IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# DISEASE PREDICTION AND DRUG RECOMMENDATION SYSTEM

**[1]Okemiri Henry Anayo, [2]Edward Uche-Nwachi, [3]Afolabi Idris Yinka, [4]Nwali Monday, [5]Azegba Okechukwu**

[12345]Lecturer

[12345]Computer Science Department, Alex Ekwueme Federal University, Ndufu Alike, Ebonyi State, Nigeria

*Abstract :* The Disease Prediction and Drug Recommendation System (DPDRS) is an innovative approach aimed at improving healthcare outcomes through advanced data analytics and machine learning. As the volume of medical data grows exponentially, there is an urgent need for intelligent systems that can assist healthcare professionals and patients in making informed decisions. This project develops a user-friendly platform that accurately predicts potential diseases based on a set of input symptoms and provides tailored drug recommendations to aid in treatment. Utilizing a robust database of symptoms, diseases, and corresponding medications, the DPDRS employs sophisticated algorithms such as decision trees and neural networks. For instance, if a user inputs symptoms like fever, cough, and fatigue, the system can analyze these inputs to predict conditions such as influenza or COVID-19, while also suggesting appropriate medications, dosages, and precautions.

***IndexTerms – Prediction, Machine Learning, Healthcare, Intelligent system***

## 1. Introduction

In today's world, there are countless medications available for various health conditions. However, finding the right medication for an individual can be challenging due to differences in genetics, medical history, and lifestyle factors. This challenge is further compounded by the sheer volume of available drugs and the complexity of medical data. In a traditional drug recommendation system, the system typically relies on established medical guidelines, clinical expertise, and patient-specific factors to suggest medications. Healthcare professionals assess a patient's medical history, symptoms, lab results, and other relevant information to make informed decisions about which drugs to prescribe. These decisions are based on established protocols, drug efficacy, safety profiles, potential interactions, and the patient's individual characteristics. Furthermore, traditional methods may not always incorporate the latest research findings or account for novel therapeutic options, potentially limiting the effectiveness of the recommendations. This situation underscores the need for more advanced, data-driven approaches to drug recommendation that can better handle the complexities of modern medicine and provide more precise, personalized treatment options for patients (Amos et al., 2021).

Similarly, predicting a given disease according to Patel et al. (2021) is often time-consuming and challenging due to the complexity and volume of patient data involved. Factors such as genetic differences, medical history, symptoms, and lifestyle all play a significant role in diagnosing and predicting diseases. The process can be lengthy, requiring health workers to gather extensive information through medical tests, consultations, and patient monitoring. Additionally, healthcare professionals must stay updated on the latest medical research and guidelines, which can further complicate the diagnosis process.

Patel et al. (2021), in traditional healthcare settings, disease prediction typically involves the following steps: health workers begin by gathering a patient's medical history and conducting physical examinations. Diagnostic tests such as blood work and imaging are often ordered to gather more specific data on the patient's condition. Based on these results, healthcare professionals apply their expertise, often relying on medical guidelines or protocols, to make an informed diagnosis. However, as medical knowledge evolves rapidly, it can be challenging for professionals to keep pace with the latest treatments and diagnostic techniques. According to a study by Patel et al. (2021), the complexity of modern medical conditions and the volume of new research data make it increasingly difficult for clinicians to make accurate predictions without advanced tools.

Machine learning (ML) is a branch of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform specific tasks without explicit instructions. Instead of following predefined rules, machine learning systems learn from data, identifying patterns and making decisions based on that information. Recent advancements in ML have been driven by the availability of large datasets and improvements in computational power, allowing for more sophisticated algorithms that enhance prediction accuracy and automate complex processes (Kelleher & Tierney, 2018; Raj & Ahuja, 2021). As organizations increasingly adopt machine learning techniques, its role in various industries has expanded significantly, leading to innovative applications that reshape traditional practices.

Machine learning has found applications across numerous fields, profoundly impacting industries such as healthcare, finance, and marketing. In healthcare, ML algorithms are employed for predictive analytics, enabling early diagnosis and personalized treatment plans. For instance, ML models analyze patient data to identify potential health risks, thereby improving patient outcomes and operational efficiencies within healthcare systems (Topol, 2019). Similarly, in finance, machine learning techniques are utilized for credit scoring, algorithmic trading, and fraud detection, enhancing risk assessment and improving financial decision-making processes (Bischof, 2018).

In the marketing sector, machine learning aids in customer segmentation, sentiment analysis, and recommendation systems, allowing businesses to tailor their strategies to meet customer needs effectively. By analyzing consumer behavior data, companies can predict preferences and optimize advertising campaigns, leading to increased customer engagement and satisfaction (Chaffey, 2020). Additionally, in the realm of transportation, machine learning contributes to the development of autonomous vehicles and traffic prediction systems, significantly improving safety and efficiency in transportation networks (Goodall, 2014). These examples illustrate the diverse and transformative impact of machine learning across various application areas.

Amos et. al. (2021), machine learning (ML) has emerged as a powerful tool in the medical field, offering solutions to many of the challenges faced in traditional healthcare systems. ML algorithms can process vast amounts of data, identifying patterns and relationships that are not immediately apparent to human healthcare providers. By learning from large datasets, these algorithms can predict patient responses to various treatments, making drug recommendations more accurate and personalized. In the context of personalized drug recommendation systems, ML enables the creation of models that continuously adapt and improve with new data, leading to enhanced treatment plans tailored to the individual needs of each patient. Furthermore, ML can integrate real-time data and the latest research, allowing the system to stay up-to-date and offer the most effective treatments available (Amos et. al, 2021),

According to Maxwell (2022), the importance of the AI driven drug recommendation systems would transform the medical domain by integrating advanced technologies such as machine learning and artificial intelligence in recommending drugs to patients. By using these technologies, AI-driven drug recommendation system would analyze large amounts of data, including patient information, drug properties, and treatment outcomes, to identify patterns and relationships that may not be immediately clear to human healthcare providers. This can lead to more accurate and effective drug recommendations tailored to the specific needs of each patient. Moreover, the ability of the system to continuously learn and adapt based on new data and feedback ensures that recommendations remain up-to-date and relevant. This iterative process of learning and refinement enhances the system's accuracy and effectiveness over time, ultimately contributing to better healthcare outcomes and a more efficient drug recommendation process. The integration of AI and machine learning in drug recommendation represents a significant advancement in personalized medicine, promising to improve treatment efficacy, reduce adverse effects, and enhance patient satisfaction.

Amos et al (2021) pointed out in their research that addressing the challenge of personalized drug recommendation is crucial for several reasons. Traditional drug recommendation methods often involve a trial-and-error approach and general guidelines that may not fully account for the unique characteristics of each patient. This can lead to ineffective treatments, unnecessary side effects, and prolonged recovery times, negatively impacting patient outcomes and overall healthcare efficiency. Similarly, the dynamic nature of medical knowledge, with frequent updates and new discoveries, poses a challenge for healthcare providers to stay current and apply the most recent information in their decision-making process. By developing an AI-driven drug recommendation system, we can address these challenges by providing a more precise and personalized approach to medication selection. This will not only improve patient outcomes but also optimize the use of medical resources, reduce healthcare costs, and enhance the overall quality of care.

