SAVITRIBAI PHULE PUNE UNIVERSITY

A PROJECT REPORT ON

TECHCARE

SUBMITTED TOWARDS THE PARTIAL FULFILMENT OF THE REQUIREMENTS OF

BACHELOR OF ENGINEERING(Computer Engineering)

BY

Adesh Tajane	**403080**
Ronish Zadode	**403086**
Ashutosh Kedar	**402065**
Prabhat Pandey	**403049**

UNDER THE GUIDANCE OF

Prof. Shilpa Sonawani(Internal Guide)

Department Of Computer Engineering
MAEERs MAHARASHTRA INSTITUTE OF TECHNOLOGY
Kothrud, Pune 411 038

2015-2016

MAHARASHTRA ACADEMY OF ENGINEERING AND EDUCATIONAL RESEARCH'S

MAHARASHTRA INSTITUTE OF TECHNOLOGY

PUNE

DEPARTMENT OF COMPUTER ENGINEERING

C E R T I F I C A T E

This is to certify that

TECHCARE

Submitted by

Adesh Tajane	**403080**
Ronish Zadode	**403086**
Ashutosh Kedar	**402065**
Prabhat Pandey	**403049**

is a bonafide work carried out by Students under the supervision of Prof. Shilpa Sonawani and it is submitted towards the partial fulfilment of the requirement of Bachelor of Engineering (Computer Engineering).

Prof. Shilpa Sonawani
Internal Guide
Dept. of Computer Engg.

Dr. V. Y. Kulkarni
H.O.D.
Dept. of Computer Engg.

Dr. L. K. Kshirsagar
Principal
Maharashtra Institute of Technology

Signature of Internal Examiner

Signature of External Examiner

PROJECT APPROVAL SHEET

A Project title

TECHCARE

Is successfully completed by

Adesh Tajane	**403080**
Ronish Zadode	**403086**
Ashutosh Kedar	**402065**
Prabhat Pandey	**403049**

at

DEPARTMENT OF COMPUTER ENGINEERING

MAHARASHTRA INSTITUTE OF TECHNOLOGY

SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE

ACADEMIC YEAR 2017-2018

Prof. Shilpa Sonawani	Dr. V. Y. Kulkarni
Internal Guide	H.O.D.
Dept. of Computer Engg.	Dept. of Computer Engg.

Abstract

The project is an application of data mining and machine learning techniques to the time series data of the disease development based on the temperature changes associated. The project also includes high end application of latest methodologies of prediction. It is a real time prediction model. It includes application on most widely used android platform with three tier application.

Keywords:
Data Mining, Risk Prediction, Machine Learning.

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on 'TECH-CARE'.

We would like to take this opportunity to thank our internal guide Prof. Shilpa Sonawani for giving us all the help and guidance we needed. We are really grateful for her kind support. Her valuable suggestions were very helpful.

We are also grateful to Dr. V. Y. Kulkarni, Head of Computer Engineering Department, Maharashtra Institute of Technology,Pune for her indispensable support, suggestions.

Adesh Tajane

Ronish Zadode

Prabhat Pandey

Ashutosh Kedar

(B.E. Computer Engineering)

Contents

List of Figures

List of Tables

Chapter 1

SYNOPSIS

1.1 PROJECT TITLE

The title of our project is 'TechCare'

1.2 PROJECT OPTION

The option of our project is Inhouse

1.3 INTERNAL GUIDE

The internal guide of our project is Prof. Shilpa Sonawani.

1.4 SPONSORSHIP AND EXTERNAL GUIDE

Our project is Inhouse.

1.5 TECHNICAL KEYWORDS

A.Hardware

- Realtime risk prediction
- Android phone

B.Architecture

- Client Server Architecture
- Realtime Cloud Database

C.Networking

- Three tier

1.6 PROBLEM STATEMENT

To create a real-time machine learning model to predict the risks in disease occurrence and growth of diseases affected by climate and to plan the daily routine accordingly based on the weather forecast.

1.7 ABSTRACT

The project is an application of data mining and machine learning techniques to the time series data of the disease development based on the temperature changes associated. The project also includes high end application of latest methodologies of prediction. It is a real time prediction model. It includes application on most widely used android platform with three tier application.

1.8 GOALS AND OBJECTIVES

1. To be able for users to make plans according health preferences.
2. To provide the user with before-hand risk analysis of specific diseases and allergies.
3. To improve general health status of the user.

