UNIVERSITY OF MINES AND TECHNOLOGY TARKWA

FACULTY OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

A PROJECT REPORT ENTITLED

PHARMIT: A MEDICAL DIAGNOSTIC AND PHARMACY LOCATOR SYSTEM

BY

ISAAC KWAME KONDO

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE IN

COMPUTER SCIENCE AND ENGINEERING

PROJECT SUPERVISOR

……………………………….

DR. VINCENT M. NOFONG

TARKWA, GHANA

JUNE 2020

[Document title]

# DECLARATION

I declare that this project work is my work. It is being submitted for the degree of

Bachelor of Science in Computer Science and Engineering at the University of Mines and Technology (UMaT), Tarkwa. It has not been submitted for any degree or examination in any other University.

………………………………………….

(Signature of Candidate)

………….... day of June, 2018.

# ABSTRACT

Medical diagnosis is the stepwise procedure taken to identify the disorder that affects a patient. In addition to laboratory test results, knowing various signs and symptoms is a prerequisite to successfully categorize the health condition of a patient, hence allow for medical decision and prognosis. An erroneous diagnosis causes health complications of a patient since the medication will be based on the results of the diagnosis. The pain with which diseases such as malaria and cholera inflict encourages patients to self-diagnose, especially when the symptoms are repetitive. In Ghana, in-person consultation with a doctor by a patient is the primary way to get diagnosed, and this increases consulting and waiting time since many patients in the hospital causes a bottleneck. Fatigue arises from the medical practitioners as they attend to many patients, and this is a significant contribution to erroneous diagnosis. Locating licensed pharmacies enhances easy and swift routing to buy drugs to alleviate pain. Hence, this project seeks to apply the Naïve Bayes machine learning principle to provide an accurate diagnosis for patients who choose to self-medicate and also provide the locations of pharmacies in their locality.

In memory of my sister

Juliet Kondo

# ACKNOWLEDGEMENTS

Profound thanks go to God Almighty for His grace, love, and mercies upon my life.

I am thankful to my supervisor Dr. Vincent M. Nofong, for his ample time to go through my work to make the necessary corrections.

# TABLE OF CONTENTS

**Content Page**

[DECLARATION i](#_Toc43321082)

[ABSTRACT ii](#_Toc43321083)

[ACKNOWLEDGEMENTS iv](#_Toc43321084)

[TABLE OF CONTENTS v](#_Toc43321085)

[LIST OF FIGURES viii](#_Toc43321086)

[CHAPTER 1 INTRODUCTION 1](#_Toc43321087)

[1.1 Problem Definition 1](#_Toc43321089)

[1.2 Project Objectives 2](#_Toc43321090)

[1.3 Scope of Project 2](#_Toc43321091)

[1.4 Expected Outcome 2](#_Toc43321092)

[1.5 Methods Used 2](#_Toc43321093)

[1.6 Facilities Used 2](#_Toc43321094)

[1.7 Project Organization 3](#_Toc43321095)

[CHAPTER 2 LITERATURE REVIEW 4](#_Toc43321096)

[2.1 Introduction 4](#_Toc43321098)

[2.2 Machine Learning 4](#_Toc43321099)

[2.2.1 Types of Machine Learning Techniques 5](#_Toc43321100)

[2.2.2 Machine Learning Algorithms 6](#_Toc43321101)

[2.3 Medical Diagnostic Systems 9](#_Toc43321102)

[2.3.1 Analysis of the Tradition Diagnostic System 9](#_Toc43321103)

[2.3.2 Software-Based Diagnostic System Analysis 10](#_Toc43321104)

[2.4 Adopted Diagnostic System in Ghana 11](#_Toc43321105)

[2.5 Proposed System 11](#_Toc43321106)

[2.5.1 Justification of the Proposed System against the Currently Adopted Systems for Diagnosing in Ghana 12](#_Toc43321107)

[2.5.2 Similarities Between the Knowledge-Based Diagnostic Systems and the Proposed System 12](#_Toc43321108)

[2.5.3 Differences Between the Knowledge-Based Diagnostic Systems and the Proposed System 12](#_Toc43321109)

[2.6 Conclusion 13](#_Toc43321110)

[CHAPTER 3 SYSTEM DESIGN AND METHODOLOGY 14](#_Toc43321111)

[3.1 Introduction 14](#_Toc43321113)

[3.2 Feasibility Studies 14](#_Toc43321114)

[3.3 Requirement Gathering and Analysis 15](#_Toc43321115)

[3.3.1 Functional and Non-Functional Requirements 15](#_Toc43321116)

[3.4 Implementation and Coding 18](#_Toc43321117)

[3.4.1 Flow Chart 19](#_Toc43321118)

[3.4.2 Symptom Input 20](#_Toc43321119)

[3.4.3 Disease Prediction Model 20](#_Toc43321120)

[3.4.4 En-Route Maps 23](#_Toc43321121)

[3.4.5 API Hosting 24](#_Toc43321122)

[CHAPTER 4 SYSTEM TESTING AND DISCUSSION 25](#_Toc43321123)

[4.1 Introduction 25](#_Toc43321125)

[4.2 Specified Requirements 25](#_Toc43321126)

[4.3 System Testing 25](#_Toc43321127)

[4.3.1 Snapshot of Views 26](#_Toc43321128)

[4.3.2 Testing the Map Feature 28](#_Toc43321129)

[4.4 Confusion Matrix 29](#_Toc43321130)

[4.5 Drawbacks and Limitations 30](#_Toc43321131)

[CHAPTER 5 CONCLUSION AND RECOMMENDATION 31](#_Toc43321132)

[5.1 Conclusion 31](#_Toc43321134)

[5.2 Recommendation 31](#_Toc43321135)

[REFERENCES 32](#_Toc43321136)

# 

# LIST OF FIGURES

**Figure Title Page**

[Figure 3.1. Use Case Diagram for the Diagnostic and Locator System](#_23ckvvd) 16

[Figure 3.2 Activity Diagram for the Diagnostic and Locator System 1](#_ihv636)7

[Figure 3.3 Flowchart for the Diagnostic and Locator System 1](#_32hioqz)9

[Figure 3.4 Training Dataset Snapshot](#_1hmsyys) 20

[Figure 3.5 Probabilities Snapshot 2](#_41mghml)3

[Figure 3.6 Geolocation Initialization 2](#_2grqrue)3

Figure 3.7 Hosting the Trained Model 24

[Figure 4.1 Dashboard View 2](#_vx1227)6

[Figure 4.2 Make Diagnosis Pane 2](#_3fwokq0)6

[Figure 4.3 Detailed Symptom Selection View 2](#_1v1yuxt)7

[Figure 4.4 View of Selected Symptoms 2](#_4f1mdlm)7

[Figure 4.5 Outcome of Diagnosis 28](#_2u6wntf)

[Figure 4.6 Snapshot of Maps 29](#_19c6y18)

[Figure 4.7 Confusion Matrix 30](#_19c6y18)

# 

# CHAPTER 1

# INTRODUCTION

## 1.1 Problem Definition

The lack of access to health facilities by patients has caused about 8 220 deaths per year in Ghana, of which 121 cases of malaria and 8 099 cases of all other diseases (Asare, 2017). Appiah (2015) stated that there are only about 1 372 major health facilities in the country, and this is to serve a population of about 25 million. Adequate time for productivity is wasted by patients who have to join long queues in hospitals to be diagnosed by medical practitioners. Moreover, one doctor serves about 9 749 persons in Ghana. The situation of fewer doctors in the country reduces the quality of time-per-patient spent. The quest to consult most patients causes fatigue, and this often leads to an erroneous diagnosis, which is not trivial but very fatal (Graber and Matthew, 2008). Also, the inability of patients to access health facilities and have a consultation with medical practitioners cultivates the urge to self-diagnose to relieve pain. Most people are known for repeating previous prescriptions when they experience symptoms of previous diseases such as malaria, cholera, and typhoid fever (Yinyeh and Alhassan, 2015). Nevertheless, anticipate the casualties involved if such intuition to self-diagnose is wrong! About 8 000 lives are lost each year in the United States of America (USA) due to incorrect diagnoses (Shojania and Dixon, 2016).