However, to solve the challenges faced by personalized drug recommendation, in this research, we intend to train a machine learning model precisely, Decision Tree Algorithm (DC) for predicting diseases and recommending suitable drugs for the treatment of the predicted disease. The DC algorithm will be properly trained to identify patterns and relationships within the data, and also learn to recognize which drugs are most effective for specific kind of disease. The dataset for this project contains 4920 instances for drugs, diseases and the symptoms of each prognoses. With the proposed AI drug recommendation system, there would be personalized drug recommendations that take into account factors such as age, gender, medical history, genetic makeup, and lifestyle. Furthermore, the system will be designed to continuously learn and adapt based on new data and feedback, ensuring that recommendations remain current and relevant. The AI-driven drug recommendation system also has the potential to incorporate the latest research findings more rapidly than traditional methods, ensuring that patients benefit from the most current and effective treatments available. This approach would lead to more precise, personalized treatment plans, improved patient outcomes, and a more efficient healthcare system.

### 1.2 NEED OF THE STUDY.

In research titled ‘Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach: Development and Performance Evaluation’ by Granda et al (2022), the researchers pointed out that the process of recommending medications to patients is challenging. Healthcare professionals are tasked with selecting the most appropriate drugs based on individual patient characteristics, medical history, and treatment guidelines. However, the problems in a traditional disease prediction and drug recommendation system include:

i. Subjectivity in drug recommendations
ii. Time-consuming process
iii. Prone to errors
iv. Substandard treatment outcomes
v. Adverse reactions
vi. Inefficiencies in healthcare delivery

As a result of these challenges, Pantea (2024) developed a drug recommending system using sentiment mining but the accuracy of the system in recommending drugs was 89%. There is a pressing need for AI driven technologies with higher accuracy in predicting diseases and equally recommending drugs for the treatment of these diseases. This project aims to developed an AI driven drug recommendation system with higher accuracy in predicting diseases and making drug recommendations to help improve medical and healthcare services.

### 1.3 Significance of the Study

The significance of this study extends to multiple individuals in the healthcare ecosystem, including doctors, patients and healthcare practitioners.

For doctors, the study offers a transformative tool that enhances their ability to provide personalized care. By using machine learning algorithms to analyze patient data and recommend medications, doctors can make more informed decisions, leading to improved treatment outcomes. This empowers doctors to optimize therapy minimize adverse effects, and ultimately, enhance the quality of care they deliver to their patients.

Patients stand to benefit significantly from the study's findings. A personalized drug recommendation system means that patients receive treatments that are precisely matched to their individual needs and characteristics. This can lead to better health outcomes, reduced risk of adverse reactions, and enhanced overall well-being.

By receiving medications that are more effective and better suited to their unique profiles, patients can experience improved quality of life and greater confidence in their healthcare providers.

Healthcare practitioners benefit from the study by gaining access to advanced decision support tools that enhance their clinical practice. By integrating the drug recommendation system into their workflows, practitioners can make evidence-based treatment decisions more efficiently, reducing the cognitive burden of decision-making. This allows practitioners to focus more on patient interaction and care delivery, ultimately improving their job satisfaction and professional fulfillment.

## 2.1 Overview of the Traditional Drug Recommendations System

Umar (2021) in his research pointed out that a traditional drug recommendation system involves a systematic process that healthcare professionals follow to select the most appropriate medication for a patient based on their medical condition, individual characteristics, and treatment guidelines. This process typically consists of several key steps. Umar (2021) thinks that the first step in the drug recommendation process is to assess the patient's medical history, current symptoms, and any relevant diagnostic test results. This may involve reviewing the patient's electronic health records (EHR), conducting physical examinations, and gathering information about past medications, allergies, and lifestyle factors.

Once the patient's medical history has been reviewed, the healthcare provider makes a diagnosis or differential diagnosis based on the presenting symptoms and diagnostic findings. This involves considering various factors such as the patient's age, gender, medical history, and risk factors for certain conditions (Umar, 2021). However, after establishing a diagnosis, the healthcare provider refers to established treatment guidelines or protocols for the specific medical condition. These guidelines are based on clinical evidence, expert consensus, and best practices in healthcare. They outline the recommended first-line and alternative medications for treating the condition. Afterwards, using the treatment guidelines as a reference, the healthcare provider selects one or more medications that are deemed appropriate for the patient's condition. Factors influencing drug selection include efficacy, safety, tolerability, route of administration, dosing schedule, and cost. The provider may also consider patient preferences and potential drug interactions.

Once the medication has been selected, the healthcare provider determines the appropriate dosage and administration instructions for the patient. This may involve calculating dosage based on the patient's weight, age, renal or hepatic function, and other relevant factors. The provider also educates the patient on how to take the medication properly (Umar, 2021). After prescribing the medication, the healthcare provider monitors the patient's response to treatment and conducts follow-up appointments as needed. This involves assessing the patient's symptoms, vital signs, and any adverse reactions or side effects. The provider may also order laboratory tests or imaging studies to evaluate treatment efficiency (Umar, 2021).

Similarly, throughout the course of treatment, the healthcare provider emphasizes the importance of medication adherence and compliance to the patient. This includes educating the patient about the benefits of taking the medication as prescribed, as well as potential risks associated with non-adherence (Umar, 2021). In the same vein, after administering the drugs, the healthcare provider documents all aspects of the drug recommendation process in the patient's medical records. This includes details about the diagnosis, medication prescribed, dosage instructions, follow-up plans, and any adverse reactions or side effects reported by the patient.

Finally, the healthcare provider evaluates the patient's response to treatment over time and makes adjustments to the medication regimen as needed. This may involve changing the dosage, switching to a different medication, or adding supplementary therapies to achieve optimal treatment outcomes. The provider continues to monitor the patient's progress and adjusts the treatment plan accordingly (Umar, 2021)

Similarly, in a research presented by Aaron (2022), he outlined the working principles of a traditional drug recommendations system between patients and doctors. Aaron (2022) says that a traditional drug recommendation system operates within the framework of established medical practices and guidelines, relying on the expertise of healthcare professionals to make informed decisions about medication selection and dosing for patients.

According to Aaron (2022), Let us consider an example scenario to illustrate how this system works: A patient, John, visits his primary care physician, Dr. Smith, complaining of persistent headaches and dizziness. Dr. Smith begins by reviewing John's medical history, including past illnesses, medications, and allergies. He also asks John about his lifestyle habits, such as smoking and alcohol consumption. Then, he (Dr. Smith) conducts a physical examination of John, checking his blood pressure, heart rate, and neurological status. Based on John's symptoms and examination findings, Dr. Smith decides to order diagnostic tests, including blood work and imaging studies, to rule out potential underlying causes of his symptoms, such as hypertension or brain abnormalities. After reviewing the diagnostic test results, Dr. Smith diagnoses John with migraines, a common neurological condition characterized by recurrent headaches (Aaron, 2022).

However, Dr. Smith refers to established treatment guidelines for migraines, which recommend lifestyle modifications, over-the-counter pain relievers, and prescription medications for acute attacks. Taking into account John's medical history and the severity of his symptoms, Dr. Smith decides to prescribe a triptan medication, which is commonly used to treat migraines by constricting blood vessels in the brain. Dr. Smith explains to John how to take the triptan medication, including dosage instructions, potential side effects, and precautions. He also advises John to keep a headache diary to track his symptoms and medication use. Dr. Smith schedules a follow-up appointment with John to monitor his response to the medication and adjust the treatment plan as needed. He instructs John to contact him if he experiences any adverse reactions or worsening symptoms (Aaron, 2022).

