Chapter 1

INTRODUCTION

Machine learning is a branch of artificial intelligence that aims at solving real life engineering problems. It provides the opportunity to learn without being explicitly programmed and it is based on the concept of learning from data. It is so much ubiquitously used dozen a times a day that we may not even know it. The advantage of machine learning (ML) methods is that it uses mathematical models, heuristic learning, knowledge acquisitions and decision trees for decision making. Thus, it provides controllability, observability and stability. It updates easily by adding a new patient‘s record.

The amount of Medical data recorded in hospitals and its significance as an ever-growing source of information has been long known and proven. Though the importance of the information hidden in these records has never been doubted, this data has mostly been used only for clinical purposes. Only recently has this been properly mined for valuable information to be used for research and to develop systems that assist the medical fraternity. Mostly, the systems that make use of this information are domain specific systems that predict diseases restricted to their area of specialization (like heart, brain etc.). But these systems are limited and are not applicable to the whole medical dataset. Our system uses this vast storage of information so that diagnosis based on this historical data can be made. This system aids medical diagnosis in the whole dataset by computing the probability of occurrence of a particular ailment from the medical data. The system mines the data using a unique algorithm which increases accuracy of such diagnosis by combining Neural Networks and Differential Diagnosis all integrated into one single approach. The strengths of kNN, Feed Forward algorithm, SOM, back propagation and P2P Grid Architecture are used to make the system unique and effectively enhanced symptoms.

1.1 Definitions

Machine Learning- It is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

Differential Diagnosis: In medicine, a differential diagnosis is the distinguishing of a particular disease or condition from others that present similar clinical features. Differential diagnostic procedures are used by physicians to diagnose the specific disease in a patient, or, at least, to eliminate any imminently life-threatening conditions. Often, each individual option of a possible disease is called a differential diagnosis (e.g. acute bronchitis could be a differential diagnosis in the evaluation of a cough, even if the final diagnosis is common cold).

Neural Networks- Such systems "learn" to perform tasks by considering examples, generally without being programmed with task-specific rules. For example, in image recognition, they might learn to identify images that contain cats by analyzing example images that have been manually labeled as "cat" or "no cat" and using the results to identify cats in other images. They do this without any prior knowledge of cats, for example, that they have fur, tails, whiskers and cat-like faces. Instead, they automatically generate identifying characteristics from the examples that they process.

Multilayer perceptron (MLP) is a feed-forward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses back-propagation for training the network. MLP is a deep learning method.

1.2 Project Outline

Chapter 1: We propose a system for diagnosing liver based on its symptoms and patients history and predict the probable diseases causing the symptoms.

Chapter 2: This system provides the interface for doctors to enter the symptoms and appropriate lab report values and diagnose the patient. Visualization of the most probable disease is shown for better understanding.

Chapter 2

REVIEW OF LITERATURE

2.1 System Study

A detailed study to determine whether, to what extent, and how automatic data-processing equipment should be used; it usually includes an analysis of the existing system and the design of the new system, including the development of system specifications which provide a basis for the selection of equipment.

2.2 Proposed Work

Our proposed system works on diagnosing diseases using the symptoms related to Liver. We present KNN, ANN Back Propagation for differential diagnosis of diseases. KNN, ANN helps in predicting the final possible diseases with recurring classification of data. The main architecture of KNN is presented, it includes the following three major phases: (1) Input of data; (2) Dataset classification; (3) Prediction of diseases.

2.2.1 Problem Statement

- Diagnosing the patient and treating for the permanent cure has been challenging forever. Patient would have undergone irreversible damage to his health condition even before identifying the actual disease he has been suffering from.
- Thus permanent cure can be achieved by a patient when he is diagnosed and treated right within the golden period of the disease.
- Developing an automated system for diagnosing the patient provided with the patient's conditions, test results and history and listing out accurately the most appropriate diseases causing the symptoms can improve the rate of diagnosing drastically, increase the probability of permanent cure and thus saving millions of life.

