**UNIVERSITY OF EMBU**



**TITLE: AI-BASED SYMPTOM CHECKER SYSTEM**

**JULIUS MOYI LOMUKET**

**REG: B141/20076/2020**

**Report to be Submitted to the Department of Computing and Information Technology In**

**Particular Fulfilment of the Requirements for the Degree of Bachelor of Science In Information Technology**

**Date of Submission**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

# Declaration

I, Lomuket Julius Moyi, hereby declare that this project AI-Based symptom checker is my own original work, completed in fulfilment of the requirements for the Bachelor of Science in Information Technology at University of Embu.

I also assure that this project work is based on my own work research, and all sources of information or contributions from other individuals are acknowledged accordingly. In addition, this work has never been submitted or presented for any academic award or any other degree or qualification at this institution or any other institution.

Name Signature Date

……………………………………… …………………… ………………….

Supervisor’s Name Signature Date

……………………………………… …………………… ………………….

# Dedications

I dedicate this project to my family that is my father Samwel Lomuket, my mother Christine Lomuket and my brother Moses Arupe whose their unwavering support and encouragement have been instrumental throughout my academic journey.

I also dedicate this work to my supervisor Dr. Consolata Gakii, whose invaluable insight and knowledge have shaped my academic personal growth. I affirm that this project would not have been possible without their dedication and inspiration.

To my friends and colleagues who also stood by me, providing support and sharing in both struggles and successes, this achievement is ours.

# Abstract

This project presents the development of AI-Based symptom Checker targeting at providing preliminary health assessments to users based on their reported symptoms. Due to the rapid rising demands for accessible healthcare, especially in underserved places, this system is designed to assist those individuals in identifying key potential health issues and look for appropriate care. The primary objective is to improve diagnostic accuracy and reduce unnecessary medical visits, ultimately reducing healthcare burdens.

The system utilizes machine learning algorithms, specifically Random Forest algorithms, to analyze symptom input and generate initial diagnoses. Additionally, Natural Language Processing (NLP) is integrated to interpret symptom descriptions, allowing users to communicate symptoms naturally. The data that was used for model training was sourced from Kaggle website, followed by preprocessing, feature selection, and model validation to ensure reliability.

The expected results include increased diagnostic accuracy by at least 20% compared to existing tools, reduction in unnecessary medical visits by approximately 30%, and a valuable contribution to digital healthcare innovations. This solution holds potential to improve patient care accessibility and efficiency, establishing a foundation for future AI- driven diagnostic tools.

# Acknowledgement

I would like to express my sincere gratitude to everyone who has contributed to the successful completion of this project.

First and foremost, I am deeply appreciative to my supervisor, Dr Consolata Gakii, for her guidance, valuable insights, and encouragement throughout the research and development phases. Her expertise and constructive feedback have been instrumental in shaping this project.

I am also grateful to my instructors and lecturers, whose dedication and knowledge provided a strong foundation for this work. Their mentorship has been a source of inspiration and support in my academic journey.

Special thanks also goes to my family and friends, whose support encouragement and guidance kept me motivated through every challenge.

Finally, I would like to acknowledge my peers and colleagues for their collaboration. Working alongside such talented people has been a rewarding experience, and their feedback ideas have been improved this project.

Thank you to every individual who made this journey possible.

# CHAPTER ONE

# 1.0 The Background Information

In recent days, health sectors have been experiencing high technological advancement like for example Artificial Intelligence playing an increasingly important role in enhancing patient in final results. Some of these areas in healthcare where Artificial Intelligence has shown a great impact is the development of an intelligent system that can help in the diagnosis and management of health conditions. A study by Abideen (2022) show that in a report from the Word Health Organization, it says that Pakistanis have trouble getting health care because they do not have enough money. More than 60 million people live below the poverty line and cannot afford to be wrongly diagnosed, which is another big problem in the medical field. It also indicates that there are numerous treatments available for various diseases, but there is no place where a person can get the details of the diseases and their treatments. Among those intelligent systems that can perform these tasks of diagnosis and management of health condition is Artificial Intelligent- Powered symptom checker. A tool that has been purposely developed to examine patient’s symptoms and give initial health assessment thus assisting both the patient health providers

AI-Based symptom checker, this tool employs machine learning algorithms and natural language processing to be basically able to interpret individual inputs that are symptom description and then try to relate them against large medical databases. By doing this they can provide potential diagnoses then suggesting the next step to take and able to recommend if there is any further medical assessment or consultation is required. A study by Zain(2022) indicates that an update has been recently added to Google in which Google plays doctor by identifying your medical symptoms. One can easily ask Google for medical advice and search for certain symptoms. Instead of searching for a whole condition a person can search for a certain symptom like ‘my head hurts’, and Google will provide conditions, treatments and home remedies about it. Although this feature will massively benefit the public, Google still urges the fact that the person should seek medical advice as precaution and in case of serious situation in which the ailment itself is deadly.

This approach not only provides individuals with very quick insights but also alleviates control and reduces unnecessary visits for minor health conditions. Research by Harad et al,(2024) shows that the current symptoms checker-generators that produce differential diagnoses based on the user inputs from the patients themselves before they encounter a clinician are potentially promising tools to reduce diagnostic errors. It also demonstrated that the accuracy of the symptom checkers has been improved from 2015 to 2022.

The global medical health care system faces a lot of shortcomings like minimal access to medical professionals, long waiting queues and cost increase. Research by Abideen et al, (2022) stated that

Advanced Medication found saving of 10,000 per person, 29 percent of patient’s had less intense therapy, 15 percent reduction in patient’s medication. The work of Advanced Medication is aimed at decreasing the cost of medicine. Hence saving lives through technology will save billions of dollars.

Even with the promise of Artificial intelligence in health care still there are shortcomings to current symptoms checkers tools. Some of these systems fail to provide accurate results, for example some lack comprehensive data sets for example for a rare disease or in conditions that occur with non-specific symptoms. Also, there is a limitation based on patient privacy and ethical consequences of depending on Artificial Intelligence for very sensitive health for very sensitive health-related decisions. More techniques are needed to improve the accuracy of Artificial Intelligence-Based symptom checkers and make sure they are created in a manner that aligns with moral guidelines and meets the user’s requirements.

# 1.1 Problem of the statement

A symptom is a partial indication of a certain disease or a health condition that can be experienced by an individual

Various approaches or techniques have been implemented and used to provide users with initial assessment and recommendations.

Research conducted by Harada et al, (2024) proposed a symptom diagnosis generator that produces a differential diagnosis based on the inputs from patient’s themselves before they encounter a clinician which of course is a good approach. Another study by Abadeen et al, (2022) suggested E-Healthcare system whose major motivation is to reduce the rate of misdiagnosis and save time. The system eliminates all worries about the legitimacy and integrity of the results produced by implementing machine learning and AI approaches. However, this technique has a problem whereby the output of the symptom checker lucks quality because of the following, symptom interpretation, input sequencing and symptom listing that needs to be addressed to improve the quality of the results.

This study is proposing to come up with an Artificial Intelligence-Based symptom checker that had been trained using real-world patient data to enhance its accuracy symptom checker output.