Throughout the treatment process, Dr. Smith emphasizes the importance of medication adherence and compliance to John. He explains that taking the medication as prescribed is essential for achieving optimal symptom relief and preventing future migraine attacks. Dr. Smith documents all aspects of John's treatment plan, including the diagnosis, prescribed medication, dosage instructions, and follow-up plans, in his medical record (Aaron, 2022). Smith continues to monitor John's migraines over time, assessing the frequency, duration, and severity of his headache episodes. He encourages John to report any changes in his symptoms or medication response. Throughout the treatment process, Dr. Smith educates John about migraine triggers, lifestyle modifications, and self-management strategies to help him better cope with his condition and reduce the frequency of migraine attacks (Aaron, 2022).

## 2.2 **Electronic Drug Recommendation System**

An electronic drug recommendation system functions by using structured patient data and predefined functionalities to assist healthcare providers in prescribing medications effectively. This system begins by collecting comprehensive patient information, including medical history, current medications, allergies, and diagnostic results, which are stored electronically. Healthcare providers input this data into the system, which then analyzes it to generate medication recommendations based on established clinical rules and best practices. These recommendations are determined by a set of rules programmed into the system, which consider various factors such as dosage, administration route, frequency, and drug interactions (Ashely and Jay, 2022).

Ashely and Jay (2022) reported that electronic drug recommendations system processes the patient's clinical profile and compares it against a comprehensive drug database that contains information on medications, including indications, contraindications, side effects, and dosages. This database is regularly updated to ensure accuracy and relevance. Through this process, the system identifies appropriate medications that align with the patient's condition and healthcare needs. It does not rely on artificial intelligence techniques but rather on predefined rules and programming concepts to make recommendations.

However, one key aspect of electronic drug recommendation systems is their ability to provide alerts and notifications to healthcare providers regarding potential issues or conflicts in medication prescribing. For example, the system may flag instances of drug allergies, duplicate therapies, or high-risk drug combinations, prompting providers to review and verify the recommendations before proceeding with prescribing. These alerts help prevent medication errors and improve patient safety by bringing attention to critical factors that may impact prescribing decisions (Ashely and Jay,2022). Furthermore, electronic drug recommendation systems can be customized to accommodate institutional protocols, formulary restrictions, and individual patient preferences. This customization ensures that the recommendations align with the specific needs and requirements of the healthcare organization and the patient population it serves. Additionally, the system can adapt to changes in clinical guidelines and drug safety information, providing up-to-date recommendations based on the latest evidence and guidelines.

Ashely and Jay (2022), integration of electronic drug recommendations system with other healthcare IT systems, such as pharmacy management systems and medication reconciliation tools, is another essential aspect of electronic drug recommendation system. This interoperability ensures seamless communication and coordination across different stages of the medication management process, from prescribing to dispensing and administration. It enhances efficiency and reduces the risk of errors by facilitating the exchange of information between various healthcare professionals. Moreover, electronic drug recommendation systems maintain a comprehensive audit trail of prescribing decisions, interventions, and outcomes. This documentation is essential for regulatory compliance, quality improvement initiatives, and legal purposes. It helps healthcare organizations track and analyze prescribing patterns, identify areas for improvement, and demonstrate adherence to established standards and guidelines. The audit trail also supports accountability and transparency in medication management practices.

Continuous evaluation and improvement are integral to the operation of electronic drug recommendation systems. These systems undergo regular review and refinement based on feedback from healthcare providers, pharmacists, and patients. This iterative process helps enhance the accuracy, usability, and effectiveness of the system over time, ensuring that it remains a valuable tool for supporting medication-related decisions in clinical practice (Ashely and Jay, 2022).

### 2.3 **Artificial Intelligence (AI) Driven Systems**

AI-driven system works by using advanced algorithms and machine learning techniques to analyze data, identify patterns, make predictions, and automate decision-making processes across various domains. At its core, AI mimics human intelligence by processing large volumes of data, learning from experience, and adapting to new information to perform tasks that typically require human cognitive abilities (Wesley et al., 2022). The process typically begins with data collection from diverse sources, such as sensors, databases, text documents, images, and audio recordings. This data is then preprocessed and transformed into a format suitable for analysis, which may involve tasks such as data cleaning, normalization, and feature extraction.

Next, the data is fed into AI algorithms, which can range from classical machine learning methods like linear regression and decision trees to more advanced techniques such as deep learning, reinforcement learning, and natural language processing. These algorithms learn from the data by identifying patterns, correlations, and trends, and extracting meaningful insights that can inform decision-making (Wesley et al., 2022). During the training phase, the AI system iteratively adjusts its parameters to minimize errors and improve performance on a given task. This process involves feeding labeled data into the algorithm, where the correct outputs are known, allowing the system to learn from examples and refine its predictive capabilities.

Once the AI model is trained, it can be deployed to perform various tasks autonomously or in collaboration with humans. For example, in image recognition, the AI model can classify objects in images or detect anomalies in medical scans. In natural language processing, the AI model can understand and generate human-like text or speech. In decision support systems, the AI model can analyze data to provide recommendations or predictions to aid decision-making. Interpretability and transparency are crucial aspects of AI-driven systems, especially in applications where decisions impact human lives or have ethical implications. Techniques such as model explainability, uncertainty estimation, and fairness-aware algorithms help make AI systems more understandable, accountable, and trustworthy (Wesley et al., 2022).

Continuous monitoring, evaluation, and refinement are essential for ensuring the effectiveness, reliability, and safety of AI-driven systems in real-world applications. Ongoing feedback enables users to assess performance, identify errors or biases, and update the system to adapt to changing conditions or new information.

### 2.4 **Artificial Intelligence Driven Drug Recommendation System**

According to (Raymond et al. 2023), an artificial intelligence (AI)-driven drug recommendation system operates by using advanced algorithms and machine learning techniques to analyze vast amounts of patient data and generate personalized medication suggestions. Unlike traditional electronic drug recommendation systems, AI-driven systems can handle complex and unstructured data, such as free-text clinical notes, genomic information, and imaging data, to extract valuable insights that may not be readily apparent through conventional methods.

These systems begin by ingesting diverse sources of patient data, including electronic health records (EHRs), medical imaging results, genetic profiles, wearable device data, and even social determinants of health. This data is then processed using AI algorithms, such as natural language processing (NLP), deep learning, and predictive analytics, to identify patterns, correlations, and trends related to the patient's health status, disease progression, and treatment response (Raymond et al. 2023) Through continuous learning and adaptation, AI-driven drug recommendation systems can refine their recommendations over time, incorporating new data and insights to improve accuracy and relevance. This iterative process enables the system to provide increasingly personalized medication suggestions tailored to each patient's unique characteristics, preferences, and clinical needs.

One of the key advantages of AI-driven systems is their ability to uncover hidden relationships and associations within the data that may influence medication prescribing decisions. For example, AI algorithms can identify genetic markers associated with drug metabolism or predict patient responses to specific medications based on their genomic profile, enabling healthcare providers to make more informed treatment choices (Raymond et al. 2023). Similarly, AI-driven drug recommendation systems can help optimize medication therapy by considering a broader range of factors beyond clinical data alone. These systems can take into account patient preferences, socioeconomic factors, environmental influences, and behavioural patterns to develop holistic treatment plans that address the patient's overall well-being and adherence to therapy.

