python: Doctor bots often rely on Python due to its extensive libraries for machine learning (e.g., TensorFlow, PyTorch, scikit-learn) and natural language processing (e.g., NLTK, spaCy).

   Integrated Development Environment (IDE): Recommended IDEs include PyCharm, Jupyter Notebook, or VS Code, providing a user-friendly interface for coding and debugging.

2. **Machine Learning Frameworks and Libraries:**

   TensorFlow or PyTorch: Leading frameworks for building and training neural networks, essential for tasks like natural language understanding and image recognition.

   Scikit-learn: Library for classical machine learning algorithms such as SVMs, decision trees, and ensemble methods.

   NLTK and spaCy: Libraries for natural language processing tasks such as tokenization, part-of-speech tagging, and named entity recognition.

3. **Data Handling and Visualization:**

   Pandas: Data manipulation and analysis library, crucial for handling structured data such as patient records or clinical datasets.

   Matplotlib and Seaborn: Libraries for creating visualizations to understand data distributions, trends, and model performance metrics.

System analysis is a structured approach to studying a system whether it involves business processes, technology, or organizations by examining its components, interactions, and governing processes. The primary objective of system analysis is to understand how a system functions, identify potential problems or inefficiencies, and propose improvements. The first step is defining and scoping the system, including its boundaries, components, and step is defining and scoping the system, including its boundaries, components, and stakeholders. This process involves understanding the system's goals and objectives to set the stage for further analysis.

Efficient development and deployment of a doctor bot system leveraging machine learning (ML) requires rigorous system analysis and careful consideration of software requirements. At the core of this process is defining clear functional and non-functional requirements. Functionally, the bot must support natural language understanding (NLU) to interpret medical queries accurately, requiring robust NLP algorithms capable of entity recognition, intent classification, and context-aware dialogue management. Data handling capabilities are crucial, necessitating efficient data preprocessing pipelines to clean, normalize, and integrate heterogeneous medical data sources such as electronic health records (EHRs) and medical imaging datasets.

Machine learning models, selected based on task-specific requirements (e.g., classification, regression), should be scalable and capable of continual learning to adapt to new medical insights. Non-functional requirements encompass performance metrics (e.g., response time, accuracy), scalability to handle concurrent user queries, and adherence to healthcare data privacy regulations (e.g., HIPAA) to ensure patient confidentiality. Integration with existing healthcare IT infrastructure for seamless data exchange and interoperability is essential, requiring adherence to standards like HL7 for data interchange and FHIR for healthcare data representation.

Real-time processing capabilities are essential for interactive user experiences. The system should incorporate streaming and batch processing frameworks like Apache Kafka or Apache Flink for handling continuous data streams and performing near-real-time predictions. Ensuring compliance with healthcare regulations such as HIPAA (Health Insurance Portability and Accountability Act) is paramount. The system must implement robust security measures to protect patient data and maintain confidentiality throughout data storage, transmission, and access.

Continuous monitoring and evaluation are crucial to assess model performance, refine algorithms, and ensure alignment with evolving medical practices. Overall, optimizing system efficiency involves a comprehensive approach to software design, encompassing robust architecture, scalable algorithms, stringent data governance, and adherence to regulatory standards, thereby enhancing the doctor bot's effectiveness in supporting healthcare delivery.

CHAPTER 4

DESIGN

4.1 Purpose

This chapter provides a comprehensive overview of the design of the suggested system, describing its general architecture from conception to specific improvements in later design stages. It explores both static and dynamic characteristics, carefully capturing the actions of particular objects. This documentation is essential for steering the project's progress since it serves as a roadmap for the implementation and testing phases. It is important to recognize that the details discussed below are anticipated to change as the design progresses in order to accommodate new needs and insights. The conceptual design forms the basis upon which successive design stages add increasingly detailed layers to enhance it. This chapter is an invaluable resource for the development team, giving them insight into the architecture and operation of the system and guiding their decisions and actions in the implementation and testing phases.