2.2.2 Existing System

- Design an automated medical decision making system and help doctors in minimizing misdiagnosis.
- The database for a specific disease and related symptoms is created.
- Later proposed an algorithm to predict the Swine Flu disease on several attributes.
- They have observed the outcomes for machine learning algorithms such as Bayesian classification, Decision Tree, and compared it with the neural networks algorithms such as dynamic node creation (DNC) and feed-forward neural network construction.
- The system is more reliable and faster than the conventional systems.
- The problem statement only focuses on a single disease i.e. Swine Flu.
- Each disease will have to have its own algorithm to be precise and accurate.

2.2.3 Proposed System

We propose to design an automated diagnosing system for diagnosing various kinds of diseases from which liver can be affected. This system takes clinical data as input and gives out diagnose predicting the probable diseases causing the symptoms.

2.3 Scope of the project

Most of the time it takes a lot of time in diagnosing a disease and sometimes its even misdiagnosed. The motive of this project is to diagnose disease using symptoms with high rate of accuracy and less execution time. We hope to bridge the gap of lack of knowledge during crucial times. We've made use of trending technology i.e Machine Learning. We aim to collect relevant data related to our fields of study. Train this data as per our proposed algorithm and modules and predict the disease that's causing the symptoms.

Chapter 3

SYSTEM REQUIREMENTS SPECIFICATION

3.1 Functional Requirements

The functional requirements for a system describe what the system should do. These requirements depend on the type of software being developed, the general approach taken by the organization when writing requirements. The functional system requirements describe the system function in detail, its inputs and outputs, exceptions and so on.

Functional requirements are as follows:

- Develop a robust solution which can predict the vulnerability of a disease given basic symptoms, patients history and lab reports etc.
- Train the model against enough data sets to maintain the accuracy level above 90%
- Optimize the model to rise the accuracy level even further
- Provide data visualization features for the customers to get more valuable insight of the patient health
- Expose the solution over the cloud as a service so that the solution can be re-usable by any third party applications.

3.2 Non-Functional Requirements

Nonfunctional requirements, as the name suggests, are requirements that are not directly concerned with the specific functions delivered by the system. They may relate to emergent system properties such as reliability, response time and store occupancy. Alternatively, they may define constraints on the system such as capabilities of I/O devices and the data representations used in system interfaces.

The nonfunctional requirements are as follows:

- Should be easier to access it from the various browsers available.

- Response time of the applications should reflect the real time observations.
- The algorithm should never fail in any of the test cases.
- There shouldn't be any security concerns on the merged data.
- Each user's activity should be separated from the other user's activities

3.3 Hardware Requirements

- System : Intel processor with 2.2GHz and above.
- Hard Disk : 120 GB.
- Input Devices : Keyboard, Mouse
- Ram : 4 GB

3.4 Software Requirements

- Operating system : Windows 10
- Coding Language : Python
- IDE : PythonIDLE

Chapter 4

SYSTEM DESIGN

4.1 Design Overview

The Design Summary allows you to quickly access design overview information, reports, and messages. By default, the Design Summary appears in the Workspace when you open a project, and it displays information specific to your targeted device and software tools.