# 1.2 Justification

As a result of increasing global demand for health care services has led to rapid development of innovative solutions to enhance patient care. Some of these advancements are like AI-Based symptom checker which have a potential to change the way people live by providing initial health assessment, especially those individuals who live in underserved places or regions where it is very difficult to access healthcare. By having the potential to reduce the problems experienced on the health care systems by simply giving patients more firsthand or immediate information concerning the health conditions, this advanced technology can play a critical role within the healthcare system by acting as critical tool for future healthcare delivery. Research conducted by Hammoud, Mohammed et al, (2024) suggest that every day, millions of people turn to the internet for health information and treatment advice. For instance, in Australia approximately 80% of people search the internet for health information and approximately 40% seek web-based guidance for self-treatment. It also showed that half of patients investigated their symptoms on search engines before visiting emergency rooms. This research is proposing to develop and validate an AI-Based symptom checker system that will be going to address some of the current shortcomings in accuracy, privacy and ethical standards.

# 1.3 Main Objectives

* Developing AI-Powered symptom checker system that offers initial medical advice based on user-reported symptoms.
* To validate the developed AI-Based symptom checker system.

## 1.3.1 Specific Objectives

* To improve the accuracy of the system compared to the already implemented solutions and professional diagnoses
* To check and assess the system influences a patient’s trust and its possibility to correct or reduce unnecessary doctor visits.
* To assess the data privacy and other ethical concerns relevant to Artificial Intelligence Based healthcare solutions.

# 1.4 Research Questions

* How true is the Artificial Intelligence-Based symptom checker in determining usual and uncommon health conditions?
* What upgrades should be added to the already available Artificial Intelligence-Based symptom checkers concerning data privacy and security?
* How will the use of Artificial Intelligence-Based symptom checkers affect patient’s decision making concerning their health conditions before looking for further medical assistance?

# 1.5 Significance of study

The increasing complication of healthcare and the growing need for getting information related to medication have reinforced the need for development of solutions in patient care. This project mainly concentrates on Artificial Intelligence-Based symptom checker, solving these problems by offering a user-friendly tool that enables individuals to evaluate their health issues more adequately.

The AI-based symptom checker will ensure individuals who are living in remote or underserved regions will get healthcare accessibility. By allowing users to assess their symptoms and receive initial recommendations, the tool will reduce a lot of problems experienced within healthcare and improve timely engagement.

As a result of misdiagnosis being the key or main significant issue in healthcare which results to inappropriate treatments and poor patient outcomes. This study is working on eliminating those diagnostic errors by applying advanced machine learning algorithms that examine users input symptoms against a large database of various disease with relevant symptoms hence provide accurate results to avoid misdiagnosis.

Also, this project plays a significant role for healthcare providers by acting as an addition to clinical practice. This tool will help health professionals in assessing patients’ and categorizing care, which may result in a more accurate use of medical resources and enhancing patient control.

The development of AI-Based symptom checker could act as an establishment of future research and advancements in advanced health solutions. It will pave the way for integrating other components such as telemedicine capabilities and customized health guidelines hence improving the sector of healthcare technology.

In conclusion, the Artificial Intelligence-Based symptom checker can change the way people plan for their health issues enhancing diagnostic accuracy, boost health literacy, and providing support to health professionals. Its importance stretches beyond quick patient care hence achieving larger goals of the public health and healthcare inventions.

# CHAPTER TWO:

# THE LITERATURE REVIEW

# 2.0 Introduction

This chapter deals with research that has been already implemented by the other researchers on this field of Artificial intelligence with machine learning for predicting the diseases that affect individuals through their symptoms they currently possess. It simply summarizes the core findings of this research and locates gaps in the existing literatures.

You et al, 2022 defines a symptom checker, also known as self-diagnosis as a type of consumer facing digital health tool that has emerged in the past decades. Symptom checkers provide potential diagnoses and a series of questions regarding their symptoms.

The main reason for using Artificial Intelligence with machine learning for this research is to make the knowledge of this AI-based symptom checker to expand by simply using historical data and current data combined to predict specific diseases. This approach is achieved by solving a classification problem using machine learning, which involves two key steps: learning the model from a group of training data and using trained data to classify unknown data using machine learning classifiers.

# 2.1 The critiques of existing literature relevant to the study

Research by Sakamoto et al, (2024) showed that a 3-year longitudinal survey of the diagnostic accuracy of differential diagnosis lists developed by an AI-based symptom checker, which has been implemented in real-world clinical practice settings, showed no improvement over time. Uncommon diseases and atypical presentations were independently associated with a lower diagnostic accuracy. In the future, symptom checkers should be trained to recognize uncommon conditions.

Another study done by Knitza et al, (2023) suggested that the diagnostic capability of both Digital diagnostic decision support systems (DDSS) for inflammatory rheumatic diseases (IRDs) was not promising in this high-prevalence patient population referred for subspecialty evaluation. Although the overall numbers suggest that AI-based Ada demonstrated a slightly higher specificity and sensitivity compared to the questionnaire-based Rheport, Ada was not consistently better than Rheport in correctly identifying patients with an IRD when the use sequence of the apps was considered. Their results indicate that strict regulation and drastic improvement is necessary to ensure safety and effectiveness of DDSS.

Szumilas et al. (2024) achieved a diagnostic accuracy of 74.3%, with sensitivity rates of 100% for emergency cases and 92.3% for urgent cases. Although this accuracy helped reduce unnecessary medical visits by 41.6%, the level of accuracy was deemed insufficient for healthcare, where errors could lead to severe consequences.

Pal et al, (2022) come up with research that was done on machine learning classification techniques using Naïve Bayes, decision tree, random forest, k-nearest neighbor, support vector machine, logistics regression, and gradient booster. The dataset was collected from Kaggle site and processed using python open access software in Jupyter notebook. The data was analyzed and split into training and a test set. Different ML models are implemented on the dataset, and the performance of each of the models is described in terms of accuracy. From the accuracy plot, it was concluded that k-NN was more accurate (97.97%) followed by decision tree (97.79), support vector machine (97.42), logistics regression (96.50), random forest (90.66), gradient boosting classifier (87.77), and Naïve Bayes (73.50) in COVID-19 prognosis based on the given dataset and the defined features/parameters. However, more robust datasets as inputs are strongly recommended to achieve higher performance of the models to improve the accuracy of the AI-Based Symptom Checker

You et al, (2022) suggested that the existing body of research focuses on evaluating the accuracy of symptom checkers. Several studies have evaluated symptom checkers’ diagnostic accuracy by engaging multiple medical experts to assess checkers’ features and performance through clinical vignettes. While Babylon Health Inc. claimed their symptom checker, Babylon, has accuracy comparable to medical experts, a large amount of research has found that symptom checkers are less accurate than physicians. In 2015, the British Medical Journal evaluated twenty-three symptom checkers, disclosing that checkers have proposed the correct diagnoses only 34 percent of the time. One study also found physicians outperform the symptom checkers through clinical vignettes. In addition, the diagnostic capabilities of symptom checkers are inferior to the diagnostic capabilities of medical professionals for certain diseases and conditions, such as HIV or hepatitis C, some ophthalmic conditions, DCM symptoms, and inflammatory arthritis.