4.2 System Architecture

Within the system architecture of a doctor bot, Decision Trees and KNN serve distinct roles in processing medical data and providing intelligent responses. Decision Trees operate by recursively partitioning the dataset into smaller subsets based on the value of attributes, aiming to create subsets that are as homogeneous as possible with respect to the target variable (e.g., disease diagnosis). Each internal node of the tree represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label or decision. This hierarchical structure allows the bot to follow a sequence of if-then rules to classify patient symptoms or medical records accurately. Decision Trees are particularly useful in healthcare for their interpretability, enabling clinicians to understand the reasoning behind a diagnosis or treatment recommendation.

On the other hand, KNN is a non-parametric algorithm that classifies objects based on their similarity to neighboring objects in the feature space. In the context of a doctor bot, KNN calculates the distance between a new input (e.g., patient symptoms) and all other data points in the training dataset. It then selects the k nearest neighbors (data points with the closest distance) and assigns the most frequent class among them to the new input. This approach is for instance, to recommend treatments based on similarities between patient profiles or to predict the likelihood of a patient developing a specific condition based on similar historical cases.

Both Decision Trees and KNN contribute to the doctor bot's ability to analyse complex medical data, provide accurate predictions, and support clinical decision-making. Their integration within the bot's system architecture enhances its capability to handle diverse healthcare scenarios, from initial symptom assessment to personalized treatment recommendations, thereby improving patient care and healthcare efficiency. Machine learning architecture is cutting-edge and high-pressure field. Therefore, you need to have an excellent IT resume. Simply put you are at the centre, ensuring the system works well. Selecting the optimum storage for all phases of a machine learning project is more important than you except. What usually happens is that teams produce numerous datasets versions and experiment with various model designs. When a model designs. When a model is promoted to productions, it must make predictions on fresh data efficiently. The ultimate aims is to have well-trained model operating in production that adds Ai to an applicant.

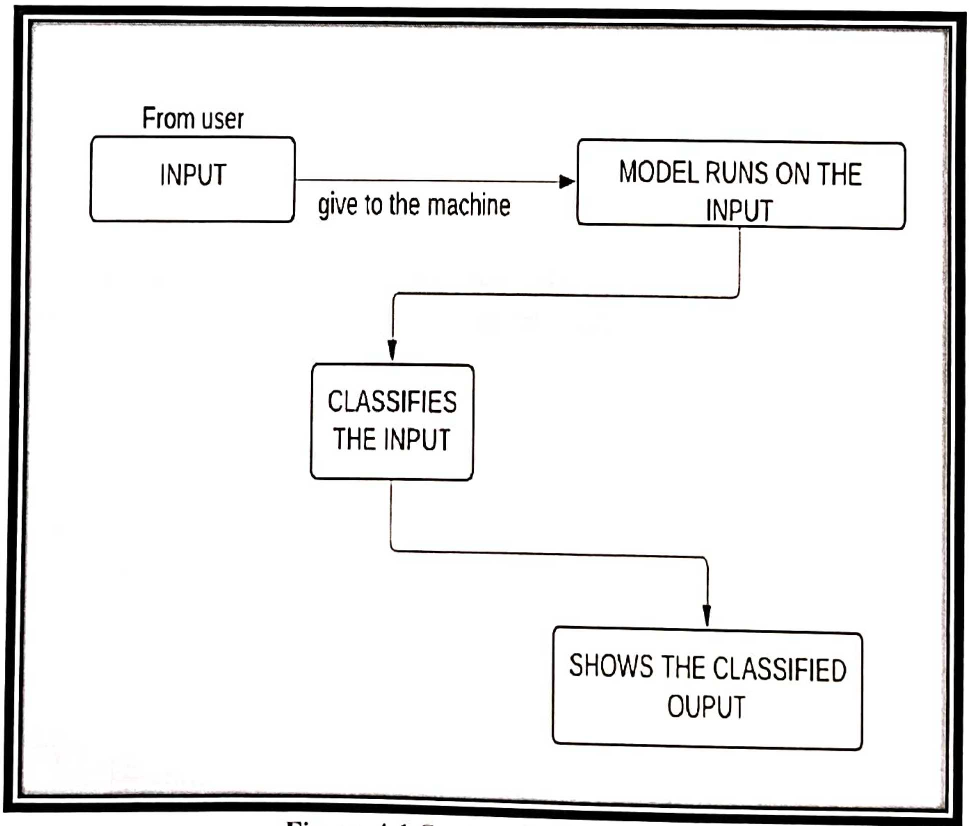


Figure 4.1 System Architecture

4.3 Modules Division

4.3.1 Dataset

The "Symptom_Description" dataset typically consists of information linking symptoms to possible diseases or conditions. It usually includes a comprehensive list of symptoms and their corresponding descriptions, detailing how each symptom manifests in a patient. This dataset is crucial for doctor bot systems as it helps them understand and interpret user inputs regarding symptoms.