Machine learning is a branch of artificial intelligence. In the medical domain, diagnostic classification and treatment are the main jobs for a physician, and a significant focus of machine learning is to recognize intricate patterns and make decisions based on data on its own (Oguntimilehin *et al*., 2013). With all these said, a medical diagnostic software application is developed on models built with machine learning principles, to aid in the accurate and quick diagnosis of patients having the symptoms of malaria, and cholera, to reduce the death rate due to errors in diagnosing which is fatal.

## 1.2 Project Objectives

The project's objective is to develop a software system that will provide an accurate diagnosis after a patient shows symptoms, and also locate a pharmacy for the purchase of prescribed medications to reduce diagnostic errors and drug abuse.

## 1.3 Scope of Project

The software system would be limited to the usage of android devices, a patient who has literacy in computing devices, and also a patient that experiences the symptoms of malaria and cholera. Moreover, due to the location of the patient, licensed pharmacies at Nkwanta were pinned on google maps.

## 1.4 Expected Outcome

By the completion of this project, it is expected that patients can diagnose with ease and with maximum accuracy to avoid further complications. Also, help locate pharmacies around.

## 1.5 Methods Used

The following methods were used to complete this project:

1. Review of relevant literature.
2. Consultation with medical physicians.
3. Machine Learning (Supervised Learning) with Python Programming Language.
4. Consultation with lecturers.
5. Training data set from goggle.

## 1.6 Facilities Used

These facilities were used in the course of the project:

1. The university library.
2. Internet.
3. GitHub.
4. Sci-kit learn.
5. Flutter and Flask.

## 1.7 Project Organization

Chapter one entails the problem statement, project objectives, the scope of the project, expected outcome, methods and facilities used, and the project organization. Chapter two has literature reviews on medical diagnostic systems. Chapter three comprises of the system's analysis and design, and the methodology used. Chapter four states the implementation process.

Chapter five has a conclusion and recommendations on the system.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Introduction

Medical diagnosis is the identification of a disorder that affects a patient based on proven clinical data (Oguntimilehin *et al*., 2013). A list of symptoms alone seldom leads to an accurate diagnosis (Baerheim, 2001). Hence accurate diagnosis involves not only symptoms from patients but also the patient's clinical history and physical examinations where necessary. Medical practitioners lay basis on information gathered from patients to determine the specific disease and further make prognosis and prescriptions for curing them.

Moreover, extensive knowledge base, the ability to comprehend all available information, and having a well-trained consultation and examination skills are the fundamental skills required by clinicians to excel in providing correct diagnosis (Graber and Matthew, 2008). Unfortunately, these skills vary among various clinicians and are cultivated by long term practice, which involves a try and error process- which is fatal.

The advancement of technology has seen most hospitals in Ghana using various software applications to manage the data of patients instead of using folders. Nevertheless, this cannot solve the queuing problems in the hospitals due to fewer health facilities and practitioners working, and this puts much pressure on them. This project is going to see to the development of a software application that would give an accurate diagnosis and help locate nearby pharmacies. This would control the rate of erroneous self-diagnosis and also reduce drug abuse cases.

## 2.2 Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) in computer science, which is the imitation of human intelligence by a computer system. ML is a paradigm that learns from experience for the improvement of future performance. One main objective of this field is to learn methodologies automatically. ML is used in medical diagnosis for disease prediction, data analysis, and therapy planning (Pavithra and Jayanthi, 2018).

ML serves as a platform that incorporates computer systems with the medical field for quality treatment and efficient diagnosis.

Machine Learning can be divided into various categories, and also has some algorithms that enable systems to work intelligently.

### 2.2.1 Types of Machine Learning Techniques

There are many categories of these techniques, and this is due to the difference in:

1. The processing and display of input and output data.
2. The type of problem to be solved.

*Supervised learning*

Under the supervised learning technique, results are predicted based on training from a dataset. Output values are predicted based on a given input data or vector. These output values can either be continuous or discrete. The continuous value is associated with *regression* problems, while the discrete value is for *classification* problems. Classification deals with a qualitative target value that has a correspondence to a class, and its output is categorical. This gives a Yes or No prediction. For example," is this spot, contagious?". While Regression answers "How much" and "How many" questions (Meherwar and Maruf, 2017).

*Parametric model*

The predictive function of this model is a combination of a fixed number of parameters. The learning stage with datasets is the first step, and a predictive step follows it in which new input data is classified. This model does not need training data after learning.

Hence, training data is discarded. An example is the Neural Network.

*Non-parametric model*

With the non-parametric model, the dataset after training is not discarded but maintained for prediction. Support Vector Machines (SVM) and Nearest Neighbour (NN) algorithms are the most commonly used non-parametric models.

*Unsupervised learning*

Classification of data under unsupervised learning is by finding out similarities between the input data, and this is also known as the density estimation. Clustering falls under unsupervised learning since it makes clusters based on similarities.

*Semi-supervised learning*

The semi-supervised learning is a combination of supervised and unsupervised learning techniques, where labelled data is not obtained, i.e., this technique uses both labelled and unlabelled data for training (Pavithra and Jayanthi, 2018). The various categories of these techniques are:

1. Generative model.
2. Graph-based model.
3. Two-step model.
4. The low-density region-based decision boundary model.

*Reinforcement learning*

This technique uses the Markov Decision Process (MDP), which explores various possibilities on a test data and finds a correct output using the evaluative feedback process; this is learning with a critic. The algorithm is not shown how to get the right answers when wrong but is informed when faulty. This is also identified by delayed reward and trial-and-error. A clear difference between reinforcement learning and supervised learning is that, with reinforcement learning, accurate inputs and outputs are not offered, and it also focuses on on-line performance. However, precise inputs and outputs are provided under supervised learning.

### 2.2.2 Machine Learning Algorithms

*Naïve Bayes Classifier*

The Naïve Bayes classifier is categorized under the supervised learning method and is based on the Bayesian Theorem. It assigns a finite set of labels to data previously unknown, and it assumes independence of features mutually, i.e., it considers each element contributes separately. The Naïve Bayes classifier can be applied on large datasets and is for binary and multi-class problems. This classifier is successfully known for classifying text documents, and spam filtering.

*Merits of the Naïve Bayes Classifier*

1. The Naïve Bayes classifier is highly scalable, simple and is implemented efficiently.
2. The classifier is not sensitive to unimportant features.
3. It has a lesser training period due to smaller training dataset.
4. It is easy to implement.

*Demerits of the Naïve Bayes Classifier*

1. Considering continuous variables is difficult.
2. It assumes that all attributes are mutually independent.
3. It tends to assign zero probability to variables not observed in the training data set. (Zero Frequency).
4. There are higher chances of loss of accuracy.

*Support Vector Machine*

The support vector machine (SVM) is a supervised learning model that analysis data before classification. SVM performs classification by constructing an optimal hyperplane of which is used to separate two different classes and can be controlled when the condition is true (Pavithra and Jayanthi, 2018). The hyperplane is a linear line that separates a set of data, and support vectors are vectors that have data points near to the hyperplane.

*Merits of the Support Vector Machine*

1. Support vector machines are accurate classifiers.
2. Due to the usage of subsets of training points, it is more efficient.
3. It is useful in a high dimensional space.
4. SVM is memory efficient.