4.2 System Architecture

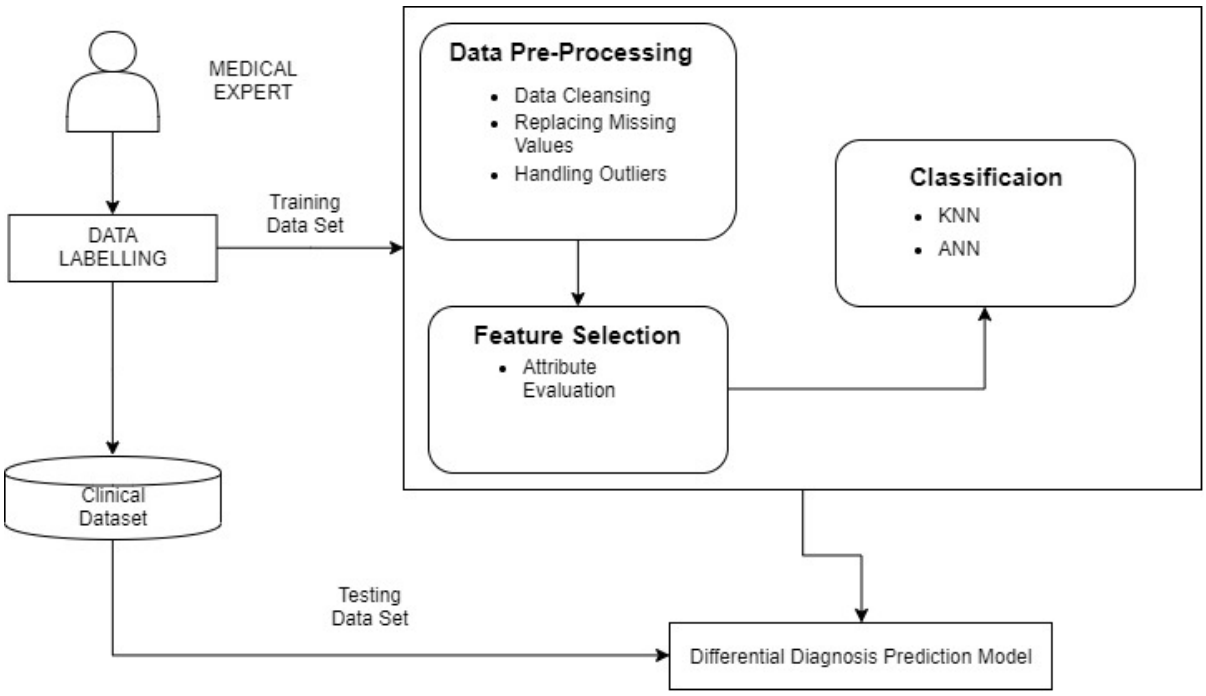

Fig 4.2 System Architecture

4.3 Data Flow Diagrams

4.3.1 Data Flow Diagram - Level 0

A **data-flow diagram** (DFD) is a way of representing a flow of a data of a process or a system (usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow, there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart.

Level 0 describes the overall process of your project. It takes the input as disease data and processes through algorithm to get a trained model.

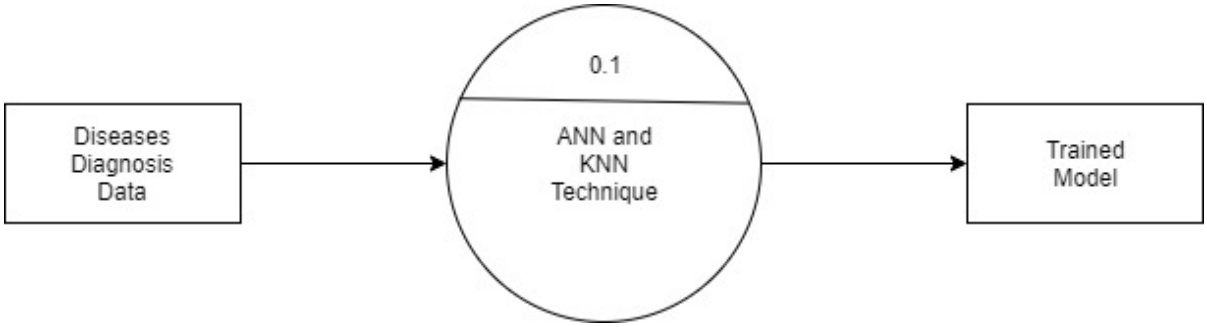

Fig 4.3.1 DFD-Level 0

4.3.2 Data Flow Diagram - Level 1

Level: 1 Describes the data flow of our project from one instance to the next. It takes the input of database and processes the data, performs the neural network algorithm to get a classified data.