On the other hand, the "Disease_Drug" dataset pairs diseases or medical conditions with recommended drugs or treatments. This dataset is essential for doctor bot systems to suggest potential treatments or medications based on identified diseases or conditions from the symptom analysis. By combining these datasets, doctor bot systems can effectively simulate diagnostic processes and provide relevant treatment recommendations based on user-provided symptoms.

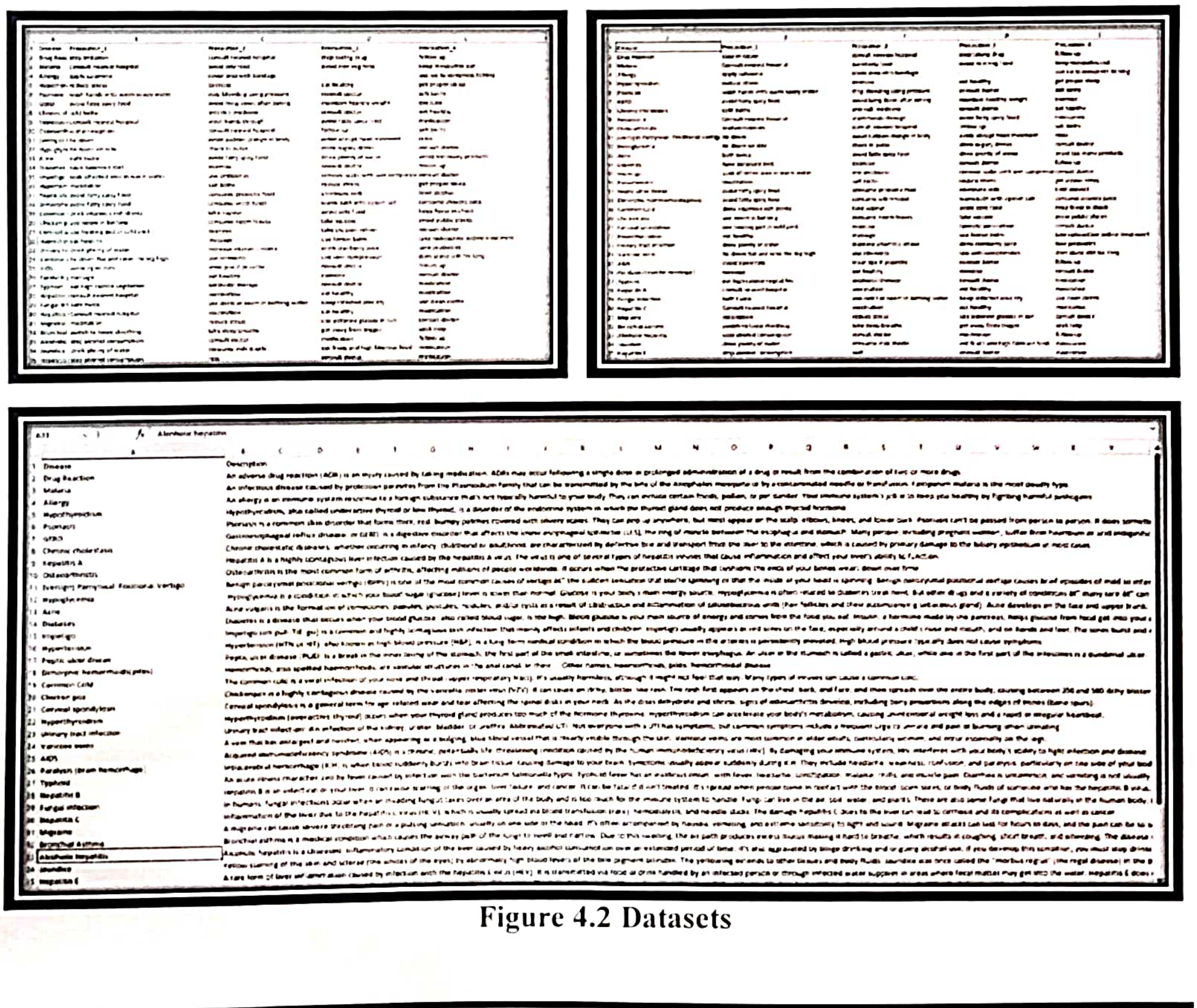


Figure 4.2 Datasets

4.3.2 Training and Testing

Training and testing machine learning models such as decision trees (DT) and k-nearest neighbors (KNN) using CSV files involves several structured steps to ensure effective model development and evaluation. Decision trees are hierarchical structures that recursively partition data based on feature attributes, aiming to classify or predict outcomes by creating a tree-like model of decisions.

To start, the CSV file serves as the primary data source, containing rows representing instances (cases or observations) and columns representing features (attributes or variables). Each row typically corresponds to a patient case with associated symptom descriptions and possibly other relevant data points like demographic information or medical history.

During the training phase, the CSV data is split into two main subsets: a training set and a testing set. The training set, usually larger than the testing set, is used to build the models. For decision trees, the algorithm recursively splits the data based on feature values to create nodes that best differentiate between classes (or outcomes). For KNN, the algorithm stores all instances and their class labels to classify new instances by comparing them to the k-nearest neighbors based on a chosen distance metric (e.g., Euclidean distance).

4.3.3 Decision Tree Model

The decision tree algorithm processes the extracted features. It starts at the root node and progresses through internal nodes based on predefined rules derived during the model training phase. At each node, a decision is made based on the feature values, directing the flow of decision-making towards subsequent nodes or leaf nodes.

As the decision tree algorithm traverses the nodes, it eventually reaches a leaf node. Each leaf node corresponds to a specific medical condition or disease diagnosis. The decision tree's structure ensures that the path taken through the nodes aligns with the input features provided, leading to a conclusive diagnosis at the leaf node.

The final output of the decision tree algorithm is the diagnosis or medical condition associated with the input symptom descriptions. This output is generated based on the path taken through the decision nodes and reflects the most probable condition given the input data.