*Demerits of the Support Vector Machine*

1. The support vector machine algorithm cannot be extended to solve multi-class problems.
2. The algorithm has a long training time on large data sets.
3. There is no probabilistic explanation for the classification of data points.
4. It underperforms when there is much noise in the data set.

*Neural Network*

The Neural Network (NN) is an algorithm that has been developed to mimic the structure and functionalities of the biological neurons. The NN has dendrites for information collection, the soma for information processing, and the axon for the output. The architecture for the artificial neural network was first presented by McCulloch and Pitts (1943), and many architectures have been developed. Each node is initially fed with a large amount of input data, get passed unto the hidden layers, and then an output is generated.

*Merits of Neural Networks*

1. There is no restriction on input variables.
2. The neural network algorithm can model non-linear problems.
3. It has appreciable fault tolerance.
4. It has a higher parallel processing ability.

*Demerits of Neural Networks*

1. For maximum accuracy, a large amount of training data is required.
2. The algorithm is more computationally expensive.
3. The duration of the network is uncertain.
4. It has a high hardware dependency.

*Random Forest Algorithm*

The Random Forest Algorithm is an advanced version of the Decision Tree algorithm (DT), of which trees randomly grow in a selected subspace. The Random Forest Algorithm solves the overfitting problems the DT faces with its training data. Random Forest Algorithm, unlike the DT, has trees in various subspaces, and this helps generalize its classification in complementary ways, and the combined classification can be improved monotonically (Ho, 1998).

*Merits of the Random Forest Algorithm*

1. The implementation of the algorithm is fast and easy.
2. There is no overfitting in handling extensive data.
3. It can solve both classification and regression problems.
4. There are methods for balancing errors in datasets where there are imbalanced classes.

*Demerits of Random Forest Algorithm*

1. More computational resources are required and consumed.
2. It is time-consuming with large decision trees.
3. There is little control over what the model does.
4. It does precisely well with classification problems but not as for regression problems.

## 2.3 Medical Diagnostic Systems

Generally, there exist two central medical diagnostic systems in operation in Ghana. This involves having an in-person consultation with a doctor in the hospital, which can be termed as the traditional system of diagnosing. The other method, which involves the use of computer software applications to aid in diagnosing a disease either by the doctor or the patient and this can also be termed as the Clinical Diagnosis Support Systems (CDSS) or Computer-Aided Diagnosis System (CADs).

### 2.3.1 Analysis of the Tradition Diagnostic System

The traditional diagnostic system depicts a scenario of a patient who visits a hospital, and his or her vitals taken. There is a consultation initiated by a nurse who asks for symptoms and discomforts. The information is documented in a folder or saved in the records of the patient. The last phase of the consultation process is done by the doctor, who further asks for symptoms and matches it with the clinical history of the patient to synthesize coherently all that the patient has said to aid in diagnosing (Graber and Matthew, 2008).

*Merits of the Traditional Diagnostic System*

Firstly, the presence of the patient in the hospital makes it easier for physical examinations and laboratory tests to be taken. Moreover, patients can be treated with urgency when the medical situation gets worse.

*Demerits of the Traditional Diagnostic System*

Firstly, much time is spent in consulting a patient, since a patient can give a diversity of symptoms that makes comprehension very difficult. Secondly, there is a higher tendency of giving erroneous diagnosis due to fatigue by the health practitioner on duty, which arises from the fact that there is a poor doctor to patient ratio in the country. Furthermore, there is a reduction in the quality of health services rendered to the patients in Ghana.

### 2.3.2 Software-Based Diagnostic System Analysis

The quality of medical services can readily and increasingly be improved by using advanced software in diagnosing a patient. The advancement in technology and research of AI has made this possible. The following are the various types of Software-Based Diagnostic Systems:

1. Expert Systems. They are a branch of AI and can be defined as intelligent computer programs that use a knowledge base and inference engine to solve severe problems and provide a diagnosis (Russell and Norvig, 2002). These systems are also of the kind of Knowledge-Based Systems, i.e., they are systems that have a database of various diseases, and uses rules and procedures (an inference engine), to make predictions or diagnosis based on patient's input symptoms (Yinyeh and Alhassan, 2015). Examples of such systems are MYCI, DENDRAL, PXDES, CaDet.
2. Machine Learning Systems. These systems are also a branch of AI that deals mostly with learning (training using datasets) since most researchers agree that a requirement for any intelligent behaviour is learning (Oguntimilehin *et al*., 2013). Examples of Machine Learning systems are PReDicT, AML.
3. Differential Diagnosis Generators. These are also intelligent computer programs that suggest diagnosis based on past clinical data. These systems, unlike the machine learning systems, provides a list of possible diagnosis which serves as a guide to the medical practitioner. An example is the Isabel Clinic Application.

*Merits of Software-Based Diagnostic Systems*

1. Software-Based diagnostic systems are more efficient and accurate in delivering a diagnosis, i.e., less prone to error per precise training data.
2. Advanced systems such as the Expert systems can handle cases as complex as that which would have demanded a solution from an expert practitioner.

*Demerits of Software-Based Diagnostic Systems*

1. Most of the deployed systems have awful user interfaces and uses much medical jargon hence, making the user experience terrible and uncomfortable, yielding a low understanding of the system.
2. An attempt to capture a broader scope of diseases causes the database to be loaded.

## 2.4 Adopted Diagnostic System in Ghana

Ghana, as a developing country and economically poor by Western standards, is now advancing towards the introduction of technology into its medical sector, i.e., the use of software applications for managing data and consultations (Drislane *et al*., 2014). At hand, the country still uses the traditional method of diagnosing patients. A patient has to visit a hospital, which wastes a lot of productive time.

## 2.5 Proposed System

The proposed system would help solve the problem of wrongful diagnosis, and wasting productive time in the hospitals with ailments such as cholera and malaria. The system will give an accurate diagnosis based on symptoms selected, and quickly locate pharmacies nearby as an option.

Features of the Proposed System

1. User-friendly interface with less use of medical jargon.
2. The system would be efficient and accurate in diagnosing patients.
3. The software would be a mobile application.
4. The software would be a crossed-platform application, i.e., the python three programming language would be used alongside the Scikit-learn library.

### 2.5.1 Justification of the Proposed System against the Currently Adopted Systems for Diagnosing in Ghana

The system is going to be more efficient, accurate, and economical. Diagnosing patients would be a matter of classification. Accurate diagnosis by the system would help reduce wrongful diagnosis by those who choose to self-diagnose. There is going to be a reduction in pressure on both medical facilities and doctors in that, patients suffering from diseases such as choleraandmalariacan comfortably interact with the system and get an accurate diagnosis and locate pharmacies quickly. Time saved can be used for a productive task, and hence the cost of having consultations in private and costly health facilities would be reduced.

### 2.5.2 Similarities Between the Knowledge-Based Diagnostic Systems and the Proposed System

The various systems, such as the Knowledge-based systems, the Differential Diagnostic System, and the proposed system, are all software-based applications. They take patient symptoms to find a diagnosis. Hence, it would be usable even in the rural parts of the country.

### 2.5.3 Differences Between the Knowledge-Based Diagnostic Systems and the Proposed System

With the individual differences, the Knowledge-Based system, unlike the proposed system, uses a database and inference engine to diagnose diseases. The issue with this system is that the database becomes huge when many conditions or symptoms are to be captured.

Furthermore, the Differential Diagnostic system functions similarly to the proposed system. It functions by providing a list of possible diagnosis to the medical practitioner to aid in diagnosing. In contrast, the proposed system offers the most likely, and specific disease per the symptoms provided.