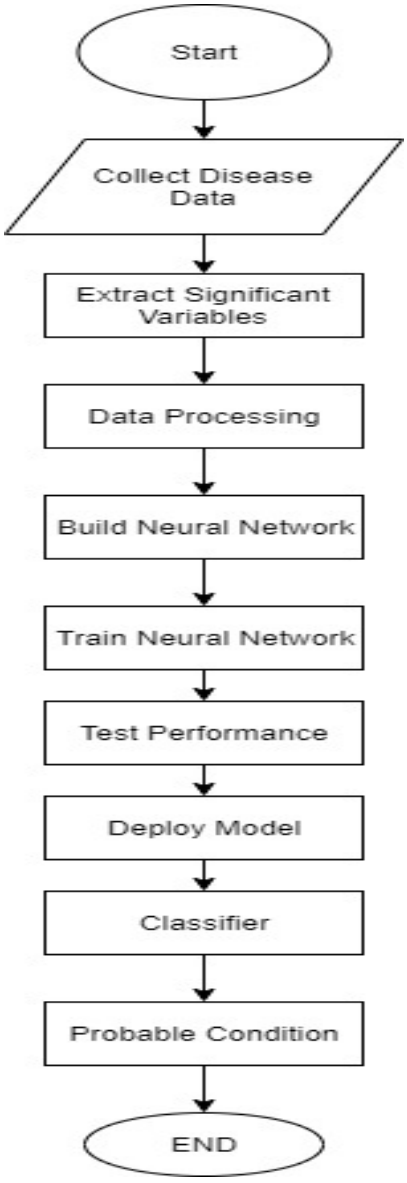

Fig 4.3.2 DFD-Level 1

4.3.3 Data Flow Diagram - Level 2

Level: 2 Describes the final step of our project. The classified data is used to take the inputs and give a probability of the disease occurring and the new data shall be updated in the database.