1.9 RELEVANT MATHEMATICS ASSOCIATED WITH THE PROJECT

1.9.1 MATHEMATICAL MODEL

A mathematical model is a description of system using mathematical concepts and language. A model may help to explain a system and to study the effects of different components, and to make predictions about behaviour.

Mathematical models can take many forms, including but not limited to dynamical systems, statistical models, differential equations, or game theoretic models. These and other types of models can overlap, with a given model involving a variety of abstract structures. In general, mathematical models may include logical models.

In many cases, the quality of a scientific field depends on how well the mathematical models developed on the theoretical side agree with results of repeatable experiments. Lack of agreement between theoretical mathematical models and experimental measurements often leads to important advances as better theories are developed.

Mathematical Model for Android Application 'Techcare':

- I: Set of Inputs
- O: Set of outputs
- F: Functions
- Sc: Success cases
- Fc: Failure cases

- I= {I1, I2, I3} where,
  - I1= Patient Datasets for model training.
  - I2= User Daily Health data.
  - I3= Weather forecast.
- O: {O1, O2} where,
  - O1= Risk prediction of the disease.
  - O2= Daily planning.

- F: {F1} where,
  - F1= Artificial Neural Network to analyse and predict the risk of disease occurrence based of weather forecast.
- Sc: {Sc1, Sc2, Sc3, Sc4} where,
  - Sc1:1. Proper training database based on the conditions and climate.
  - Sc2. Correct details entered by the user.
  - Sc3. Accuracy of weather forecast maximum.
  - Sc4. Proper ANN Model development.
- Fc: {Fc1, Fc2, Fc3, Fc4} where,
  - Fc1. Improper training Database based on the conditions and climate.
  - Fc2. Incorrect details entered by the User.
  - Fc3. Accuracy of weather forecast minimum.
  - Fc4. Improper ANN Model development.

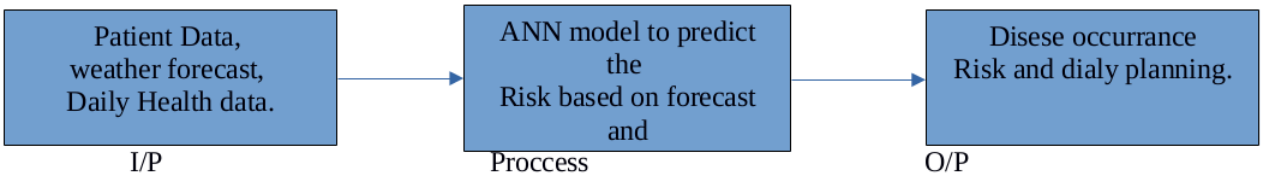

Figure 1.1: Mathematical model

1.10 NAMES OF CONFERENCES / JOURNALS WHERE PAPERS CAN BE PUBLISHED

- IEEE/ACM Conference/Journal 1
- Conferences/workshops in IITs

- Central Universities or SPPU Conferences
- IEEE/ACM Conference/Journal 2

1.11 PLAN OF PROJECT EXECUTION

Table 1.1: Plan of Project Execution

From	To	Task	Status
27–06–2017	30–06–2017	Group Formation and finalization	Done
01–07–2017	15–07–2017	Topic Search	Done
16–07–2017	20–07–2017	Preliminary Information Gathering	Done
21–07–2017	28–07–2017	Project Dicsussion with Project Coordinator and topic finalization	Done
01–08–2017	25–08–2017	Synopsis preparation and submission	Done
25–08–2017 19–09–2017	10–09–2017 24–09–2017	Detailed Literature Survey .	Done .
10–10–2017	30–10–2017	Preparing Interim report	Done
01–11–2017	15–12–2017	Language Study	In process
03–01–2018	20–01–2018	Androids Study	Not Done
21–01–2018	04–03–2018	Coding and Implementation	Not Done
05–03–2018	21–04–2018	Testing	Not Done
13–04–2018	21–04–2018	Final Documentation	Not Done
25–04–2018	20–05–2018	Final Project Report	Not Done

Chapter 2

TECHNICAL KEYWORDS

2.1 AREA OF PROJECT

The area of our project is Machine Learning and Data Mining.

2.2 TECHNICAL KEYWORDS

A.Hardware

- risk prediction
- Android phone

B.Architecture

- Client Server Architecture
- Realtime Database Connectivity

C.Networking

- Server

Chapter 3

INTRODUCTION

3.1 PROJECT IDEA

The overall Idea of the project is to provide user with before hand daily health related risks based on user's daily routine and weather changes as well as historic health data. The project's idea is to formulate a efficient Electronic Health Record so that the prediction is most accurate.