## 2.6 Conclusion

Since accessibility to health amenities have been the central issue in the rural areas and some of the urban areas in Ghana, advancement in Machine Learning has proven to be better assistance in the medical field, since the technology is suited for analysing data. Classification and treatment of diseases are the main tasks of physicians, and Machine learning medical diagnostic systems can automatically recognize patterns and make intelligent decisions.

# CHAPTER 3

# SYSTEM DESIGN AND METHODOLOGY

## 3.1 Introduction

Software engineering is a detailed study and a branch of engineering which deals with the use of systematic and disciplined methods from analysing user requirements, designing, developing, testing the software to satisfy the requirements, through to the maintenance of the software. The success criteria in developing a software system are mainly based on whether the development is complete, its usefulness, usability, and being used. A good software system should be reliable, efficient, usable, and re-usable, and deliver required functionality and performance. These goals serve as fundamental requirements for the software and require a software process, a set of related activities that results in a software product, to be followed. In this project, the iterative software development model was used, and this has the following activities:

1. Feasibility Studies.
2. Requirements Gathering and Analysis.
3. Software Design and Implementation.
4. Testing.

## 3.2 Feasibility Studies

A feasibility study analysis considers if a proposed system can be implemented practically and become successful. The objective of this study is to know whether a system will meet organizational requirements, possible integration with existing software, and also an implementation using the current technology, and this should be within an estimated budget and time. Activities in this stage are acted and evaluated with the stated objectives, this produces a feasibility report, and this saves time and resources since infeasible projects and activities will not be started.

The following feasibility studies were considered for this project:

1. Technical feasibility: In this study, assessing various technologies and technical principles required for the successful completion of the project.
2. Resource feasibility: The availability of resources, such as labelled datasets and medical records of patients.
3. Schedule feasibility: The available time was assessed to determine if the project can be completed within the due date for submission with all objectives met.
4. Economic feasibility: The estimated cost in developing the project and the acquisition of data.
5. Operational feasibility: Performance of the system and its intuitiveness.

## 3.3 Requirement Gathering and Analysis

With a positive feasibility report, various description of fundamental features and functionalities of the system was retrieved from the targeted users. The questionnaire requirement elicitation technique was employed to gather various requirements. A google form was shared with targeted users, and most of the proceeded functionalities and features emanated from the majority. The analysis included:

1. Whom is the system designed for? The system is for users who use smartphones with mainly an android operating system and can read and understand the English Language.
2. How is the system going to be used? The system is going to be used by taking and processing a selected list of symptoms from the user and then produces a diagnosis. Pharmacy markers on google maps are added to assist in finding pharmacies in the locality.
3. Where would the system be used? The system is designed for mobile android platforms such as phones and tablets and can be used mainly at "Nkwanta."
4. What type of data would be inputted? A binary list by using various checkboxes.
5. What kind of output is expected from the system? The system produces diagnosis after processing the input data. A dialogue text is expected.
6. Which kind of software development cycle model is to be used? The Iterative model was implemented for the development of the system, and this model depicts several divisions of the project into several iterations. Each iteration goes through the systems requirement, design, implementation, and testing phases until the fully desired system is developed.

The various requirements listed were thoroughly analysed for their validity and incorporated into the system.

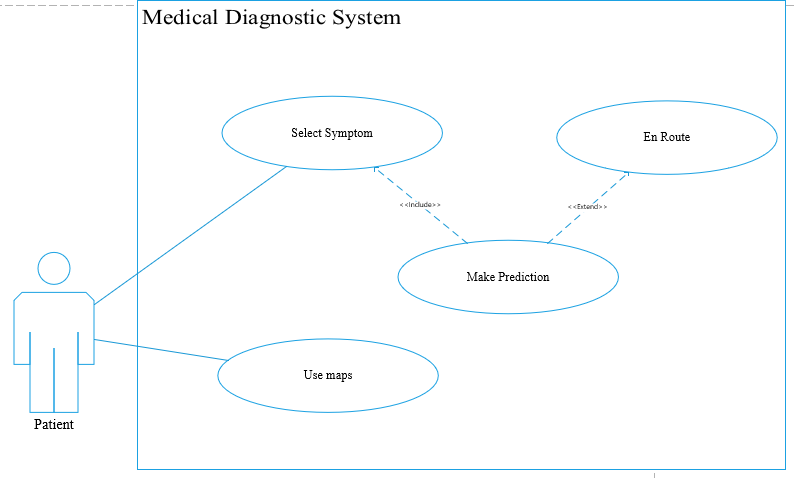
### 3.3.1 Functional and Non-Functional Requirements

The functional requirements are the functions that the system should be able to perform based on some inputs, and the non-functional requirements are the quality expected features about the system which the user has, and this includes the system's reliability, response time, security, performance, and flexibility. The system secures the user's data by resetting the list after the session is ended.

Unified Modelling Language (UML) is a standard language for modelling, and visualizing. The UML was used to model the functional requirements of the system, and in this project, the *use case* and *activity* diagrams were used.

*Use Case Diagram*

A *use case* diagram gives a graphical representation of dynamic behaviour or interactions among components of a system, i.e., the function of an actor in and out of the system. The diagram is made up of actors, a system, use cases, and various relationships between them. The actors are entities that have direct interaction or impact on the system by causing certain functions to trigger. An actor can be a user or another system, and this is represented by the actor stereotype. The system is represented with a rectangle and has a title. The actor is found outside the rectangle while the use cases with their relationships are within. The use cases are oval-shaped and have an extended or include relationship. The former means an alternative task of that use case can be done, and the latter means that use case is needed in order to do a task.

****

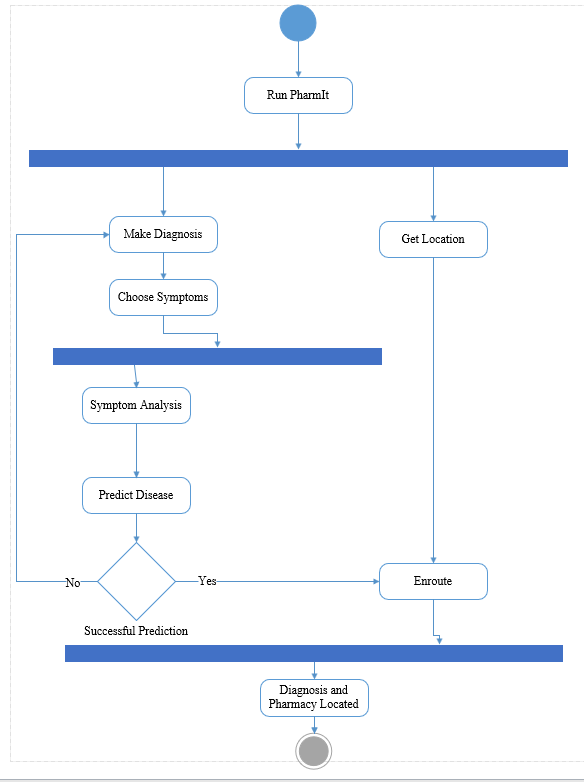
**Figure 3.1 Use Case Diagram for the Medical Diagnostic and Pharmacy Locator System**

Figure 3.1 shows the functionality of the system, and also various goals the user (patient) can achieve.

1. The user can select the desired symptoms and proceed to predict or make a diagnosis.
2. The user can use the map to locate pharmacies around.

*Activity Diagram*

The activity diagram is either the sequential or concurrent flow of activities that also models the dynamic behaviour of the system. The initial and final states of the diagram are represented by a filled circle. An action state, which is the un-interruptible action of objects, which has a rounded corner rectangle as a figure. A fork node for splitting incoming flow and a join node that joins the flow and this is represented as a thicker horizontal line.