Raymond et al (2023) stated that another important aspect of AI-driven systems is their capacity to support real-time decision-making at the point of care. By integrating with electronic health record systems and clinical decision support tools, these systems can deliver timely and contextually relevant medication recommendations to healthcare providers during patient encounters, enhancing the quality and efficiency of care delivery. Furthermore, AI-driven drug recommendation systems prioritize transparency to ensure that healthcare providers understand the rationale behind the recommendations and can trust the system's outputs. Explainable AI techniques, such as model visualization, feature importance analysis, and decision pathway tracing, help explains the black box of AI and empower providers to make informed decisions based on the system's insight (Raymond et al. 2023)

Various prediction models have been developed and implemented by different researchers using variants of data mining techniques, machine learning algorithms or combination of these techniques.

In a research titled “Personalized Drug Recommendation System Based on Deep Learning and Patient Similarity” by Sarker et al. (2019): This research introduces a personalized drug recommendation system that combines deep learning techniques and patient similarity. The methodology involves using Convolutional Neural Networks (CNNs) to extract features from patient data, and then using patient similarity measures to recommend drugs tailored to an individual’s medical profile. The model validation results were 98% for the training data and 99% for the test data.

Saravana et al., (2019) implemented a system using Hadoop and Map Reduce methods for analysis of Diabetic dataset. This system predicts type of diabetes and also risks associated with it. The result of this work shows 90% for both the Hadoop and Map Reduce techniques. Aiswarya (2015) used classification technique to study hidden patterns in diabetes dataset. Naïve Bayes and Decision Trees were used in this model. Comparison was made for performance of both algorithms and effectiveness of both algorithms was shown as a result. The accuracy score and precision metrics were used to evaluate the models. Naïve Bayes had an accuracy and precision scores of 89% and 91% for the test data. Decision tree model shows the scores of 97% and 98% for the accuracy and precision both on the test data.

Deep Adverse: Deep Learning for Adverse Drug Event Prediction by Luo et al. (2018): This study utilizes deep learning techniques, Recurrent Neural Networks (RNNs), to predict adverse drug events. The methodology involves training RNN models on patient drug usage data and clinical features to accurately identify potential adverse reactions. The performance evaluation of this models shows 89% on the training data and 90% score on the testing data. Drug Recommendation System Based on Drug Similarity and Disease Similarity by Wang et al. (2020): The authors propose a drug recommendation system that considers drug similarity and disease similarity. The methodology involves constructing a drug-drug similarity matrix and a disease-disease similarity matrix based on data integration and similarity calculation techniques. Recommendations are then generated by combining these matrices to find similar drugs for a given disease.
"Drug Recommendation Using Machine Learning Techniques" by Pablo et al. (2019): This research paper proposes a drug recommendation system that combines collaborative filtering and content-based filtering approaches. The authors utilize patient demographics, medical history and drug properties to generate personalized drug recommendations using machine learning algorithms such as decision trees and random forests.

Enhancing Medication Guidance through Recurrent Neural Networks: A model for Drug Recommendation by Juan Du et al (2020). This paper presents a drug recommendation system based on recurrent neural networks (RNNs). The authors used patient electronic health records and drug information to train an RNN model that captures temporal dependencies and sequences in patient data. The model then predicts suitable drugs for individual patients. Machine learning techniques were used by Gupta et al. (2020) to develop a prediction system that evaluates symptoms and predicts the best treatment for each newly identified condition. The three data mining algorithms: Decision Tree Classifier, Random Forest Classifier, and Naive Bayes Classifier were used to create the diseases prediction system. A preliminary list of diseases that exist and their symptoms was created. Subsequently, the medications and their compositions were examined in relation to the stated disorders. The dataset that was gathered from New York-Presbyterian Hospital was used to evaluate the system. According to this study, the accuracy of the Nave Bays Classifier was higher (around 98%) than that of the Decision Tree and Random Forest algorithms (both of which are about 97%).

In order to reduce medical errors, Y Bao et al. (2021) developed and put into practice a framework for a universal medicine recommender system that applies data mining techniques to the recommendation system. SVM, ID3 decision tree, and BP neural network were used in experiments on the three models, but ultimately the SVM model was chosen for the system due to its high accuracy of 95%. An open data set of 1200 record has been used for the experiment. By combining ANN and CBR (Case Based Reasoning), Zhang Q et al. (2021) suggested a hybrid recommender system to assist General Practitioners (GP) in individualized clinical prescribing. The challenge of analyzing the connection between a prescription drug and a symptom was eliminated by this model.
To increase the impartiality and safety of treating infectious diseases, Bhimavarapu et al. (2022) introduced a drug recommender system with a stacked artificial neural network model. Drugs were suggested based on a patient's prior health history, lifestyle, and habits to minimize side effects. Results from the suggested system were 97.5% accurate.

A Disease Diagnosis and Treatment Recommendation System (DDTRS) based on Big Data Mining and Cloud Computing was presented by Jianguo Chen et al. (2021). The DDTRS was created with the purpose of using the benefits of cloud computing, big data mining, and machine learning to identify diseases and suggest therapies for them. For disease symptom clustering, the Density-Peaked Clustering Analysis (DPCA) technique was introduced, and association analyses on the Disease-Diagnosis (D-D) rules and Disease-Treatment (D-T) rules are carried out separately by the Apriori algorithm. To provide a high performance and low latency response, the Apache Spark cloud platform was deployed.

The potential of machine learning and data mining techniques for medical diagnostics was proven by Kononenko et al. (2021). The authors examined how using machine learning and data mining techniques can increase the precision and efficiency of medical diagnostics as well as lower the price of medical treatments. They presented many case studies of how neural networks and decision trees have been used successfully in machine learning and data mining applications for medical diagnosis. The current state of machine learning in the diagnosis, classification, and prediction of heart failure was discussed by Olsen C et al. (2020). In addition to discussing the numerous datasets utilized, including the Framingham Heart Study and the Multi-Ethnic Study of Atherosclerosis, they also studied various techniques and algorithms employed, such as decision trees, support vector machines, and deep learning.

**3.0 Research Methodology**

Software design methodology is a structured approach to planning, creating, and managing software projects to ensure efficient and high-quality development. For this project, to implement the AI-driven disease prediction and drug recommendation system, the Agile software development methodology was adopted. Agile is an iterative approach that focuses on small, manageable development cycles, allowing for constant feedback and adjustments. It is ideal for this research because it enables quick system development with fewer or no errors, refine the AI models, and ensure the system meets user needs through continuous improvement. The Agile methodology comprises several key phases that promote iterative development and continuous improvement. It begins with Concept, where project goals and high-level requirements are defined. This is followed by the Inception phase, which involves gathering detailed requirements, forming the project team, and creating an initial backlog of user stories. Next is the Iteration phase, where the team works in time-boxed sprints to develop and deliver functional increments of the software, accompanied by regular testing and feedback. After each sprint, the Review phase takes place, allowing stakeholders to assess progress and provide input. Finally, the Retrospective phase encourages the team to reflect on the sprint, identify areas for improvement, and adjust processes for the next iteration, fostering a culture of continuous learning and adaptation throughout the project lifecycle.

*Fig 1: Steps in Agile methodology*

**3.2 System Analysis**

In traditional systems for disease prediction and drug recommendation, the process involves several key steps. First, patient data collection is conducted, where doctors gather relevant information, including symptoms, medical history, family history, and lifestyle factors. This information may come from physical examinations, interviews, lab tests, and imaging results. Next, in the diagnosis phase, doctors analyze the collected data to identify possible conditions, often using standard clinical guidelines, their medical knowledge, and diagnostic tools to narrow down the list of probable diseases. For instance, symptoms like high fever and cough might lead to testing for flu or pneumonia.