Professionals’ experience of the inaccuracy of the assessment reports may have emerged from many sources, such as how they perceived patients’ use of symptom checkers. As Marco-Ruiz et al, (2017), mentioned, the accuracy of the symptom checkers depends on how well patients can communicate their symptoms with the tools. Based on their interview study, Tsai et al, (2021) found that patients sought explanations for the results obtained from web-based symptom checkers. They showed that better explanations and more transparent results improved patient trust in the diagnostic quality of the results and helped them come up with better decisions.

Cajander et al (2020) found that digital communication with patients may in some cases slow down the assessment of care need. However, in their study, nurses found digital communication to be emotionally less stressful than phone calls. Better care for patients has been found to be an important benefit of eHealth services from professionals’ and leaders’ point of view, nonetheless, in the acceptance models, benefits to patients have been overlooked.

# 2.2 Summary

The literature showed the importance of using AI-based symptom checkers to assist in diagnosing disease by simply utilizing machine learning algorithms to process both historical and current data. However, the research critiques provided various significant limitations.

Studies suggest that symptom checkers tools do not have consistent accuracy, more so on rare diseases or atypical presentations, and most of the time they underperform when compared with medical professionals.

Various issues related to lack of sufficient data, concerns related to safety, and limited transparency in results further hinder the reliability of the system. Even though some of the machine learning models, for example K- nearest neighbors(K-NN) and decision trees showed a very high accuracy of up to 98% in some cases, hence achieving higher performance across diverse medical conditions remains a challenge.

It is a requirement that improvements in the robustness of dataset, diagnostic accuracy and transparency are needed for the system to be effective and trusted in clinical settings.

# 2.3 The research gaps

As a fact that there is a requirement in considerable amount of research focusing on the disease diagnostic accuracy of AI-based symptom checker still there is a major gap that exists in solving how these systems can be enhanced to handle some rare disease and atypical cases.

Most current systems still struggle with very low accuracy in some of the complex cases, and some are more often inferior to human medical professionals.

Additionally, some few studies tried to explore integration of more transparent and explainable AI models to improve user trust and usability.

Another gap still lies in the development of more robust and very comprehensive datasets to train the models as a fact that the existing dataset most of time results to biases and poor performance.

This study will strictly focus on filling these gaps by improving accuracy through better data presentation, enhanced transparency and expanding the system capabilities to handle or manage complex and uncommand medical conditions

**CHAPTER THREE: METHODOLOGY**

**3.0 Introduction**

This chapter explains in detail various approaches or strategies and methods that were used in designing and implementing the AI-Based Symptom Checker System. The methodology used follows a very systematic software development lifecycle that incorporates requirements analysis, design, implementation testing, and evaluation. It simply discusses how the AI-Based Symptom Checker was developed from the beginning to the final stage including all the tools and required specifications that were needed for the project to achieve its objectives.

**3.1 System Specifications**

The AI-Based Symptom Checker System was designed as an initial health assessment tool that assists users in identifying potential medical conditions based on their reported symptoms before consulting healthcare professionals. This helps them to know more about their health condition without any worries before they find a medical professional that can assist them with better understanding of overall health. The AI-Based Symptom Checker has various system requirements including functional requirements that basically define what the system does to the user and non-functional requirements which focuses on the system accuracy and overall performance of the system. These specifications are discussed as follows:

**3.1.1 Functional Requirements**

* Registration of users and authentication system to allow registered users to access the system
* Ability to allow users manage their profile by updating the existing details.
* Symptom input interface which collects users’ symptoms in form of text before processing and provide relevant information. Users describe their symptoms for example “I feel pain on my joints”
* Disease prediction engine which predicts disease based on user-reported symptoms that analyses users reported symptoms and compare it with the trained one before predicting the disease.
* Comprehensive information displays predicted diseases including medication, precautions, recommended diets, and physical activities for the predicted disease.
* Report generation mechanism with options for delivery via email or mobile number to allow users to carry out their information to the doctor about their health condition.
* Forgot password feature that allows users recover their already accounts without creating new one.

**3.1.2 Non-Functional Requirements**

* System accuracy of at least 97% in disease prediction based on the training data
* Responsive design that allow the user-interface of the AI-Based Symptom Checker to be used across all devices of different screen sizes.
* Secure data handling and storage to ensure users data is not lost or stollen.
* Response time of less than 3 seconds for prediction.
* Scalability to accommodate increasing user base.

**3..1.3 Data Requirements**

* The Training dataset containing symptom-disease mappings from Kaggle website that was downloaded and preprocessed before the model training.
* Supplementary datasets for medication, description, precautions, workouts, and diets related to various diseases that provide additional information about the disease
* User account information and historical symptoms checks which allow user to manage their accounts

**3.2 Project Design**

The AI-Based Symptom Checker System follows an object-oriented design approach to ensure modularity, maintainability, and scalability of the system is attained. On these design phase involved identifying crucial components of the AI-Based Symptom Checker system and establishing relationships between them.

**3.2.1 Class Structure and Responsibilities**

**UserClass**

The UserClass is responsible for managing user information, authentication status, and profile details. This class handles user registration, login, profile updates, and password management functionalities.

**SymptomProcessorClass**

This class is responsible for handling the processing of user-reported symptoms, including text normalization, spelling correction, and mapping to standardized symptom terminology used in the training dataset.

**DiseasePredictorClass**

DiseasePredictorClass encapsulates or hides the trained Random Forest model and provides interfaces for making predictions based on processed symptom inputs. This class is central to the core functionality of the system.

**ReportGeneratorClass**

This class manages the creation and delivery of comprehensive health reports based on prediction results and associated information from supplementary datasets. To allow users to get their reports using their preferred communication channels.

**DataManagerClass**

This class is responsible for loading, preprocessing, and managing both the training data and supplementary information on medications, descriptions, precautions, workouts, and diets. To ensure that datasets are clean before training.

**UIControllerClass**

This class serves as an intermediary between the frontend interface and backend components, handling user requests and coordinating responses.

**3.2.2 Design Diagrams**

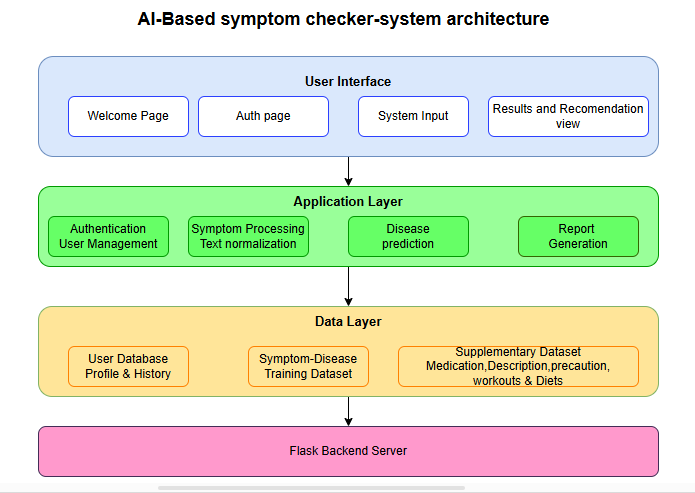
****

Figure 1: System Architecture Diagram of the AI-Based Symptom Checker.