****

**Figure 3.2 Activity Diagram for Medical Diagnostic and Pharmacy Locator System**

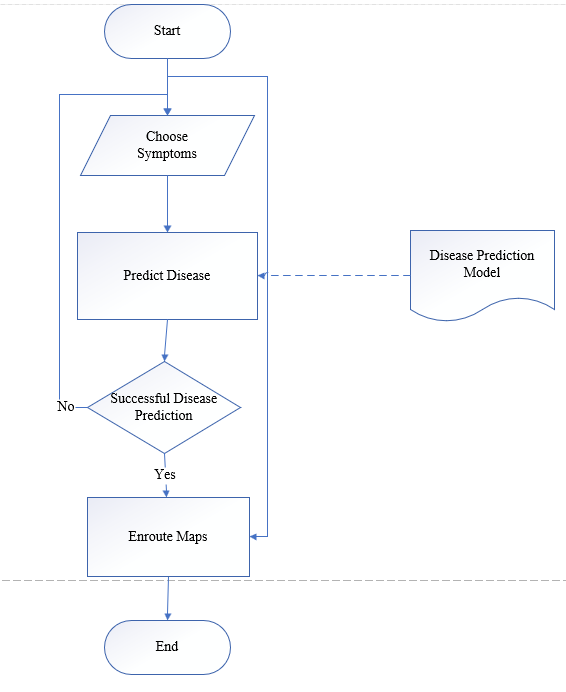
Figure 3.2 shows the activity diagram of the system. The fork node splits the flow after running the application, which displays the dashboard with two main activities, i.e., to make a diagnosis or to locate a pharmacy nearby. The fork node further splits the make diagnosis activity into the disease prediction activities. The join node synchronizes the diagnosing activity and the en-route activity. This results in successful diagnosis and en routing activity.

## 3.4 Implementation and Coding

This stage of the system shows how the application was developed and commenced in operation. It included the development of the model, system input and output data, and coherence of components for the desired output.

A flowchart is a graphical and visual representation of steps in a sequential order, which depicts the flow of algorithms and has various decisions required to perform processes. The operations of the system were modelled with a flowchart diagram. This aided in the translation of graphical ideas into codes.

### 3.4.1 Flow Chart

****

**Figure 3.3 Flow Chart for the Medical Diagnostic and Pharmacy Locator System**

Figure 3.3 shows the flow of the application and hence functions as follows:

1. The user chooses from a list of provided symptoms. These symptoms are converted to a list of binary values, i.e., (1)'s representing selected symptoms and (0)'s representing unselected symptoms.
2. The system then uses the disease prediction model to determine the disease of the patient.
3. The user proceeds to use the map after a successful disease prediction. Else, the user starts the input process again if there was an error.
4. The map function can also be accessed separately without going through the prediction process.

### 3.4.2 Symptom Input

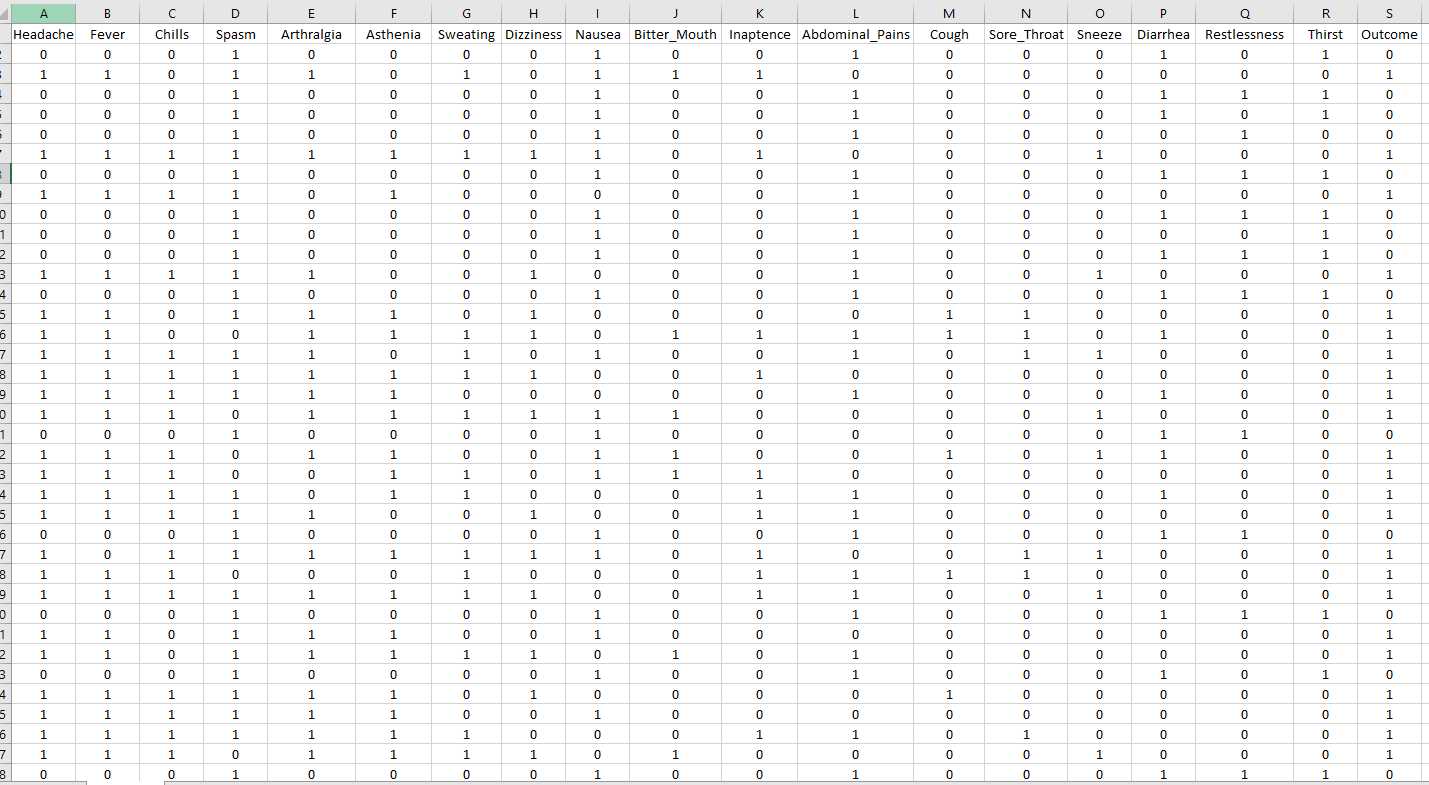
A list view of various symptoms such as headache, fever, chills, Nausea and Vomiting, Muscle Pain and Fatigue, Sweating, Chest or abdominal pain, Cough, Diarrhoea are all mandatory signs of malaria and cholera (Anon, 2017). Hence, the user chooses from these symptoms of which values are stored in a binary list. In Flutter, various symptom selected checkboxes can be "listened on selected" and then getting feedback of (1)'s and (0)'s. These values are then processed, and the respective disease output displayed.

### 3.4.3 Disease Prediction Model

Previous patient data consisting of various symptoms and their diagnoses were collected, processed, and used to build and train the Naïve Bayes model.

*Definition of Data.*

The patient data comprises 19 columns and 250 rows of which the last column from the left represents the outcome with binary values, i.e., a "0" representing the presence of malaria and a "1" representing cholera. The various patient data were converted to binary values, i.e., "1" shows the presence of such symptom by the patient, and "0" shows the patient did not experience that symptom. The 250 instances had 148 malaria cases and 102 cholera cases.