Decision trees are valued for their interpretability, as each decision node and leaf node can be traced back to specific feature criteria. This transparency allows healthcare providers or users of the doctor bot to understand how the diagnosis was reached based on the input symptoms.

Doctor bots utilizing decision trees can be continually improved through iterative updates to the training data and model parameters. This ongoing process enhances the accuracy and reliability of diagnostic decisions over time, making the bot more effective in providing medical guidance based on symptom descriptions.

4.3.4 KNN Model

Using the KNN algorithm, the doctor bot calculates the similarity (often using metrics like Euclidean distance) between the input symptom profile and each historical patient case in its dataset. This involves comparing the input features with corresponding features of past cases that are most similar in terms of symptoms.

Based on the calculated similarities, the KNN algorithm identifies the k nearest neighbors—historical patient cases from the dataset that most closely resemble the input symptoms. The value of k is pre-determined and influences the number of neighbors considered for decision-making.

Once the nearest neighbors are identified, the doctor bot employs a majority voting mechanism or weighted average to determine the most probable diagnosis or recommendation. For classification tasks (e.g., identifying a disease), the bot may classify the input based on the most frequent class among its nearest neighbors. For regression tasks (e.g., predicting severity), it may estimate the severity based on averaged values from similar cases.

Finally, the doctor bot generates an output based on the decision made. This output could range from suggesting potential illnesses or conditions matching the input symptoms to recommending specific actions such as further diagnostic tests, lifestyle changes, or immediate medical attention.

As more data is collected and incorporated into the dataset, the KNN model can be periodically retrained to improve its accuracy and relevance in decision-making. This iterative learning process ensures that the doctor bot stays updated with the latest medical knowledge and trends, enhancing its ability to provide accurate and timely recommendations.

4.3.4 Random Forest

The implementation of a Random Forest model in a doctor bot system aims to significantly enhance the accuracy and reliability of medical diagnoses and treatment recommendations. The process begins with comprehensive data collection, which involves gathering extensive patient information, including demographics, medical histories, symptom descriptions, diagnostic results, and treatment outcomes. This data undergoes rigorous preprocessing to ensure quality and consistency, involving steps such as cleaning, normalization, and transformation into a suitable format for analysis. Feature selection is a critical step where the most relevant attributes are identified to contribute to the prediction task effectively.