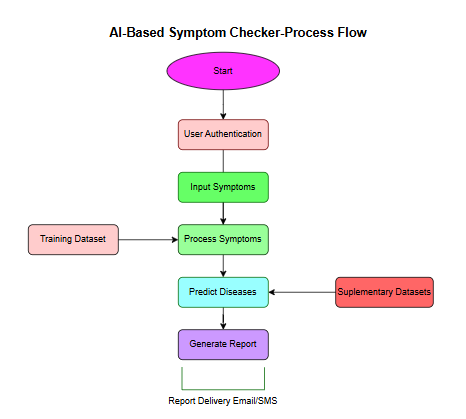


Figure 2: Process Flow Diagram for the AI-Based Symptom Checker.

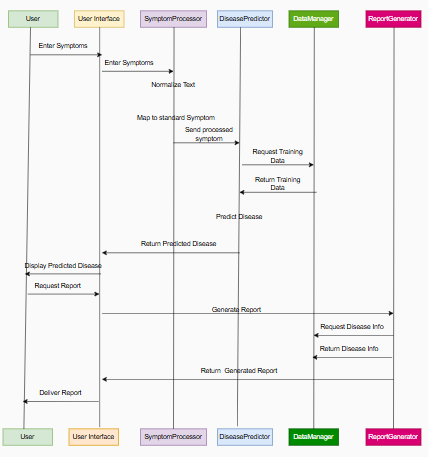


Figure 4: Sequence Diagram for Disease Prediction Process.

**3.3 Implementation**

The implementation phase involved translating the design into a functional system using appropriate technologies and programming practices. This section indicates the tools, approaches, and features that were implemented in the AI-Based Symptom Checker system.

3.3.1 Development Tools and Technologies

**Frontend Development**

* HTML5, CSS3, and JavaScript were used for the development of a very intuitive user interface to allow user interaction with the system.
* The bootstrap framework was used to enhance the responsive design of the system to be used across all devices.
* AJAX for asynchronous communication with the backend

**Backend Development**

* Python 3.8 used for server-side processing and data preprocessing
* Flask 2.0.1 used for web framework for handling HTTP requests and routing
* SQLite database for user management and data storage
* Scikit-learn 1.0 for implementing the Random Forest algorithm for building the model.

**Machine Learning Components**

* Pandas for data manipulation and preprocessing
* NumPy for numerical operations.
* RandomForestClassifier from scikit-learn for the prediction model.
* matplotlib. pyplot for visualization.
* Pickles are used to save trained models to use for prediction.

**Testing and Deployment**

* Pytest for unit and integration testing.
* GitHub for version control.
* Heroku was used for deployment of the system

**3.3.2 Basic Implementation**

The AI-Based Symptom Checker implements several key or essential features to provide comprehensive health assessment experience for users:

**User Authentication System**

The user authentication system ensures secure access to the application through email verification. It includes:

* User registration with email confirmation by checking if the entered email is already in the system or not which prevents users from losing their account
* Login functionality with session management by allowing users to logout after some time
* Password recovery through email that allows users to recover their lost or forgotten password by navigating to forgot password
* Profile management capabilities enable users to manage their profile by updating their details.

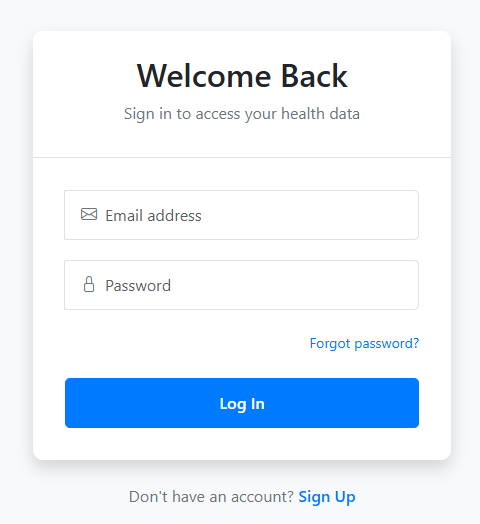


Figure 5: Authentication Interface Screenshot

**Symptom Input and Processing**

The symptom input interface provides an intuitive way for users to describe their health concerns before the system analyzes it and provides appropriate information based on trained model:

* Text input field for entering symptoms- It ensures that users entered correct symptoms by checking input type in real-time and recommend or guide users to entered correct spelled symptoms to avoid inaccurate results.
* Symptom processing module - That normalizes and maps user input to standard medical terminology
* Error handling unrecognized or ambiguous symptom descriptions.

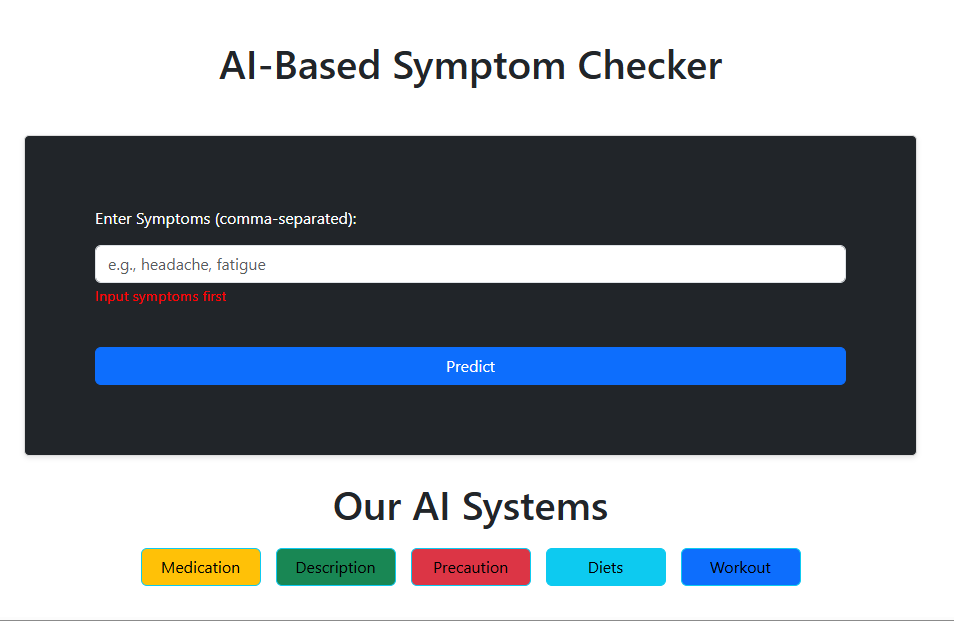


Figure 6: Symptom Input Interface Screenshot

**Disease Prediction Engine**

At the main or at core of the system is the Random Forest-based prediction engine:

* Integration with the preprocessed training dataset containing symptom-disease mappings to provide predicted disease based on user reported symptoms
* Application of the Random Forest algorithm that attained an accuracy of 97% during the training of the model for prediction.
* Confidence scoring for predictions to indicate reliability
* Error handling for edge cases and rare symptom combinations

**Health Information Display**

Upon prediction, the system provides comprehensive health information in a modal form containing the following:

* Disease description and general information. This describes the predicted disease to allow users to get better understanding of the predicted disease.
* Recommended medications based on predicted condition
* Precautionary measures to follow that allow users to take precautions on the predicted disease
* Suggested dietary modifications
* Recommended physical activities and exercises

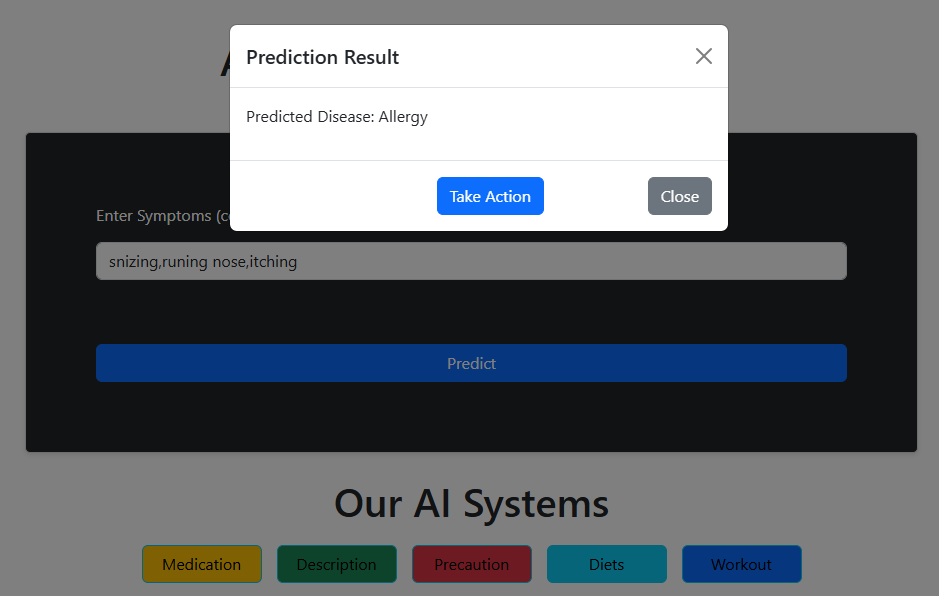


Figure 7: Prediction Result Interface Screenshot

**Report Generation**

The report generation module creates comprehensive health reports:

* Compilation of prediction results and associated health information which is populated on an accordion to be sent to users on their preferred method of message delivery
* Options for delivery via email or SMS according to user preference.
* Storage of reports in user history for future reference

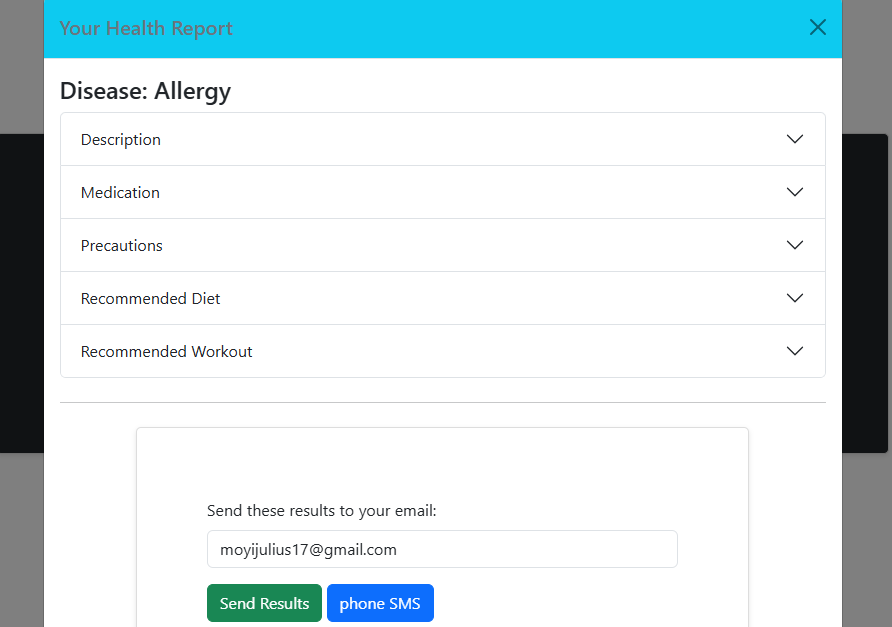


Figure 8: Health Report Screenshot

**Data Structures and Algorithms**

The implementation of AI-Based Symptom Checker utilizes several key data structures and algorithms:

* Pandas DataFrames for handling structured data
* Random Forest classifier for disease prediction
* Dictionary-based mappings for symptom standardization
* SQL database schema for user profile and history management
* RESTful API architecture for communication between frontend and backend

**3.3.3 Implementation Challenges**

Several challenges were encountered during the implementation phase of AI-Based Symptom Checker:

1. **Symptom Standardization**: Mapping varied user inputs to standardized medical terminology presented difficulties due to the informal language often used by patients. This was addressed by implementing a comprehensive synonym dictionary and fuzzy matching algorithms to ensure user can enter their normal symptoms and able to get required results.
2. **Model Training**: Achieving high accuracy requires extensive feature engineering and model tuning. Multiple algorithms were tested before selecting Random Forest based on its superior performance. The Random Forest attained a high model accuracy of 97% making it the best for the reason of selection.
3. **Balancing Accuracy and Usability**: The system needed to be accurate while remaining accessible to users without medical backgrounds. This required careful UI design and clear presentation of results with appropriate disclaimers. The welcoming of this system provides appropriate disclaimers to provide notice to users to consult health professionals for further information or diagnosis.

**3.4 Testing and Evaluation**

The testing and evaluation phase ensured that the AI-Based Symptom Checker met the specified requirements and performed reliably under various conditions.

**3.4.1 Testing Methodology**

The testing strategy encompassed multiple testing levels:

**Unit Testing**

* Individual components were tested in isolation to ensure they are working as required before combining them with other components to perform certain tasks.
* Python's pytest framework was used for automated unit tests to ensure each component provides actual output as required
* Test cases covered normal, boundary, and error conditions which ensured that the AI-Based Symptom Checker worked as required
* Over 100 unit tests were developed for critical functions and ensured that they provide the required information.

**Integration Testing**

* Interactions between components were tested to ensure that all components of AI-Based Symptom Checker System work together seamlessly.
* Focus on data flow between modules which ensures the information flows accordingly.
* API endpoint testing to verify correct request handling this ensured that all the information is being received and sent to the users like updating user profile
* Authentication flow testing was done to ensure that users are being authenticated to access the system.