**Figure 3.4 A Screenshot of the first thirty instances**

*Naïve Bayes*

The Naïve Bayes Classifier (NBC) is a supervised Machine learning algorithm of which its classification technique uses the principles of the Bayesian theorem. This classifier is preferred because of its high accuracy and speed on large datasets, and it is very reliable.

NBC is considered naïve because of the assumption of conditional independence, i.e., the independence of attributes in a given class. The Bayes theorem states that: given a variable of a class **y**and dependent attribute vector **,**through to**,**

= (3.1)

Where, = independent variables.

= class vector or outcome.

= posterior probability of given .

and = prior probabilities.

= likelihood of if the class is true.

By using the naïve conditional independence assumption, the model can be simplified to

= (3.2)

And since is a constant (1) based on the input. Then the model can further be:

(3.3)

By using the Maximum A Posterior (MAP), which is the highest posterior probability, the NBC algorithm can classify a new case with various class levels with the formula:

= (3.4)

Where, = estimated class group or outcome

While in the Bayesian statistical inference, previous experience is the basis for the prior probabilities, with the NBC, this is related to the training dataset.

In this project:

1. Each class of possible classification which is denoted by will be malaria or cholera. And the class with the highest probability i.e. will be considered as the desired class or disease.
2. () = , denotes all the 18 attributes (symptoms) that contribute to either the malaria class or the cholera class.
3. = probabilities of various attributes.
4. The Bernoulli Naïve Bayes model was used because after careful consideration. The available training dataset had all entries as binary values, which perfectly fits with the functioning of the Bernoulli's Naïve Bayes model on datasets.

Example,

If a patient chooses symptoms from the list, for instance, having headaches, fever, chills, asthenia, sweating, dizziness, bitterness in the mouth, and sore throat, the data is retrieved and stored as a binary list.

The data is then fed into the disease prediction model, which then uses the Naïve Bayes algorithm for analysis. i.e., finds the probability if symptoms belong to a particular class.

This implies that the probabilities of malaria and cholera will be computed and compared. The class with the highest probability is considered as the predicted disease.

= \*(.

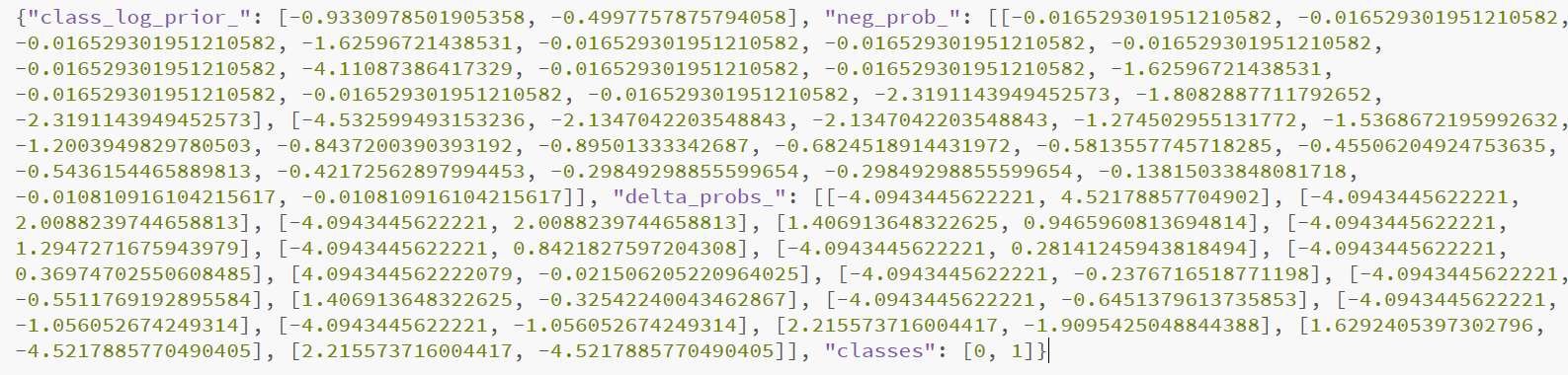
The same procedure is repeated for cholera:

= \*(.

If > ,

then the patient is diagnosed with malaria, else it is cholera.

Figure 3.5 below shows the various log probability of each class (class\_log\_prior), the negative(neg\_prob\_) and delta(delta\_probs\_) probabilities of all trained values.



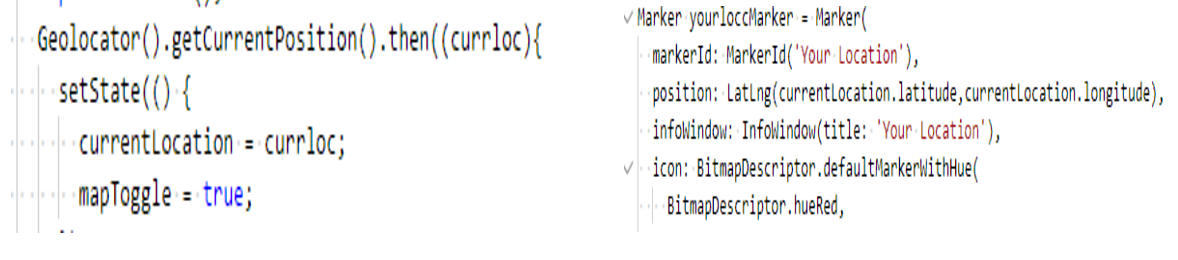
**Figure 3.5 A Snapshot of Sample Probabilities**

### 3.4.4 En-Route Maps

Google Maps Application Programming Interface (API) is an extensive package that was used in developing the flutter application. Maps API key was generated at google clouds, under map Software Development Kit (SDK), and registered in the respective pubs.yaml and XML files.

Google Maps API allows instantiation of markers at various locations, and this is done by updating each marker with various Global Positioning System (GPS) coordinates. Live device location, when permitted, is recorded by the application and shown on the map relative to the marked pharmacies in the area.

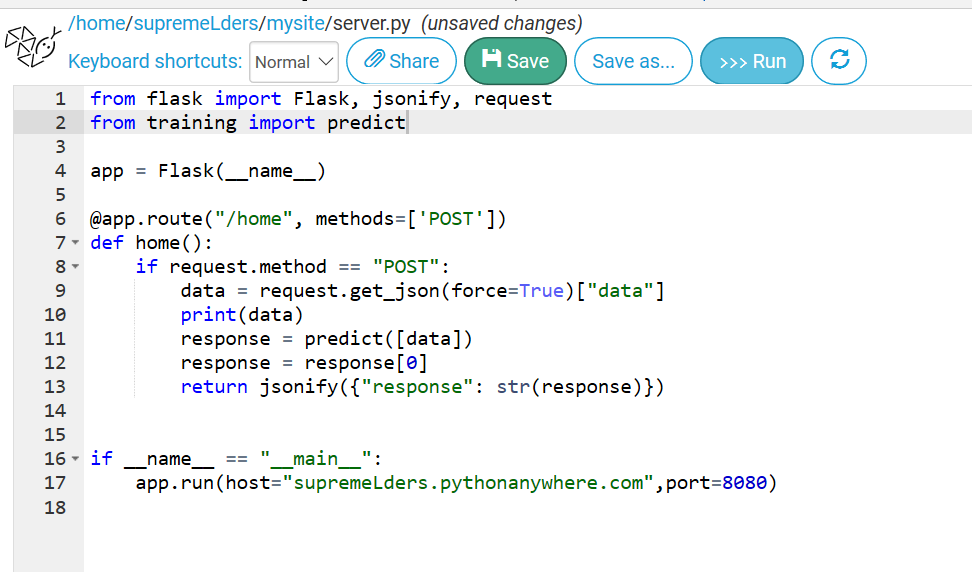
The map function can be accessed directly at a different pane in the application or launched after a successful disease diagnosis.