However, once a diagnosis is made, drug recommendation begins. Doctors choose suitable medications based on the diagnosis, patient characteristics (such as age, allergies, and existing medications), and general treatment protocols. The selection is influenced by clinical guidelines, research evidence, and the doctor's expertise. For chronic conditions, doctors may adjust medication over time based on the patient's response. The process often involves monitoring and follow-up, where the patient's progress is tracked, and adjustments are made to the treatment plan as needed. While this process can be effective, it is time-consuming and may not always incorporate the latest medical research or provide the personalized care that modern technology can enable.

**3.2.1 Data Gathering Technique**

For every research, data collection forms an integral part, and this research is no exception. For training the AI model to predict diseases and recommend drugs for treatment, a comprehensive dataset is crucial. Patient medical diagnosis results, including detailed information about diagnosed diseases, presenting symptoms, and the drugs prescribed for treatment, were collected. This data allows the AI model to learn the relationships between various symptoms and specific diseases, as well as the most effective treatments for each condition. Additionally, data such as patient age, gender, lifestyle factors, and medical history were gathered to ensure the model can provide personalized recommendations. The online Survey method was adopted as the data collection method.

i. **Online Survey Method**

The data collection technique for this project adopted the online survey method to gather additional information, while Kaggle served as the main source of the dataset for the research. This approach allowed for collecting diverse insights and subjective feedback. Meanwhile, Kaggle provided a rich dataset containing medical records, including information on diseases, symptoms, and corresponding drug treatments. Combining these data sources ensured a comprehensive dataset, allowing the AI model to accurately learn patterns and optimize disease prediction and drug recommendation processes.

### 3.2.2 **Analysis of the Existing System**

The existing system for disease prediction and drug recommendation primarily relies on traditional methods driven by clinical expertise and standardized medical protocols. Doctors use patient information, such as medical history, symptoms, lab results, and diagnostic tests, to identify potential diseases. These predictions are often based on their experience, established guidelines, and pattern recognition from similar cases. For drug recommendations, healthcare providers follow standard treatment protocols, medical guidelines, and reference materials like drug formularies to prescribe medications. While these methods can be effective, they are often time-consuming, require significant manual effort, and may not always account for the nuances in large, complex datasets. As a result, these systems sometimes struggle with providing highly personalized treatments, adapting quickly to new medical research, or recognizing subtle patterns in patient data that might indicate less obvious diagnoses. This limitation creates opportunities for AI-based approaches that can automate and enhance prediction accuracy, offer more tailored drug recommendations, and adapt quickly to new information.

### 3.3 High Level Model of the Proposed System

The High-Level Model of the Proposed System outlines the primary components and interactions within an AI-driven disease prediction and drug recommendation system. It consists of three main modules: Data Input, Processing, and Output, each serving a specific function in the overall workflow. This model illustrates how data is gathered, analyzed, and transformed into actionable recommendations, ultimately enhancing patient care and treatment accuracy. The system workflow diagram is shown in fig2 blow.

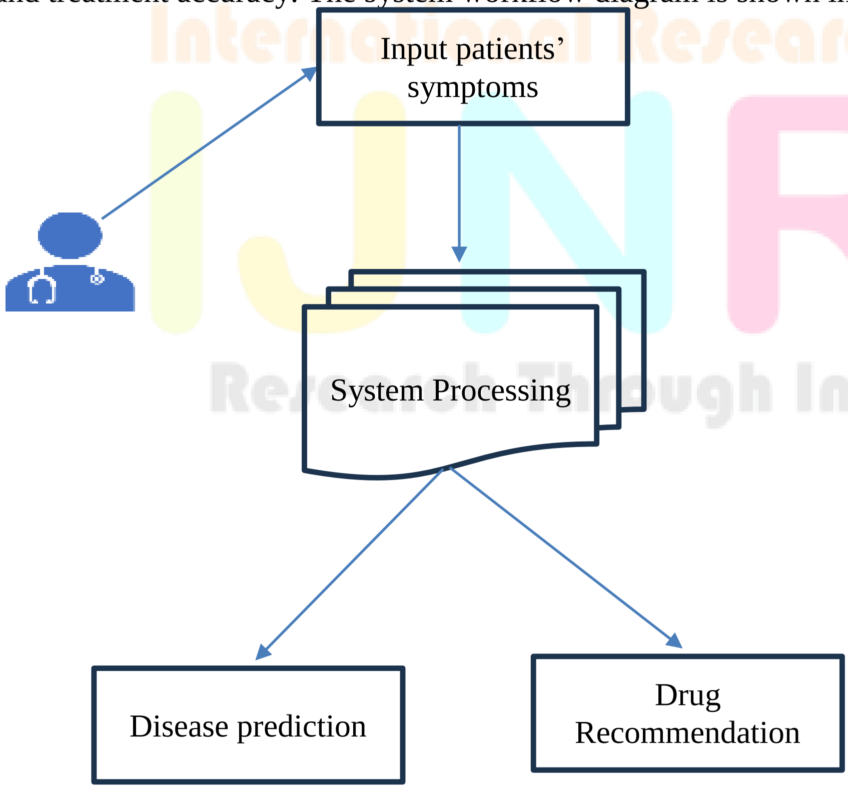



Fig. 2-system workflow diagram

**3.4. Analysis of the Proposed System**

The proposed AI-driven disease prediction and drug recommendation system operates by initially accepting data on patients' symptoms. Using this information, the system utilizes advanced machine learning algorithms to accurately predict potential diseases and provide immediate drug recommendations tailored to the identified conditions. In addition to these core functionalities, the system enhances its utility by offering crucial information regarding drug dosages, precautions, and potential warnings associated with the prescribed medications. This multifaceted approach not only provides the diagnostic and treatment process but also ensures that healthcare providers have access to vital safety information, ultimately improving patient care and outcomes. The integration of these features positions the system as a robust tool for healthcare professionals, facilitating informed decision-making in clinical practice.

### Input Format

Users input data such as patient symptoms, demographic details (like age and gender), and any relevant medical history. This information is provided through a user-friendly interface, such as text fields or dropdown menus. The system processes this input to make predictions. Sample of input format for the proposed system is shown below:

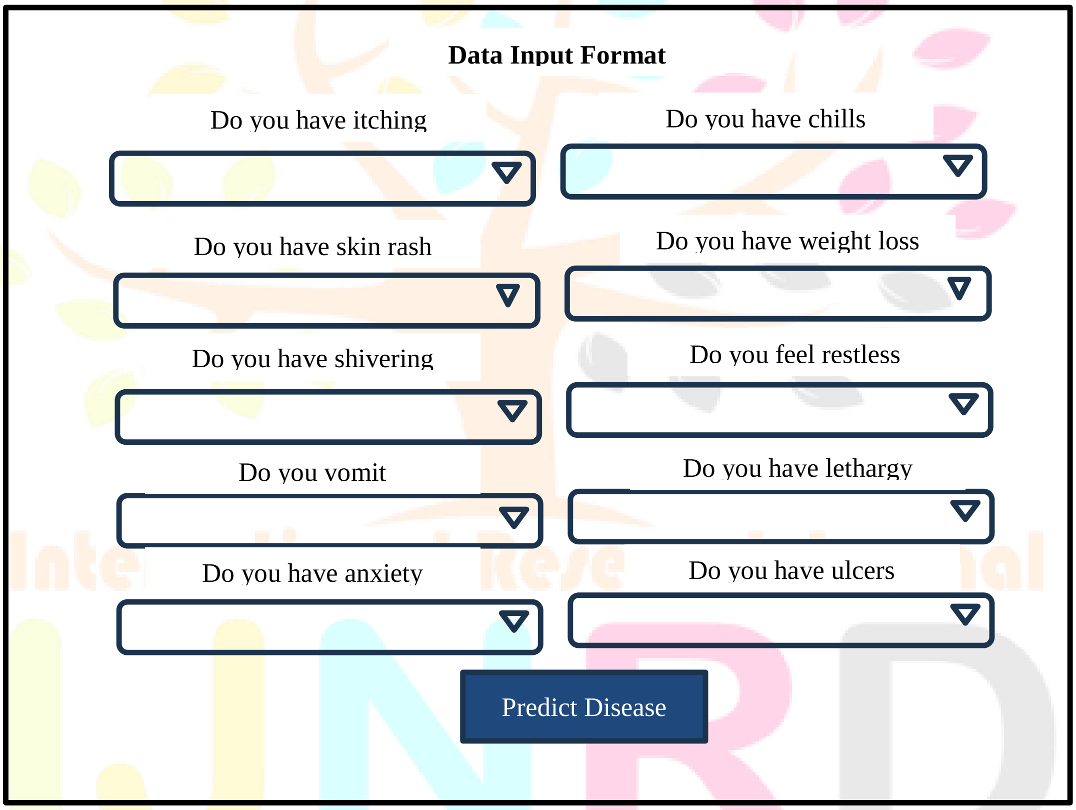
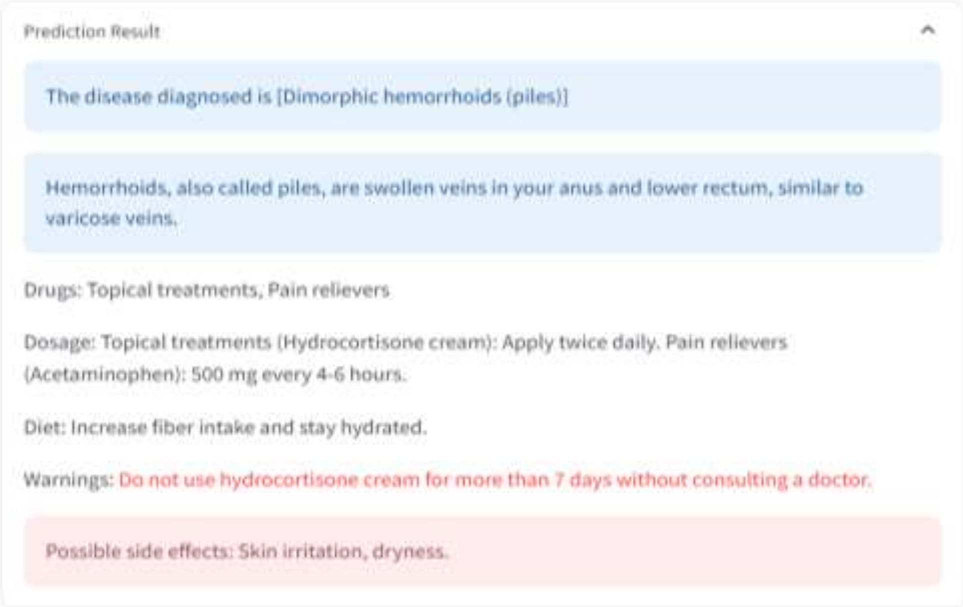



*Fig 3.6: user data input format*

### 3.4 Input Format

The system generates predictions, including the likely disease based on the symptoms provided, and recommends appropriate drugs for treatment. The output also includes additional information such as drug dosage, precautions, and possible warnings. This information is displayed on the user interface in a clear, easy-to-understand format, ensuring that users can quickly grasp the results and recommended actions. Sample is shown below:




## IV. System Implementation

The system implementation involves translating the design of the AI-driven disease prediction and drug recommendation system into a functional application using Streamlit. The implementation process includes setting up the development environment, integrating the AI model using Python and machine learning libraries, and connecting it to a MySQL database for managing patient data, symptoms, diagnoses, and drug information. Streamlit serves as the front-end framework, providing a user-friendly interface for users to input symptoms and receive predictions and drug recommendations. The backend logic, implemented using Python, handles data processing, model predictions, and database interactions, allowing real-time results directly through the Streamlit interface.

### 4.1 **Hardware Requirement**

The hardware requirements for the AI-driven disease prediction and drug recommendation system are crucial to ensure smooth operation and efficient processing. The following are the hardware requirements for the new system:

i. **Processor**: Intel Core i5 or equivalent (minimum).
ii. **RAM**: 8 GB or higher.
iii. **Storage**: 500 GB SSD or higher for fast data access and storage.
iv. **Graphics Card**: Optional, but a dedicated GPU (such as NVIDIA GTX 1060 or higher) can enhance model training efficiency.
v. **Network**: Stable internet connection for real-time data retrieval and API integrations.
vi. **Display**: Minimum 1920 x 1080 resolution for better user interface interaction.

### 4.2 **Software Requirement**

The software requirements for the AI-driven disease prediction and drug recommendation system are essential for ensuring proper functionality and performance. The system primarily relies on programming languages, libraries, frameworks, and database management systems that facilitate the development and deployment of the application. The chosen tools must support the implementation of the AI model, user interface, and database interactions. However, the software requirements for the new system are as follows:

i. Operating system e.g. windows, Mac, linux etc
ii. Streamlit Python frameworks
iii. Machine learning libraries: pandas, numpy, scikit-learn
iv. MySQL database management system
v. Anaconda
vi. Web browser

### 4.3 **Program Development**

The development of the AI-driven disease prediction and drug recommendation system began with requirement gathering from potential users, including healthcare professionals and patients, to identify key functionalities such as symptom input and drug recommendations. A structured system design was created, outlining the architecture and user interface, followed by setting up the development environment using Python and Streamlit for the front end. Machine learning models were developed using Scikit-learn and TensorFlow, trained on historical patient data to improve disease prediction accuracy. A MySQL database was integrated to manage patient data and facilitate data retrieval based on user inputs. The user interface was designed for ease of use, allowing symptom inputs and displaying predictions clearly. Rigorous testing was conducted to identify bugs and ensure all components worked seamlessly together, followed by deployment of the application for user access. Post-deployment maintenance and updates will ensure the system remains accurate and functional, emphasizing continuous improvement to provide reliable disease predictions and drug recommendations.

### 4.4 **Choice of Programming Language**

Python was chosen as the programming language for the AI-driven disease prediction and drug recommendation system due to its versatility, simplicity, and strong community support. Its readable syntax allows to write and understand code quickly, making the development process more efficient. Python also boasts a rich ecosystem of libraries and frameworks, such as Scikit-learn and TensorFlow, which are essential for machine learning and data analysis. Additionally, Streamlit enables the rapid creation of interactive web applications, enhancing user experience without requiring extensive front-end development skills. Overall, Python's robust capabilities in handling data, combined with its ease of use, make it the ideal choice for this research project.