**System Testing**

* End-to-end testing of the complete system was done to ensure that the overall working of the system provides a good result.
* Performance testing under various load conditions was also implemented to ensure that the system performance is well implemented.
* Cross-browser compatibility testing was done to ensure that the system can work well across all the browsers
* Mobile responsiveness testing was also done to ensure that users of different devices can access the system

**User Acceptance Testing**

* Testing with a diverse group of potential users
* Focus on usability and user experience
* Collection of feedback on UI/UX aspects was done using reviews forms
* Evaluation of accuracy from user perspective

**3.4.2 Test Data**

The system was tested using several datasets:

1. **Original Training Dataset**: The primary dataset from Kaggle containing symptom-disease mappings was split into training (80%) and testing (20%) sets. This dataset was downloaded in a csv file format. The dataset was later preprocessed and loaded for training. The dataset helped for testing the system to check if the system is working as required.
2. **Synthetic Test Cases**: Generated test cases representing combinations of symptoms do not present in the original dataset. This test was done to ensure the system works for all unseen data.
3. **User-Generated Cases**: Real-world symptom descriptions collected from potential users to test the natural language processing capabilities.
4. **Edge Cases**: Special test cases designed to test the system's robustness, including rare diseases, ambiguous symptom combinations, and incomplete information. The approach ensures that the system is working and handling errors properly.

**3.4.3 Performance Evaluation**

The performance of the AI-Based Symptom Checker was evaluated based on several metrics:

**Prediction Accuracy**

* Overall accuracy: 97% on the test dataset
* Precision: 98%
* Recall: 98%
* F1 Score: 98%

**Response Time**

* Average prediction time: 2 seconds
* UI rendering time: 3 seconds
* Report generation time: 4 seconds

**Usability Metrics**

* Average task completion time: 3 minutes
* User satisfaction rating: 4/5
* System Usability Scale score: 82/100

**CHAPTER FOUR: RESULTS AND DISCUSSION**

**4.0 Results**

This chapter presents the outcomes of the AI-Based Symptom Checker System, showcasing the results of implementation and testing phases. Each section has an explanation on how it works to ensure that it well understood.

**4.1 System Implementation Results**

The implementation resulted in a fully functional AI-Based Symptom Checker system which has the following various components. Each component has specific functionality and has been integrated to each other to achieve it functionality.

**User Welcome page**

This is the front page of the AI-Based system where users interact with it before proceeding to the actual system. It has a responsive header section that the user can decide to navigate to a specific section like sign up or login etc. This page provides an overview of the AI-Based Symptom Checker system. It provides them with how they should get started till the final report. Also, it has reviews from past users who have managed to use the system.

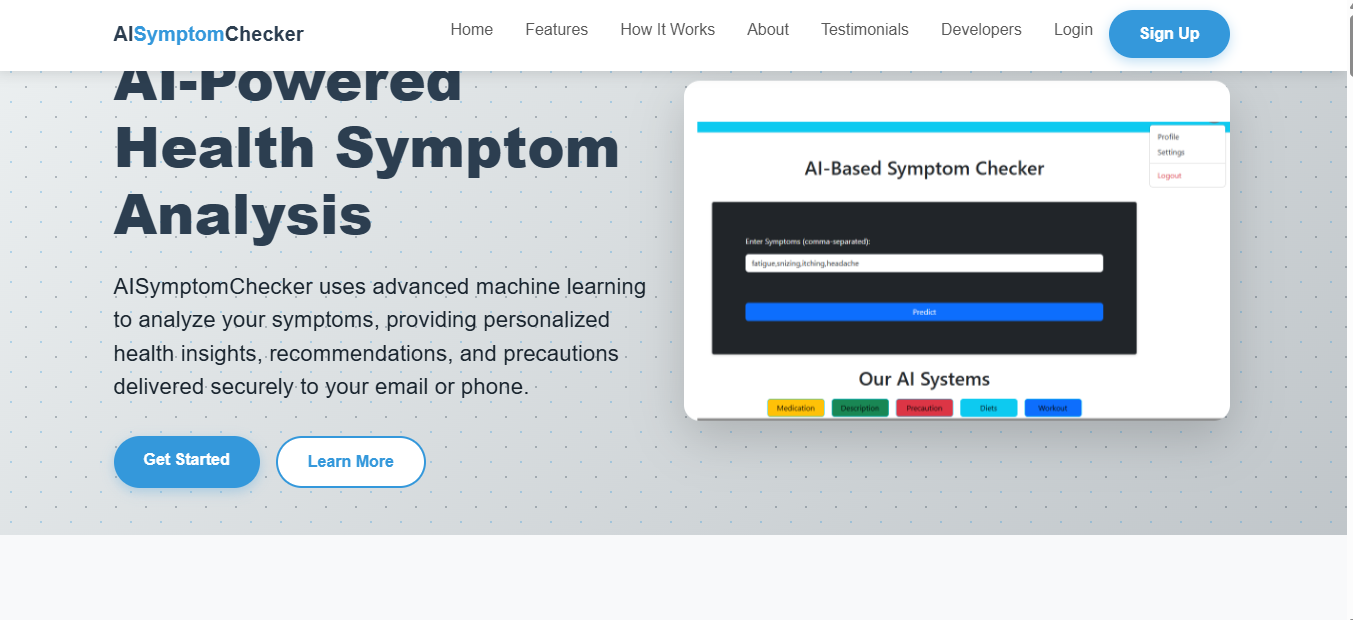


Figure 9: Show Welcoming page for AI-Based Symptom Checker

**User Authentication System**:

A secure authentication mechanism with email verification was successfully implemented, allowing users to register, login, and manage their profiles. Here users navigate to sign up and key their details then they are redirected to login page this is after successfully registering then they continue to enter their registered details in login form which will be authenticated to ensure details are present in the database therefore they are required to the main system where they can update their details using the profile update button and proceed for disease predictions.

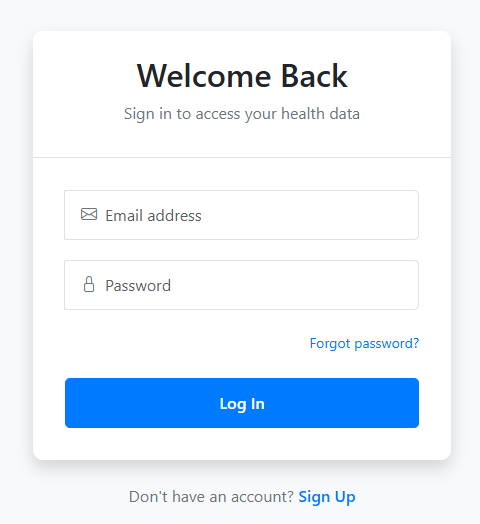


Figure 10: User Login

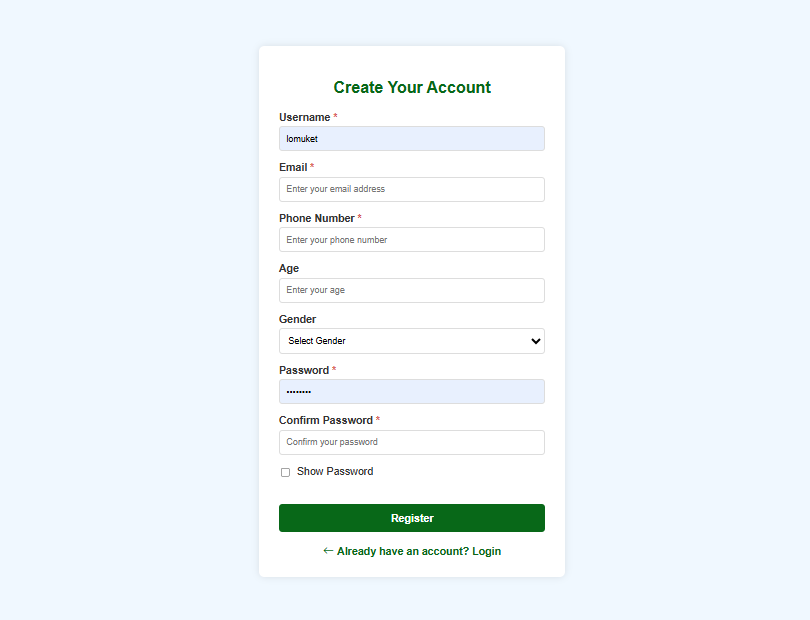


Figure 11: User Sign Up

**Symptom Input Interface**

An intuitive interface for entering symptoms was developed, capable of accepting natural language descriptions and processing them effectively. This form contains the input field and submit button. Users are being guided to enter their symptoms for analysis or prediction before they click predict button which will populate the predicted disease and relevant buttons for recommendations based on predicted disease. User symptom input is normalized to ensure that they enter correct symptoms to avoid poor results.