In training the Random Forest model, the dataset is divided into training and testing subsets, typically using an 80/20 split. The Random Forest algorithm, which is an ensemble learning method, constructs multiple decision trees during training. Each tree is trained on a random subset of the training data, ensuring diversity and reducing overfitting. The model combines the predictions from these individual trees to make a final prediction, leveraging the "wisdom of the crowd" to enhance overall accuracy and robustness.

During the training phase, hyperparameters such as the number of trees in the forest and the maximum depth of each tree are tuned to optimize performance. The model learns to classify or predict outcomes based on the training data, with each decision tree voting on the outcome and the majority vote determining the final prediction. Following training, the model's performance is evaluated on the testing subset to assess its accuracy, precision, recall, and other relevant metrics. Cross-validation techniques are often employed to further validate the model's generalizability and to mitigate the risk of overfitting.

Post-training, the model undergoes rigorous testing using unseen data to ensure it performs well in real-world scenarios. This involves assessing its ability to accurately diagnose conditions, predict treatment outcomes, and recommend interventions. Continuous monitoring and periodic retraining with new data are essential to maintain the model's accuracy and relevance over time. The deployment of the Random Forest model in the doctor bot system enables the automation of medical decision-making processes, providing reliable diagnostic assistance and personalized treatment recommendations. This not only augments the capabilities of healthcare professionals but also improves patient outcomes by ensuring timely and accurate medical interventions.

CHAPTER 5

IMPLEMENTATION

CODE

MODEL- Random Forest

```
import streamlit as st

import pandas as pd

import numpy as np

import pickle

from sklearn.preprocessing import LabelEncoder

# Load the models

with open('random_forest.pkl', 'rb') as file:

    rf = pickle.load(file)

# Load the encoders

df = pd.read_csv('dataset.csv')

label_encoders = {}

for column in df.columns:

    label_encoders[column] = LabelEncoder()

    df[column] = label_encoders[column].fit_transform(df[column])

# Load the symptom description data

df1 = pd.read_csv("symptom_Description.csv")

dfdict = df1.set_index('Disease').T.to_dict('list')

# Function to classify a new test case

def classify_new_case(model, label_encoders, new_case):

    new_case_df = pd.DataFrame([new_case])

    for column in new_case_df.columns:
```

```
le = label_encoders[column]
        unseen_labels = set(new_case_df[column].unique()) - set(le.classes_)
        if unseen_labels:
            le.classes_ = np.append(le.classes_, list(unseen_labels))
        new_case_df[column] = le.transform(new_case_df[column])
    prediction = model.predict(new_case_df)
    disease = label_encoders['Disease'].inverse_transform(prediction)
    return disease[0]
# Streamlit UI
st.title('Disease Prediction from Symptoms')
# Create input fields for symptoms
symptoms = {}
all=st.text_input('Symptoms', 'none')
alls=all.split(',')
for i in range(1,18):
    symptoms[f'Symptom_{i}'] = 'none'
for i in range(1, len(alls)):
    symptoms[f'Symptom_{i}'] = alls[i-1]
# Dropdown to choose model
model_choice = st.selectbox('Choose the model', ('Random Foresrt'))
if st.button('Predict Disease'):
    if model_choice == 'Random Foresrt':
        predicted_disease = classify_new_case(rf, label_encoders, symptoms)
    description = dfdict[predicted_disease][0]
    st.subheader('Predicted Disease')
    st.write(predicted_disease)
    disease=predicted_disease
    st.subheader('Disease Description')
    st.write(description)
```