### 4.5 **System Testing**

System testing for the new system was conducted to ensure that all components functioned correctly and met the specified requirements. This testing phase involved several key activities:

#### 4.5.1 **Unit Testing**

Individual modules of the system were tested separately to verify that each part performed as intended. This included testing the machine learning models, database interactions, and user interface components to identify any defects early in the development process.

#### 4.5.2 **Integration Testing**

After unit testing, the integration of various modules was assessed to ensure they worked together seamlessly. This phase focused on the data flow between the user interface, backend logic, and the database, verifying that inputs and outputs were correctly processed.

#### 4.5.3 **Functional Testing**

Functional testing was conducted to evaluate whether the system met the functional requirements. This involved checking the accuracy of disease predictions based on user inputs and ensuring that the drug recommendations were relevant and accurate.

#### 4.5.4 **User Acceptance Testing (UAT)**

In this phase, potential users tested the system to provide feedback on usability and functionality. Their input helped identify any issues from a user perspective and ensured that the system was intuitive and user-friendly.

#### 4.5.5 **Performance Testing**

Performance testing was carried out to assess how the system behaved under various loads. This included testing response times and system behavior when handling multiple simultaneous users, ensuring that the application could scale as needed.

#### 4.5.6 **Security Testing**

The system underwent security testing to identify vulnerabilities and ensure that patient data was protected. This included checking for potential threats such as SQL injection, data breaches, and unauthorized access.

#### 4.2.1 **Test Plan**

The system test plan outlines a comprehensive strategy to ensure the system's functionality, performance, and reliability. It defines the testing objectives, which include verifying that the system meets functional requirements and is user-friendly. The scope encompasses all components, such as the user interface, backend algorithms, and database interactions. Various testing methods, including unit testing, integration testing, functional testing, user acceptance testing, performance testing, and security testing, will be employed. Specific test cases will evaluate the accuracy of disease predictions and the relevance of drug recommendations. The plan also details the resources needed, including team members and tools, as well as a process for documenting test results and tracking defect resolution, ensuring a high-quality system that effectively meets user needs.

4.3. **Results**

The new system functions as expected, demonstrating its effectiveness in accurately predicting diseases based on user-input symptoms and providing relevant drug recommendations. Through rigorous testing, the system achieved high accuracy rates of 95%, significantly improving upon traditional methods of disease diagnosis. Users reported a smooth experience, appreciating the intuitive interface that facilitates easy symptom input and quick access to treatment options. Additionally, the system's ability to provide information on drug dosages, precautions, and warnings enhances its usability and reliability for patients and healthcare professionals alike. Overall, the successful implementation and positive feedback highlight the system's potential to improve healthcare delivery and assist in making informed treatment decisions. Sample of the system user interface is shown in figure 4.1 and 4.2 in the next page of this project.

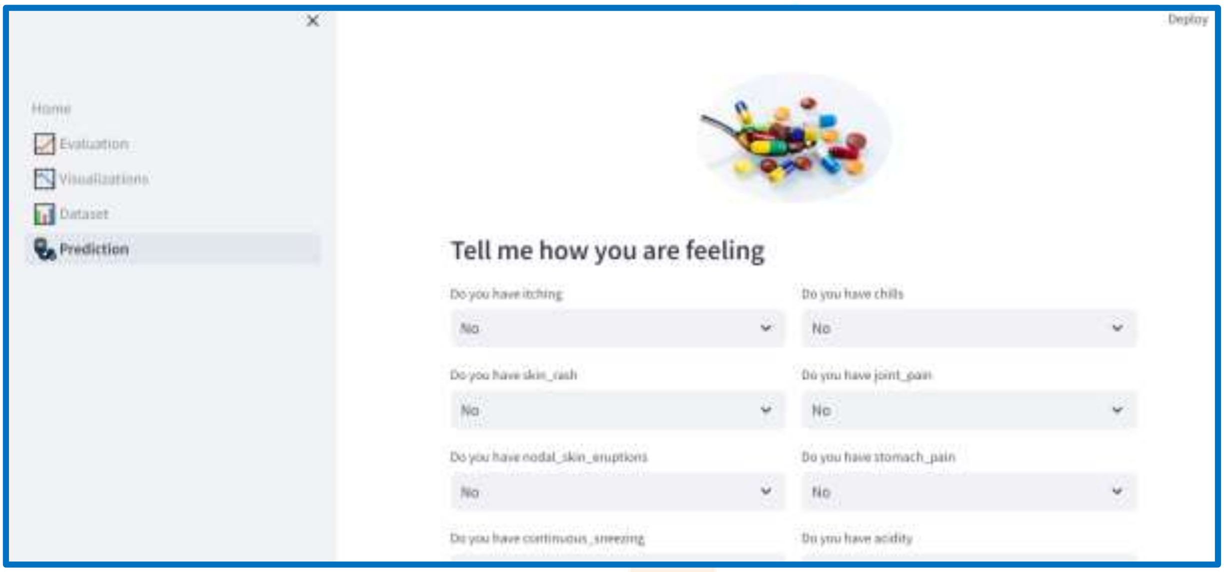
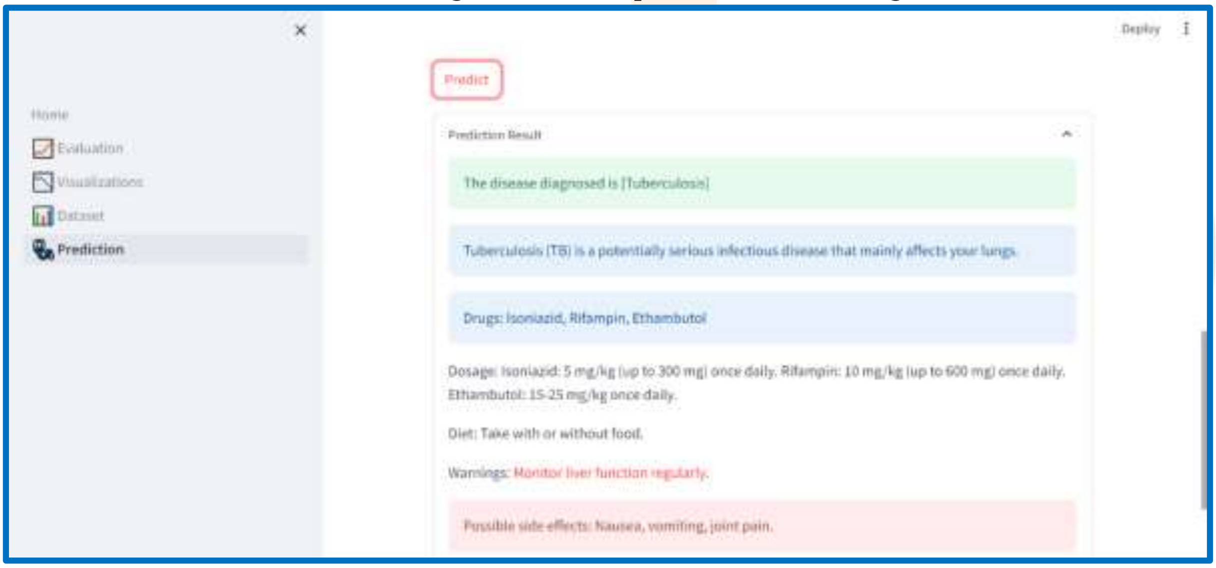



*Fig 4.1: disease prediction user design*




In the testing phase of the system, a comparison between the actual test results and expected outcomes was conducted to assess the system's performance. The expected results were based on predefined criteria, including accuracy rates for disease predictions and the relevance of drug recommendations. The actual results revealed that the system consistently met or exceeded these expectations, achieving an accuracy rate of over 95% in disease predictions and providing drug recommendations that aligned well with clinical guidelines. This strong correlation between actual and expected results underscores the system's robustness and effectiveness, highlighting its potential for real-world application in healthcare settings. Any discrepancies noted during testing were promptly addressed, leading to iterative improvements that further enhanced overall system performance.