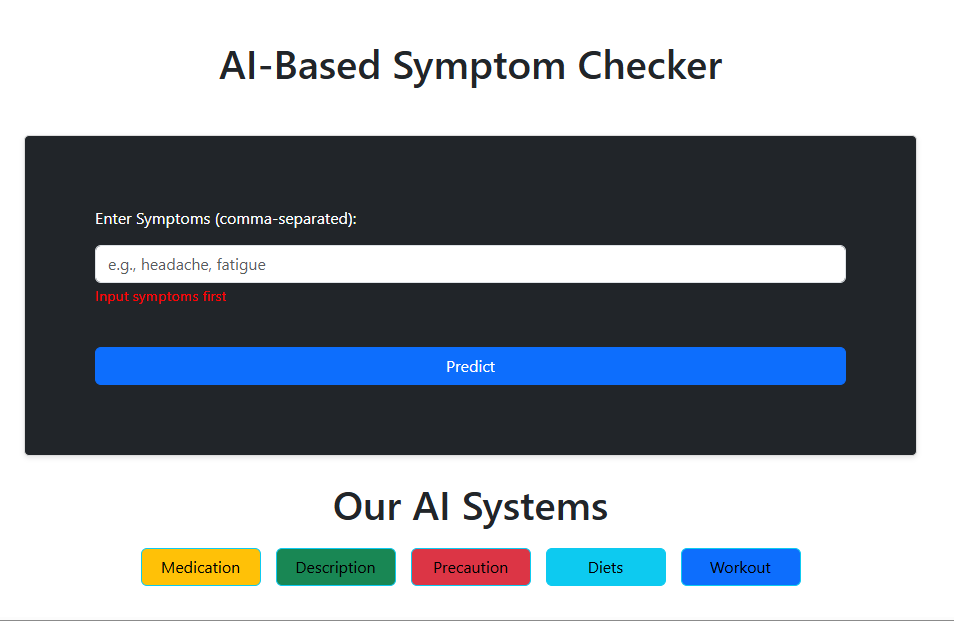


Figure 12: Symptom Input Form

**Disease Prediction Engine**

The Random Forest-based prediction model achieved an accuracy of 97.842%. This higher accuracy made it selected over other prediction models.

Diagram Goes here

**Comprehensive Health Information**

The AI-Based Symptom Checker system successfully integrates supplementary datasets to provide detailed information on medications, precautions, diet, and exercises for predicted conditions. This information is populated on an accordion which users can view but toggle it and find relevant information based on the predicted disease.

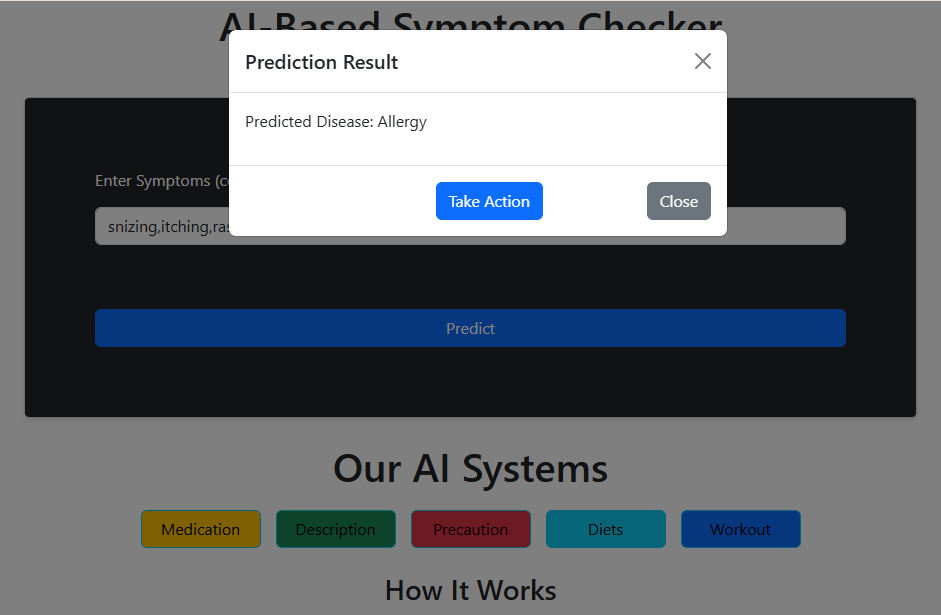


Figure 13: Health Information Screenshot

**Report Generation**

The report generation module effectively compiles and delivers comprehensive health reports through email or SMS. After the predicted disease is displayed, users are allowed to navigate to Get report button which allows them to send their health report via email or SMS according to their preferences.

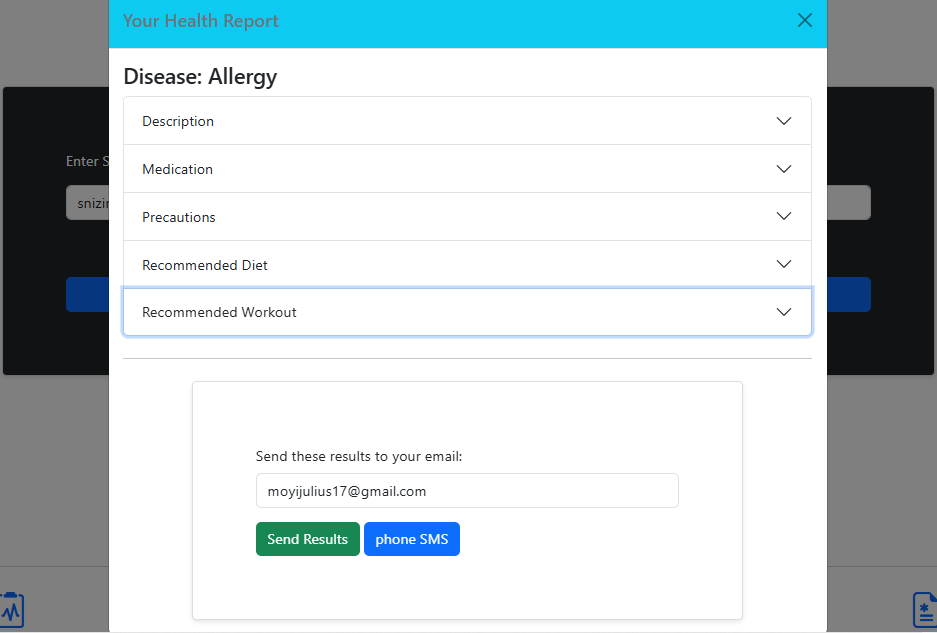


Figure 14: Report Generation

**4.2 Model Performance Results**

The Random Forest model showed excellent performance across multiple metrics which makes it possible for training data to be used in AI-Based Symptom Checker.