CHAPTER 6

RESULTS

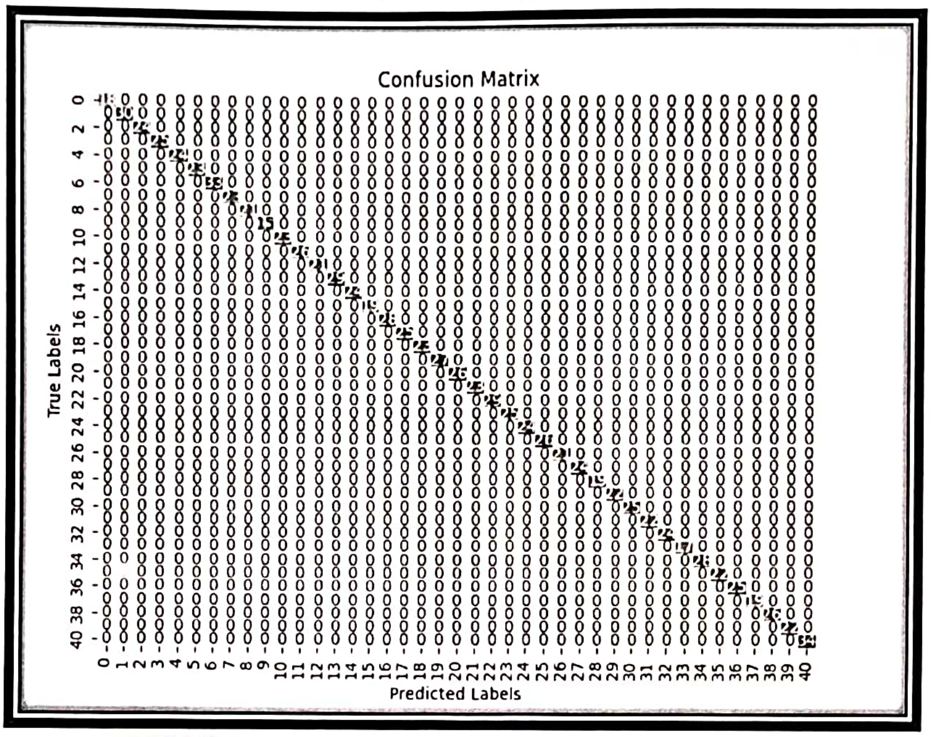


Figure 6.1 Confusion Matrix

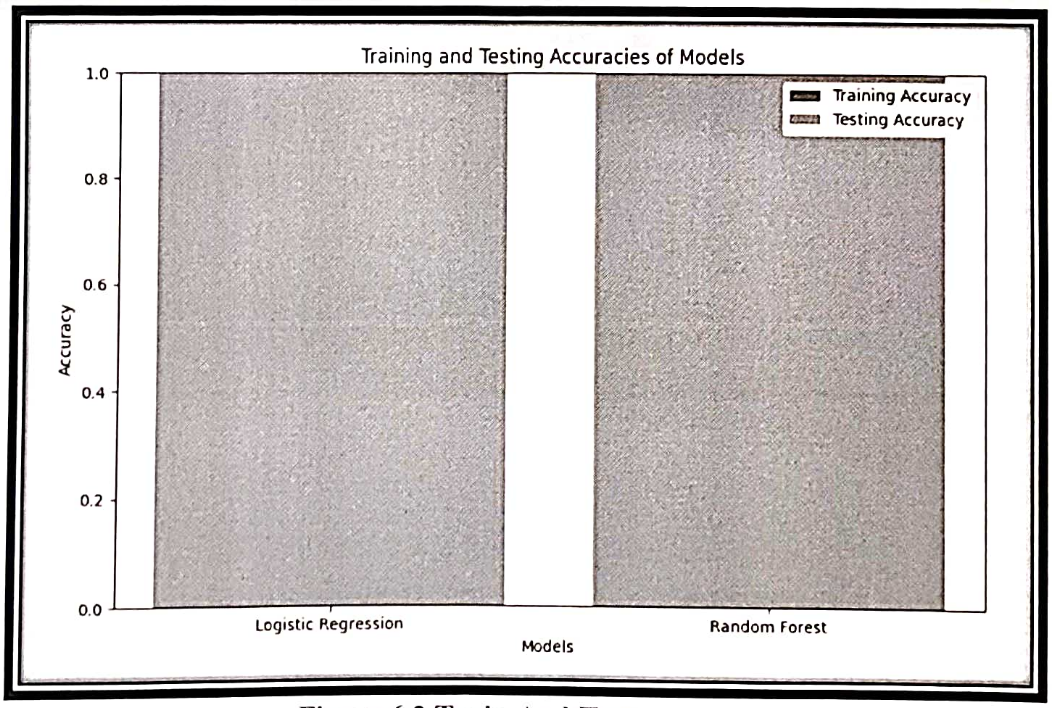
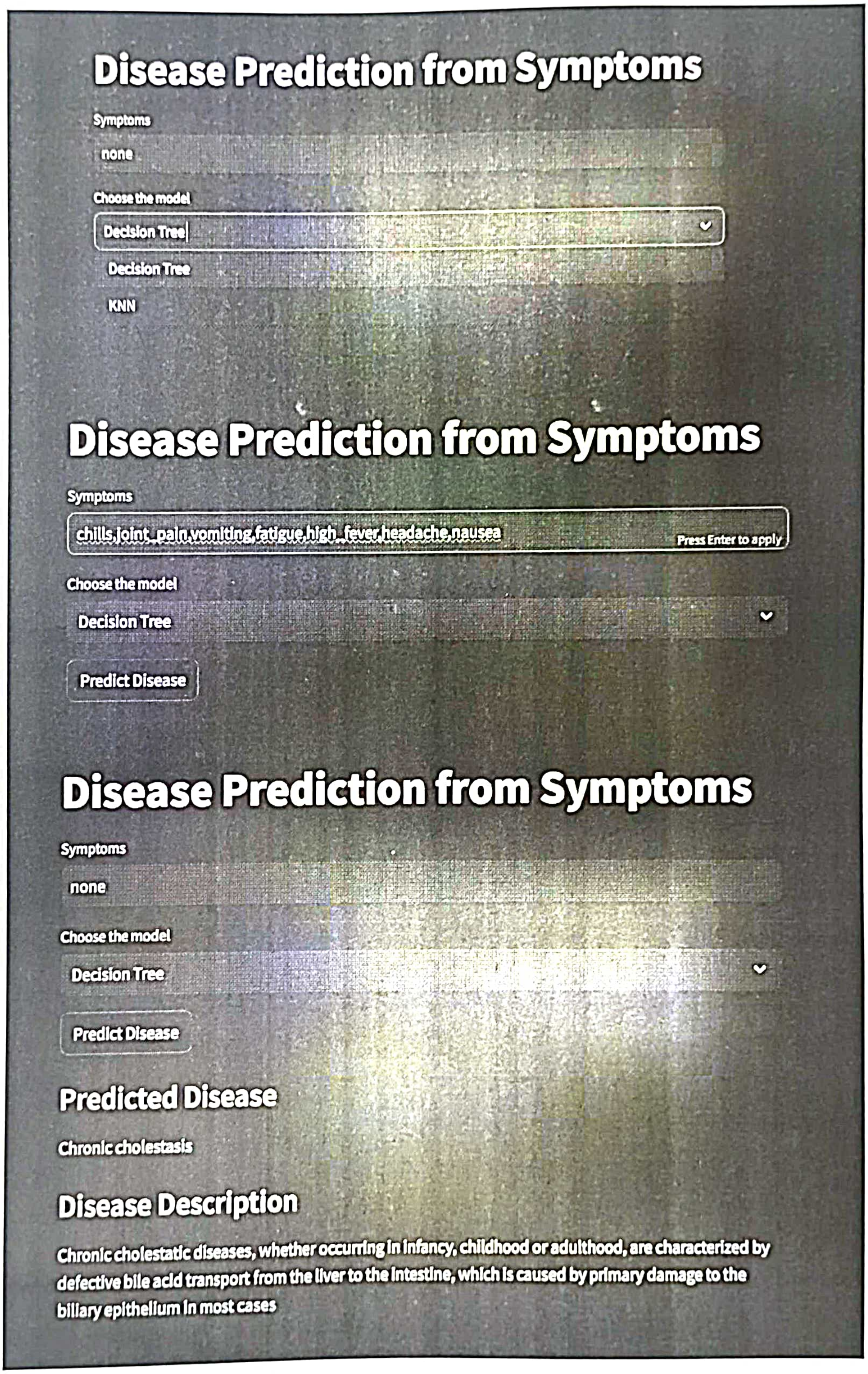


Figure 6.2 Train And Test


Disease Prediction from Symptoms
Symptoms
none
Choose the model
Decision Tree
Decision Tree
KNN
Disease Prediction from Symptoms
Symptoms
chills,joint_pain,vomiting,fatigue,high_fever,headache,nausea
Press Enter to apply
Choose the model
Decision Tree
Predict Disease
Disease Prediction from Symptoms
Symptoms
none
Choose the model
Decision Tree
Predict Disease
Predicted Disease
Chronic cholestasis
Disease Description
Chronic cholestatic diseases, whether occurring in infancy, childhood or adulthood, are characterized by defective bile acid transport from the liver to the intestine, which is caused by primary damage to the biliary epithelium in most cases