**Figure 3.6 An Initialization of Geolocation Feature.**

### 3.4.5 API Hosting

The Flask microweb framework was used to generate a Restful API, which serves as the server containing the model for prediction. The model was hosted with the online server "pythonanywhere.com".



**Figure 3.7 Hosting the Trained Model**

Figure 3.7 shows the snapshot of the trained model on the server, which passes predicted results to the flutter application in a JSON format and hence converts the result into the respective format, i.e., 0 represents malaria, and 1 represents cholera.

# CHAPTER 4

# SYSTEM TESTING AND DISCUSSION

## 4.1 Introduction

After the development of a system, the system is required to function as per the requirement and objectives. Hence the system was tested based on its functionality, performance, and reliability.

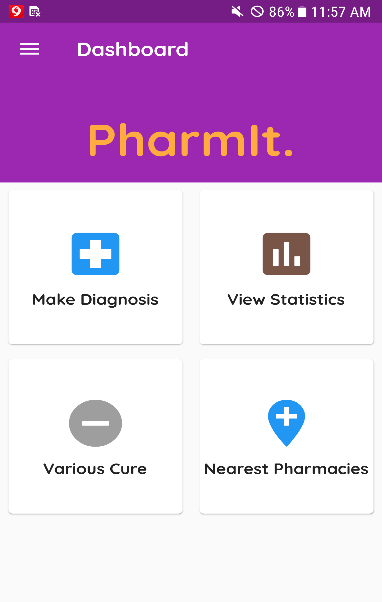
## 4.2 Specified Requirements

From the requirements elicited, the system was required to predict patient's diseases, i.e., malaria or cholera, based on their symptoms—moreover, aid in locating a pharmacy in the nearby location of the user. The disease prediction model built on the Naïve Bayes algorithm was used for the prediction, and the Google Maps API was used with markers plotted on various pharmacies in the location.

## 4.3 System Testing

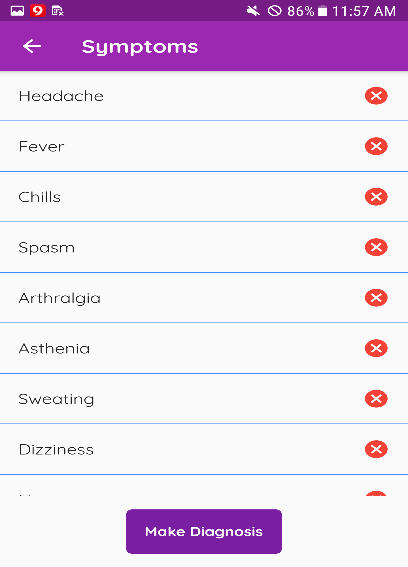
The front end of the mobile application has a list view of various symptoms, which are the major causes of both malaria and cholera, which is required by the patient to select or check them if the patient experiences any of such symptoms. By default, each state of the symptom's checkbox is set to "0" and changes to "1" if selected. Tapping on the "make a diagnosis" button submits various symptoms in a binary list to the model. The model then finds the disease with the highest probability based on the received data and then makes a final prediction.

### 4.3.1 Snapshot of Views

****

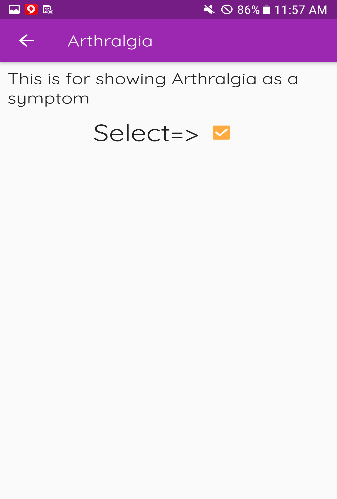
**Figure 4.1 Dashboard View**

Figure 4.1 shows the dashboard where the user chooses to either make a diagnosis or locate pharmacies nearby.



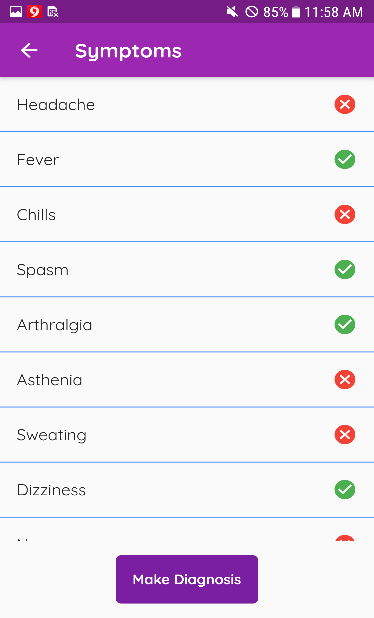
**Figure 4.2 Make Diagnosis Pane**

Figure 4.2 shows the list view of symptoms stored in the system. The user has to select from the provided list if he/she experiences any of such symptoms.



**Figure 4.3 Detailed Symptom Selection View**

Figure 4.3 gives a detailed view of a symptom list. The checkbox is selected if that symptom is to be chosen.

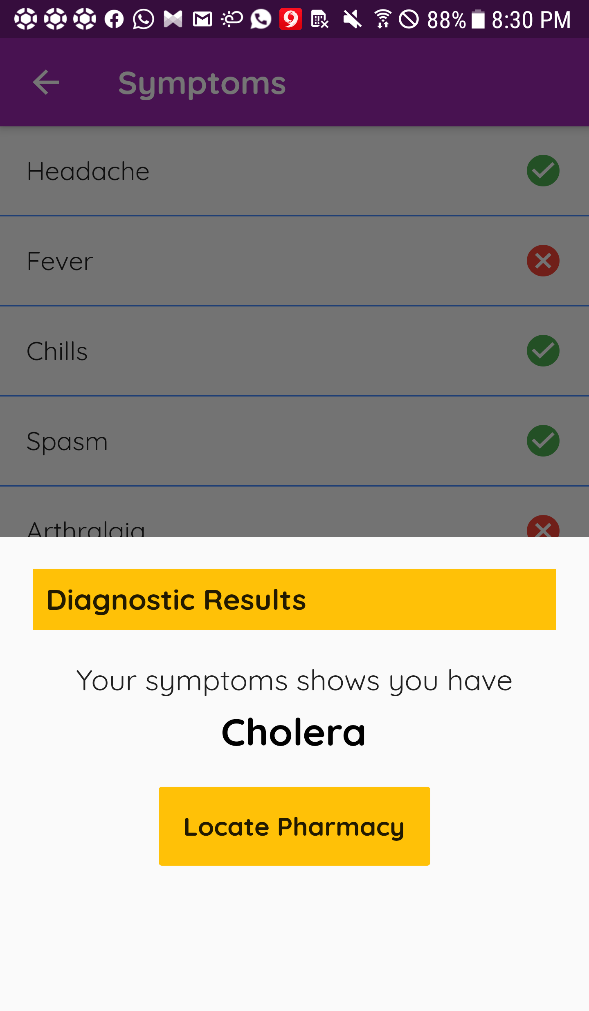


**Figure 4.4 A View of Selected Symptoms**

Figure 4.4, once selected, the notice icon changes to green, which indicates the selection of that symptom. Symptoms marked green are represented with (1) 's and the others (0) 's.

A prediction is made once the "make Diagnosis" button is pressed.