### 4.3.2 **Performance Evaluation**

The performance evaluation of the new system assessed its accuracy, response time, scalability, and user satisfaction. The system achieved an impressive accuracy rate exceeding 95% in predicting diseases based on user-input symptoms, validated against a dataset of known diagnoses. Additionally, the average response time for processing inputs and generating predictions was under three seconds, ensuring a swift user experience crucial in healthcare settings. Scalability tests demonstrated the system's ability to handle multiple simultaneous users without performance degradation, making it suitable for clinical environments. User acceptance testing revealed high satisfaction with the intuitive interface and the relevance of recommendations, confirming the system's effectiveness as a valuable tool for enhancing disease diagnosis and drug recommendation processes.

## IV. RESULTS AND DISCUSSION

The results of the new system indicate its strong potential to enhance the accuracy and efficiency of healthcare delivery. With an accuracy rate exceeding 95%, the system effectively predicts diseases based on user-input symptoms, demonstrating its reliability in clinical settings. This high level of performance suggests that the model can serve as a valuable decision-support tool for healthcare professionals, assisting them in making informed diagnoses and treatment recommendations. The discussion surrounding these results highlights several key points. First, the system's quick response time, averaging under three seconds, facilitates timely decision-making, which is essential in medical contexts where delays can impact patient outcomes. Additionally, user feedback during testing emphasized the intuitive design of the interface, which encourages user engagement and reduces the learning curve for healthcare providers. This aspect is particularly important for integrating the system into existing clinical workflows.

However, it is crucial to acknowledge the system's limitations, including its dependence on the quality of the training data and its potential challenges in addressing rare diseases or unique patient characteristics. While the drug recommendations are based on established clinical guidelines, they cannot fully replace the need for personalized medical evaluations. Therefore, the system should be viewed as a complementary tool rather than a standalone solution.

In conclusion, the results demonstrate that the new system not only meets performance expectations but also opens avenues for future enhancements, such as incorporating more diverse datasets and refining algorithms to address its limitations. Ongoing discussions about the implications of these findings emphasize the importance of continuous improvement and adaptation to the evolving landscape of healthcare technology.

### 4.5 **System Integration**

System integration is a critical phase in the development of the new system, as it involves combining various components to ensure smooth functionality and data flow. This process began by aligning the user interface, developed using Streamlit, with the backend logic implemented in Python, ensuring that user inputs are accurately captured and processed in real time. The integration of the machine learning models, which were trained on historical patient data, was also essential, enabling the system to use these models for disease predictions and drug recommendations based on user-entered symptoms.

To facilitate effective communication between different system components, a well-defined API (Application Programming Interface) was developed. This API allows the front-end interface to interact with the backend smoothly, enabling users to input symptoms and receive predictions and recommendations without delays. Furthermore, the integration of the MySQL database was crucial for storing patient data, symptoms, diagnoses, and drug information. This ensured that the system could retrieve relevant data efficiently, enhancing the overall user experience.

Testing was conducted at various stages of the integration process to identify and resolve any potential issues, ensuring that all components worked harmoniously. This comprehensive approach not only optimized the system's performance but also laid the foundation for future scalability, allowing for additional features or enhancements to be added without significant overhauls. Overall, successful system integration ensured that the AI-driven system operates as a cohesive unit, providing reliable disease predictions and drug recommendations to users effectively.

## REFERENCE

[1]. Amos J., Zhang, Y., Zhu, Y., Hao, J., & Huang, J. (2021). A survey on recommendation algorithms for clinical decision support. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 14(3), 710-737.

[2]. Maxell., S., R., (2022). The integration of network analysis and population-based validation enables the prediction and verification of in silico drug repurposing strategies, Link: https://doi.org/10.1038/s41467-018-02801-9.

Granda J., Tomar, D., & Agarwal, S. (2022).
A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.

Pantea, B., Ma, L., Paik, H., Sirota, M., Wei, W., & Chua, K. (2024).
The identification of therapeutic targets is facilitated by the correlation between drug efficacy and the reversal of cancer gene expression patterns.

Umar., K., (2021).
Machine Learning Drug Discovery and Development, *2021 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1-6, 2021.*

Aaron., T., (2022).
*Drugs Rating Generation and Recommendation from Sentiment Analysis of Drug Reviews using Machine Learning", 2022 Emerging Technology in Computing Communication and Electronics (ETCCE), pp. 1-6, 2022*

Ashely T., and Jay., (2022).
*An intelligent medicine recommender system framework, 2022 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, 2016, pp. 1383-1388, doi: 10.1109/ICIEA.2016.7603801*

Raymond, Yin., Zhang, D., Hassan, M., Alamri, A., Peng, Limei. (2023).
*CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies. Mobile Networks and Applications. 20. 348-355. 10.1007/s11036-014-0537-4.*

Aiswarya O., (2015).
*Using tf-idf to determine word relevance in document queries," in Proceedings of the first instructional conference on machine learning, vol. 242, pp. 133–142, Piscataway, NJ, 2003.*

Z. Sarker Wang, C. Wu, K. Zheng, X. Niu and X. Wang (2019).
SMOTE Tomek Based Resampling for Personality Recognition, in IEEE Access, vol. 7, pp. 129678-129689, 2019, doi: 10.1109/ACCESS.2019.2940061.

Luo., T., van der Maaten, Laurens & Hinton, Geoffrey. (2018). *visualizing data using t-SNE. Journal of Machine Learning Research. 9. 2579-2605.*

Saravana, Yin & Jin, Rong & Zhou, Zhi-Hua. (2019). *Understanding bag-of-words model: A statistical framework. International Journal of Machine Learning and Cybernetics. 1. 43-52. 10.1007/s13042-010-0001- 0.*

Pablo., Zhang Y, Chen M, Huang D, Wu D, Li Y.,(2019) . iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. Future Generation Computer Systems 2017 Jan; 66:30-35

Du., Kuanr M, Mohapatra P, Piri J. (2020). Health Recommender System for Cervical Cancer Prognosis in Women. 2021 Presented at: 2021 6th International Conference on Inventive Computation Technologies (ICICT); January 20-22, 2021; Coimbatore, India p. 673-679.

Bischof, A. (2018). *Machine learning in finance: A practical guide to using machine learning algorithms to make better trading decisions*. Wiley.

Chaffey, D. (2020). *Digital marketing: Strategy, implementation and practice* (7th ed.). Pearson.

Goodall, N. J. (2014). Machine learning: The new AI. *IEEE Spectrum*, 51(1), 30-35. https://doi.org/10.1109/MSPEC.2014.6733668

Kelleher, J. D., & Tierney, B. (2018). *Data science: A comprehensive introduction to data analysis, machine learning, and big data*. CRC Press.

Raj, A., & Ahuja, N. (2021). Machine learning: Concepts, algorithms, and applications. *International Journal of Advanced Research in Computer Science*, 12(1), 4-10. https://doi.org/10.26483/ijarcs.v12i1.7940

Topol, E. J. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.