CONCLUSION AND FUTURE EXTENSION

Doctor Bot, integrating machine learning models such as decision trees and k-nearest neighbors (KNN) into a doctor bot that accepts both text and voice inputs, and provides outputs in text and voice formats, significantly enhances its versatility and usability in healthcare applications. By accepting text inputs, the bot can analyze symptom descriptions efficiently, utilizing decision trees to classify and recommend diagnoses based on structured decision paths. Simultaneously, voice inputs enhance user interaction by allowing natural language input, which the bot processes using KNN to match spoken symptoms with similar historical cases for accurate diagnosis predictions. Outputting results in both text and voice formats ensures accessibility and ease of use for users across different communication preferences and situations, facilitating seamless integration into healthcare environments. This multifaceted approach not only improves user engagement and satisfaction but also reinforces the bot's reliability and effectiveness in delivering timely and informed medical recommendations.

For future extension, we can improve the available dataset to get more accurate and appropriate results and add new features such as giving the remedies for the predicted disease and suggesting the doctors available nearby the user in that domain.

REFERENCES

[1]. M. Athish, E. Solomon, R. Indumathi, T. Manoj Prasath and R. Vinodhini, "Online Doctor Consultation and Automatic Medicine Dispensers in Using Machine Learning gap. 1-5, Doi: 10.1109/ADICS58448.2024.10533555. 2024

[2]. A. Goel, Satyam and S. Sharma, "Artificial Intelligence based Healthcare Chat Bot System," 2023 8th International Conference on Communication and Electronics Systems (ICCES)Coimbatore, India,2023, pp.1362-1366, doi:10.1109/ICCES57224.2023.10192727.

[3]. T. A. H. Shaikh and M. Mhetre, "Autonomous AI Chat Bot Therapy for Patient with Insomnia," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-5, Doi: 10.1109/I2CT54291.2022.9825008.

[4]. B. Dinesh, P. Chilukuri, G. P. Sree, K. Venkatesh, M. Delli and K. R. Nandish, "Chat and Voice Bot Implementation for Cardio and ENT Queries Using NLP," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 124-130, Doi: 10.1109/ICIDCA56705.2023.10099942.

[5]. M. G, "Mega Bot – The Healthcare Chatbot," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1386-1395, Doi: 10.1109/ICESC51422.2021.9533025.

[6]. R. Kaladevi, S. Saidineesha, P. Keerthi Priya, K. M. Nithiya and S. Sai Gayatri, "Chatbot For Healthcare Using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-4, Doi: 10.1109/ICCCI56745.2023.10128261.

[7]. A. K. Tripathy, R. Carvalho, K. Pawaskar, S. Yadav and V. Yadav, "Mobile based healthcare management using artificial intelligence," 2015 International Conference on Technologies for Sustainable Development (ICTSD), Mumbai, India, 2015, pp. 1-6, Doi: 10.1109/ICTSD.2015.7095895.

Vision

To become a premier institute transforming our students to be global professionals.

Mission

M1: Develop competent Human Resources, and create state-of-the-art infrastructure to impart quality education and to support research.

M2: Adopt tertiary approach in teaching – learning pedagogy that transforms students to become professionally competent technocrats and entrepreneurs.

M3: Nurture and train students to develop the qualities of global professionals.

Department of Computer Science and Engineering

Vision

To impart quality education in the field of Computer Science and Engineering with emphasis on innovative thinking, communication and leadership skills to meet the global challenges in IT paradigm.

Mission

M1: Focus on student centric approach through experiential learning and necessary infrastructure.

M2: Develop innovative thinking, communication and leadership skills by creating conducive environment and relevant training.

M3: Enrich students by developing the traits of global professionals.

BRIDGE TO THE FUTURE

CAMBRIDGE

INSTITUTE OF TECHNOLOGY

CAMBRIDGE INSTITUTE OF TECHNOLOGY

K. R. PURAM, BENGALURU - 560036