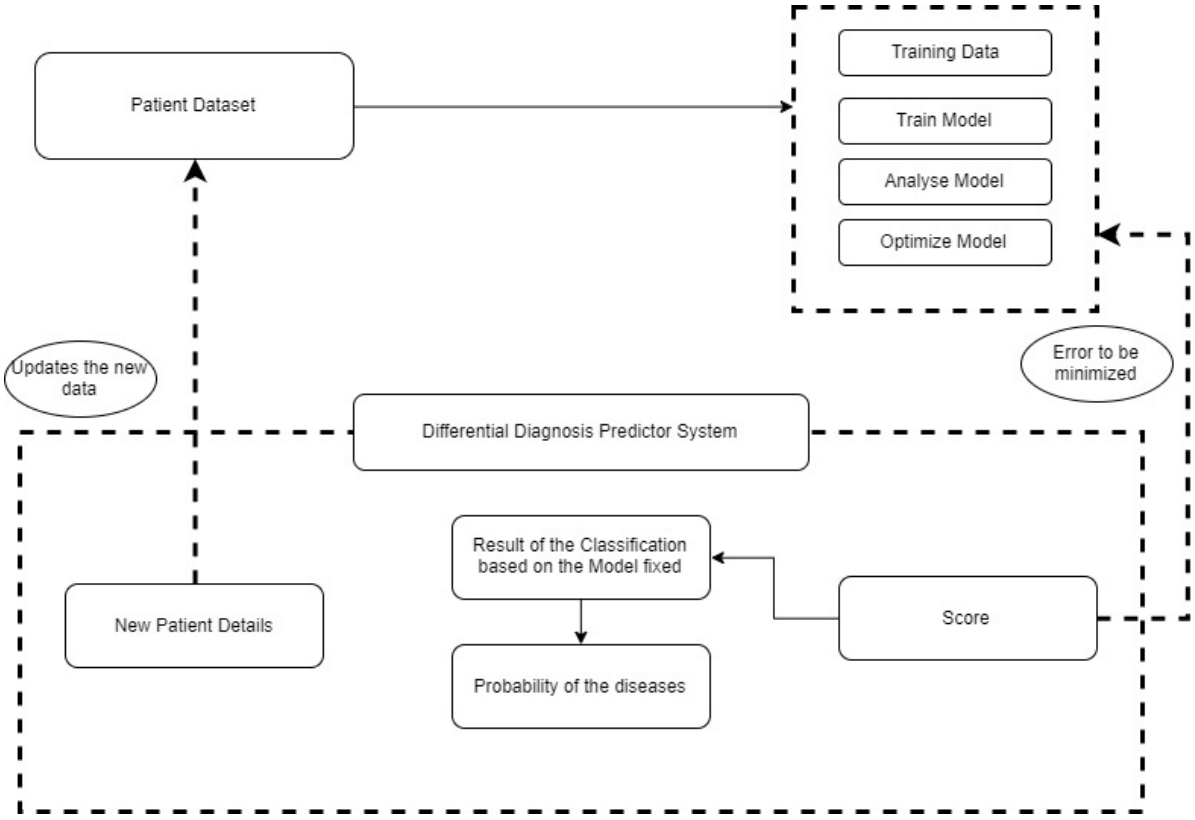

Fig 4.3.3 DFD-Level 2

4.4 Use Case Diagrams

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

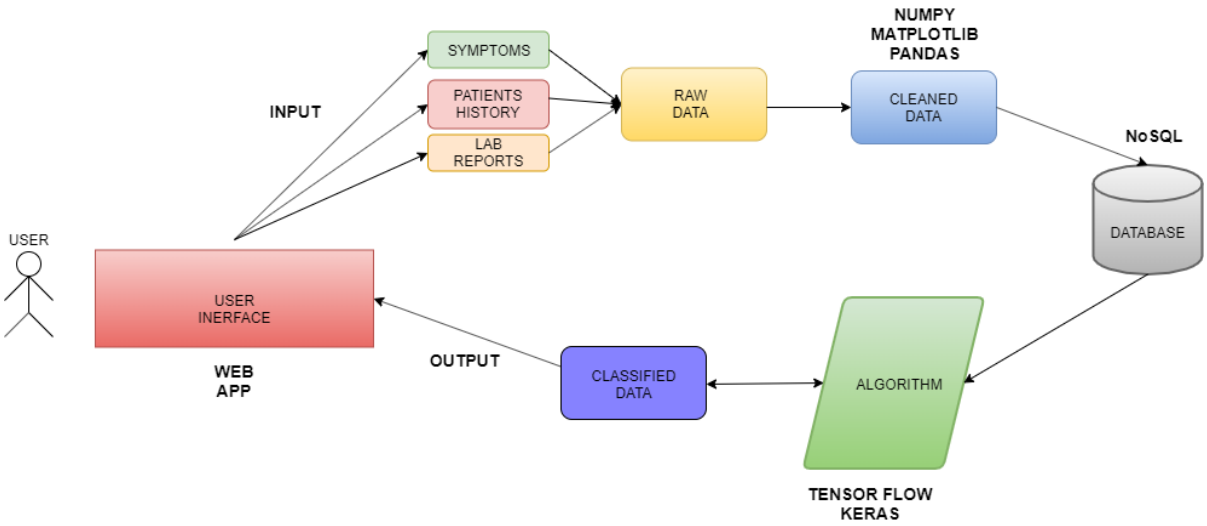

Fig 4.4 Use case diagram

4.5 Class Diagrams

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

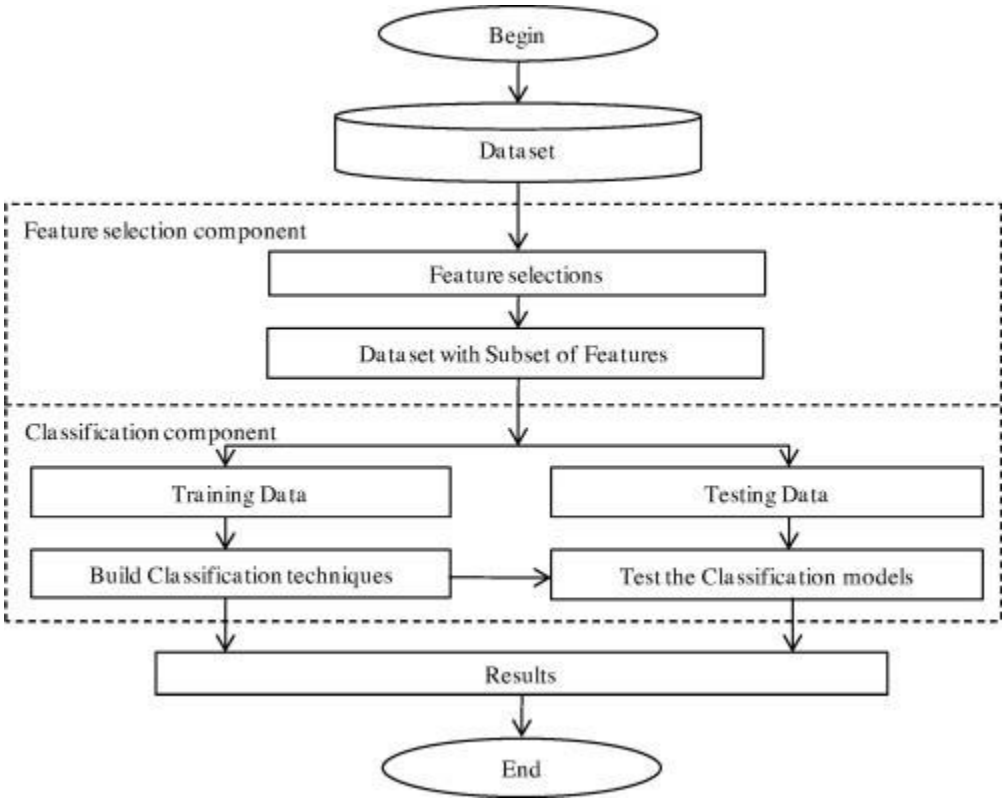

Fig 4.5 Class diagram

4.6 Sequence Diagrams

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

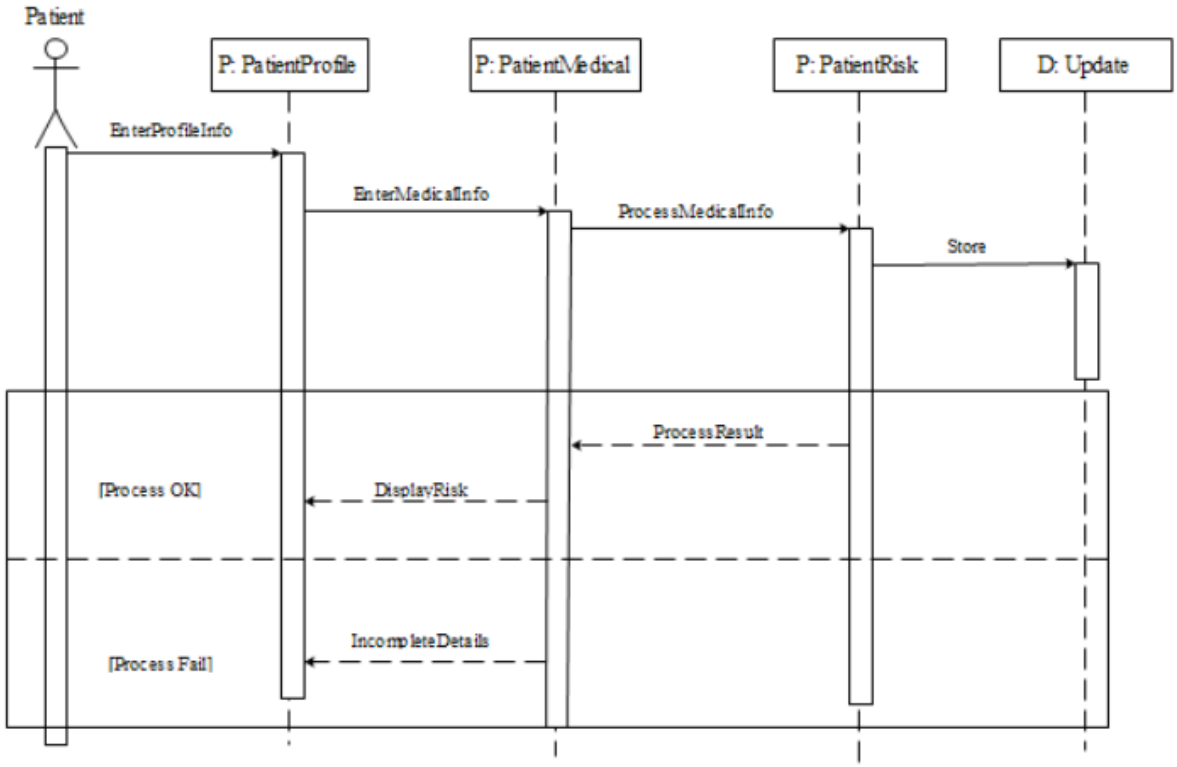

Fig 4.6 Sequence diagram

4.7 Activity Diagrams

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

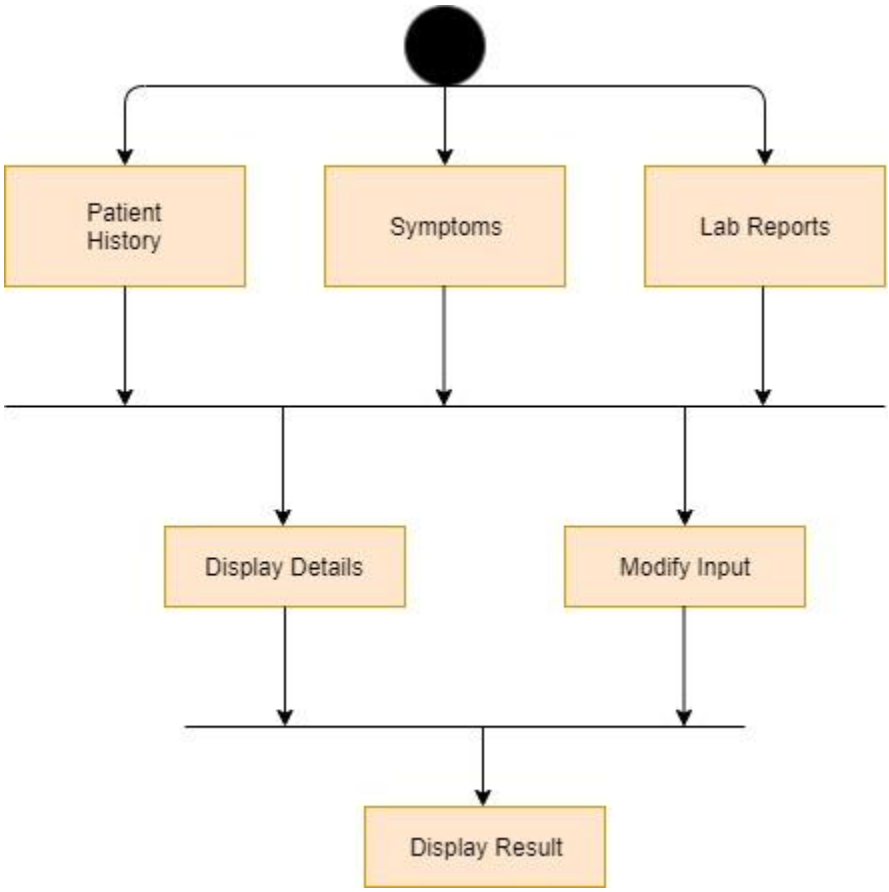

Fig 4.7 Activity diagram

4.8 Modules of the Project

4.9.1 Module 1

Module Name: - **Data Pre-processing Phase**

Functionality:-The function of data preprocessing is of collecting raw data and then transform this data into a format that can be processed by the python libraries.

Input:-Disease prediction database.