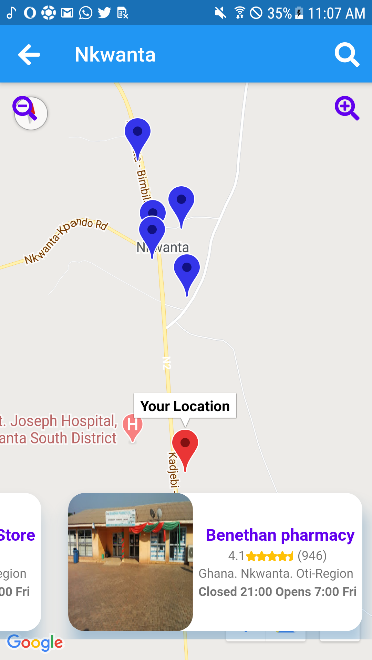
Figure 4.5 below shows the popup after tapping on the make diagnosis button. The result of the predicted disease is gotten from the model that is hosted online, which receives the (symptoms) and displays the output.



**Figure 4.5 Outcome of Diagnosis**

## 4.3.2 Testing the Map Feature

The map takes in the user's current location as an initial camera viewpoint and hence displays various pharmacies in the locality with their markers. The current user's location is indicated with a different marker relative to the plotted pharmacies on the maps. The user must allow users location agreement to allow the application to determine the current location of the device while displaying the pharmacies.



**Figure 4.6 A Snapshot of Maps**

## 

## 4.4 Confusion Matrix

A confusion matrix that provides details about the correct and incorrect classifications of each class is used in Figure 4.7 to present the results of the model. The diagonal elements represent the number of points that the predicted label is equal to the true label, and nondiagonal values are the mislabeled values. It incorrectly classified five malaria cases as cholera cases and two cholera cases as malaria cases but classified both 33 and 60 malaria and cholera cases, respectively.

The accuracy of the model can be calculated by taking the average of the diagonal values in the matrix.

Accuracy = (TP + TN) / (TP + TN + FP + FN) (4.1)

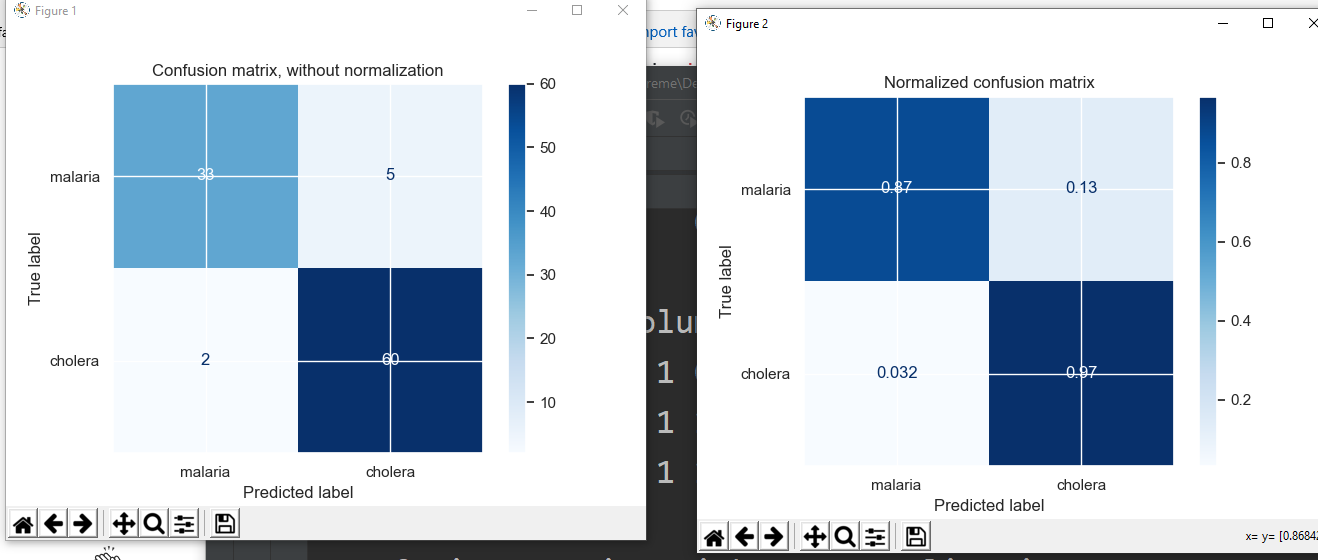
Where, TP = True Positives.

TN = True Negatives.

FP = False Positives.

FN = False Negatives.

The accuracy of the model is found to be 93.0%.



**Figure 4.7 Confusion Matrix of Results**

## 4.5 Drawbacks and Limitations

1. The system is limited to predicting only malaria and cholera.
2. The system requires an internet connection.
3. The system is limited to android operating devices.

# CHAPTER 5

# CONCLUSION AND RECOMMENDATION

## 5.1 Conclusion

An application that predicts diseases based on the patient's symptoms and provides the various location of pharmacies in the locality of the patient shows the relevance and application of Machine Learning in the medical field. Hence, it provides accurate diagnosis, which eliminates potential human errors as possible. Easy manipulation of the system is possible since it operates on an Android OS, which is widely used.

The successful contribution of this application will reduce the risk of getting an erroneous diagnosis by those who choose to self-diagnose.

## 5.2 Recommendation

It is recommended that:

1. A wide range of diseases can be captured by getting real medical data to train the model.
2. An iOS version of the application be developed and also Natural Language Processing (NLP) to build a model to capture various symptoms.

# REFERENCES

Anon. (2017), "Disease and Conditions", *www.mayoclinic.org/diseases-conditions/malaria/symptoms-causes.* Accessed: April 27, 2020.

Appiah, D.E. (2015), "Facts and Figures", *www.moh.gov.gh/facts-figures/.* Accessed: November 8, 2019.

Asare, A.N. (2017), "Ghana Annual Report", *www.ghanahealthservice.org/.* Accessed: October 12, 2019.

Baerheim, A. (2001), "The Diagnostic Process in General: Has it a Two-Phase Structure?", *Family Practice,* Vol. 18, No. 3, pp. 243 – 245.

Drislane, W.F., Akpalu, A. and Wegdam, H.H.J. (2014), "The Medical System in Ghana", *Yale Journal of Biology and Medicine,* Vol. 87, No. 3, pp. 321 – 322.

Graber, M. L. and Matthew, A. (2008), "Performance of a Web-Based Clinical Diagnosis Support System for Internists", *Journal of General Internal Medicine*, Vol. 23, No.1, pp. 37 – 39.

Ho, T. K. (1998), "The Random Subspace Method for Constructing Decision Forests", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pp. 832 – 844.

McCulloch, W. S. and Pitts, W. (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity", *The Bulletin of Mathematical Biophysics*, Vol. 5, No.4, pp. 115 – 125.

Meherwar, F., and Maruf, P. (2017), "Survey of Machine Learning Algorithm for Disease Diagnostic", *Journal of Intelligent Learning Systems and Application,* Vol. 9, No. 1, pp. 1 – 16.

Oguntimilehin, A., Adetunmbi, A. O. and Abiola O. B. (2013), "A Machine Learning Approach to Clinical Diagnosis of Typhoid Fever", *International Journal of Computer and Information Technology, Vol.2, No. 4,* pp. 672 – 673.

Pavithra, D. and Jayanthi, A. N. (2018), "A Study on Machine Learning Algorithm in Medical Diagnosis", *International Journal of Advanced Research in Computer Science, Vol.9, No. 4,* pp. 42 – 45.

Russell, S. and Norvig, P. (2002), *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2nd edition, 325 pp.

Shojania, G. K. and Dixon, W.M. (2016), "Estimating deaths due to medical error: the ongoing controversy and why it matters", *BMG Qual Saf*, Vol. 1, No.10, pp. 423 – 425.

Yinyeh, M. O, and Alhassan, S. (2015), "Expect Health Diagnosis and Monitoring System Software for Common Medical Conditions", *International Journal of Computer Application*, Vol. 124, No.1, pp. 17 – 22.