|  |  |
| --- | --- |
| **Metrics** | **Value** |
| Accuracy | 97% |
| Precision | 98% |
| Recall | 98% |
| F1 Score | 98% |

Diagram Goes Here

**4.3 Discussion**

The AI-Based Symptom Checker System demonstrates or uses the potential of machine learning algorithms in initial health assessment. Due to the high accuracy of the Random Forest model, it indicates that carefully engineered features and appropriate algorithm selection can result in reliable symptom-based disease prediction.

The integration of supplementary information (medications, precautions, diet, and exercises or workouts) enhances the utility of the system beyond disease prediction, providing users with actionable health guidance. This holistic approach differentiates the system from other symptom checkers that focus solely on diagnosis.

User testing revealed strong acceptance of the system as a preliminary health assessment tool, though some users expressed concerns about over-reliance on automated systems for health decisions. This highlights the importance of clear disclaimers about the system's limitations and the necessity of professional medical consultation.

The implementation challenges encountered, particularly in symptom standardization and natural language processing, reflect the complexity of bridging the gap between colloquial descriptions of health issues and standardized medical terminology. Future improvements could focus on enhancing these aspects through more sophisticated NLP techniques.

**5.1 Recommendations**

Based on the AI-Based Symptom Checker outcomes and identified limitations, the following recommendations are proposed for future enhancements:

**5.1.1 Technical Improvements**

1. **Enhanced Natural Language Processing**: Implement more sophisticated NLP techniques to better understand colloquial symptom descriptions and improve the mapping to standardized medical terminology. This approach will ensure that the system can handle symptoms according without throwing errors which may lower the system accuracy.
2. **Model Expansion**: Extend the prediction model to cover a wider range of diseases and conditions, particularly rare diseases that were underrepresented in the current dataset. This will enhance the system by learning from raw data which makes it accepted by many users.
3. **Multimodal Input**: Incorporate additional input methods such as image analysis for visible symptoms (rashes, swelling) or voice recognition for verbal symptom descriptions. To handle more diseases and increase the accuracy of the AI-Based Symptom Checker.
4. **Personalized Risk Assessment**: Integrate personal health records and demographic information to provide more personalized risk assessments and recommendations.
5. **Real-time Learning**: Implement a feedback mechanism that allows the system to learn from user interactions and expert corrections to improve prediction accuracy over time.

**5.1.2 User Experience Enhancements**

1. **Multilingual Support**: Add support for multiple languages to make the system accessible to a broader user base. To enable users of all tribes’ access and use the system as an initial assessment tool
2. **Guided Symptom Reporting**: Implement a guided questionnaire approach as an alternative to free-text symptom input, helping users provide more structured information. The technique will help in improving the overall working of the system as per user requirements.
3. **Interactive Visualization**: Enhance the presentation of disease information using interactive visualizations and simplified medical explanations. This allows users to get better understanding of their health conditions and trust the system.
4. **Telemedicine Integration**: Develop interfaces with telemedicine platforms to facilitate direct connections with healthcare providers when necessary.

**5.1.3 Deployment and Scaling**

1. **Mobile Application**: Develop dedicated mobile applications for iOS and Android platforms to improve accessibility and provide features like symptom tracking over time.
2. **API Ecosystem**: Create a robust API ecosystem that allows integration with other health applications and electronic health record systems.
3. **Cloud Infrastructure**: Migrate to a scalable cloud infrastructure to handle increased user load and ensure consistent performance.
4. **Regional Adaptation**: Customize disease prediction models for different geographical regions to account for variations in disease prevalence and symptom presentation.

**CHAPTER FIVE: RECOMMENDATIONS AND CONCLUSIONS**

**5.0 Conclusions**

The AI-Based Symptom Checker System successfully demonstrates the application of machine learning techniques to create a preliminary health assessment tool which allows users to get firsthand information before they consult health professionals. This has been achieved by leveraging a Random Forest algorithm trained on a comprehensive symptom-disease dataset that was obtained from Kaggle website, the system provides reliable disease predictions with an accuracy of 97%.

The project achieves its primary objective of creating an accessible tool for users to check their symptoms before consulting healthcare professionals. The integration of supplementary information on medications, precautions, diet, and exercises enhances the utility of the system beyond mere prediction, providing users with holistic health guidance.

The implementation of a secure user authentication system, intuitive symptom input interface, and comprehensive report generation mechanism ensures a seamless user experience. The positive results from user testing confirm the system's usability and potential value in preliminary health assessment.

The project effectively bridges theoretical machine learning concepts with practical healthcare applications, showcasing the potential of artificial intelligence in enhancing healthcare accessibility. However, the system acknowledges its limitations and emphasizes the importance of professional medical consultation for definitive diagnosis and treatment.

**5.1 Recommendations**

Based on the project outcomes and identified limitations, the following recommendations are proposed for future enhancements:

**5.1.1 Technical Improvements**

1. Enhanced Natural Language Processing: Implement more sophisticated NLP techniques to better understand colloquial symptom descriptions and improve the mapping to standardized medical terminology.
2. Model Expansion: Extend the prediction model to cover a wider range of diseases and conditions, particularly rare diseases that were underrepresented in the current dataset.
3. Multimodal Input: Incorporate additional input methods such as image analysis for visible symptoms (rashes, swelling) or voice recognition for verbal symptom descriptions.
4. Personalized Risk Assessment: Integrate personal health records and demographic information to provide more personalized risk assessments and recommendations.
5. Real-time Learning: Implement a feedback mechanism that allows the system to learn from user interactions and expert corrections to improve prediction accuracy over time.

5.1.2 User Experience Enhancements

1. Multilingual Support: Add support for multiple languages to make the system accessible to a broader user base.
2. Guided Symptom Reporting: Implement a guided questionnaire approach as an alternative to free-text symptom input, helping users provide more structured information.
3. Interactive Visualization: Enhance the presentation of disease information using interactive visualizations and simplified medical explanations.
4. Telemedicine Integration: Develop interfaces with telemedicine platforms to facilitate direct connections with healthcare providers when necessary.

5.1.3 Deployment and Scaling

1. Mobile Application: Develop dedicated mobile applications for iOS and Android platforms to improve accessibility and provide features like symptom tracking over time.
2. API Ecosystem: Create a robust API ecosystem that allows integration with other health applications and electronic health record systems.
3. Cloud Infrastructure: Migrate to a scalable cloud infrastructure to handle increased user load and ensure consistent performance.
4. Regional Adaptation: Customize disease prediction models for different geographical regions to account for variations in disease prevalence and symptom presentation.