Output:-Processed dataset with attributes and values which are not dependent.

Libraries used: Python libraries namely numpy and pandas.

4.9.2 Module 2

Module Name:-Training the model.

Functionality:-The function of this phase is to train the model with the data which is available after processing using an algorithm.

Input:-Attributes of the dataset with their values.

Output:-The model after the learning process will try to predict the possibility of a disease from the given list of symptoms which are attributes in this case.

Algorithms Used:-KNN algorithm and ANN back-propagation algorithm

4.9.3 Module 3

Module Name: - Output Phase.

Functionality:-In this module we will display the possible disease on the application designed by us.

Input: - The predicted disease with its probability.

Output: - Display of the disease with its probability and a brief description of the disease along with the possible symptoms which may further occur.

BIBLIOGRAPHY

[1] Murali Ravuri, Anitha Kannan, "From expert system to machine-learned diagnosis models, Proceedings of Machine Learning Research , Machine Learning for Healthcare,85,1–16, 2018.

[2] Dhaval Raval, Dvijesh Bhatt, "Medical diagnosis system using Machine Learning, International Journal of computer science , Machine Learning for Healthcare, Volume 7,1–6, 2016.

[3]Dr A N Jayanthi, Ms D Pavithra, "A Study on Machine Learning Algorithm in Medical Diagnosis", International Journal of Advanced Research in Computer Science, Volume 9, No 4, July-August 2018.

[4] Shaik Razia,N Vamsi Krishna,"A Review On Disease Diagnosis Using Machine Learning Techniques" , International Journal of Pure and Applied Mathematics, Volume 117 No. 16 2017, 79-85.

[5] Alex Yee," Machine Learning Applied to the Detection of Retinal Blood Vessels, Stanford University.