3.2 MOTIVATION OF THE PROJECT

It is evident from the historic data as well as human knowledge that human general health is closely related to the climatic conditions and the occurrence of diseases is greatly affected by the climate change. This inference gave us the motivation to apply advanced computer science concepts like machine learning and data mining to the health data available to extract certain knowledge and apply it to the current realtime health data so that future occurrence of diseases can be accurately predicted.

3.3 LITERATURE SURVEY

1.Data Driven Analytics for Personalized Healthcare Jianying Hu,Adam Perer , and Fei Wang, Springer International Publishing Switzerland 2016 C.A. Weaver et al, Healthcare Information Management Systems- Cases, Strategies, and Solutions, Health Informatics,DOI 10.1007

The concept of Learning Health Systems (LHS) is gaining momentum as more and more electronic healthcare data becomes increasingly accessible. The core idea is to enable learning from the collective experience of a care delivery network as recorded in the observational data, to iteratively improve care quality as care is being provided in a real world setting. In line with this vision, much recent research effort has been devoted to exploring machine learning, data mining and data visualization methodologies that can be used to derive real world evidence from diverse sources of

healthcare data to provide personalized decision support for care delivery and care management. It gives an overview of a wide range of analytics and visualization components have developed, examples of clinical insights reached from these components, and some new directions are taken.

2.Curve relativity analyse for relationship between blood pressure and atmospheric temperature using Matlab, 2009 International Conference on E-Learning, E-Business, Enterprise Information Systems, and E-Government

Although there are many evidence indicated that human blood pressure is influenced by atmospheric temperature , there are not useful method or tool use to evaluating the correlation relationship between atmospheric temperature and BP(blood pressure) still .The paper collected a serial date of atmospheric temperature and BP which were recorded by a family .The atmospheric temperature data were reported by weather bureau and the BP data were measured by oneself in electronic blood-pressure meter(HEM6000) .It was made in Japan by OMRON company . Everyday BP was measured for two times .In the morning, the blood-pressure was marked for a.m. BP then in the afternoon ,the blood-pressure was marked for p.m. BP .there are six groups data we used to experiment. All of them include a.m. SBP(shrink blood pressure),a.m. DBP(distend blood pressure), p.m. SBP, p.m. DBP LT(lowest atmospheric temperature) HT(highest atmospheric temperature) .Each of them was recorded during 90 days from day to day. The characters of them are discrete and there are some random error exist .There are no general methods ready to evaluate the relationship between them. For evaluating the relationship between them objectively .The paper used many methods and analysed from different aspect to make sure how the correlation degree are among each others .

3.Various R Programming Tools for Plotting Data, Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen

Moeller, Marc Schwartz, Bill Venables,

Various R programming tools for plotting data. Calculating and plotting locally smoothed summary function as ('bandplot', 'wapply'), enhanced versions of standard plots ('barplot2', 'boxplot2', 'heatmap.2', 'smartlegend'), manipulating colors ('col2hex', 'colorpanel', 'redgreen', 'greenred', 'bluered', 'redblue', 'rich.colors'), calculating and plotting two-dimensional data summaries ('ci2d','hist2d'), enhanced regression diagnostic plots ('lmplot2', 'residplot'), formula-enabled interface to 'stats::lowess' function ('lowess'), displaying textual data in plots ('textplot', 'sinkplot'), plotting a matrix where each cell contains a dot whose size reflects the relative magnitude of the elements ('balloonplot'), plotting "Venn" diagrams ('venn'), displaying Open-Office style plots ('ooplot'), plotting multiple data on same region, with separate axes ('overplot'), plotting means and confidence intervals ('plotCI', 'plotmeans'), spacing points in an x-y plot so they don't overlap ('space').

4.A soft-computing ensemble approach (SEA) to forecast Indian summer monsoon rainfall, Nisha Kurian, a T. Venugopal, b Jatin Singh a and M. M. Ali * b a Skymet Weather Services, Noida, India Department of Physics, Novosibirsk State University, Russia

Agriculture is the backbone of the Indian economy and contributes 16percent of gross domestic product and about 10percent of total exports. Hence, accurate and timely forecasting of monthly Indian summer monsoon rainfall is very much in demand for economic planning and agricultural practices. Several methods and models, comprising dynamic and statistical models and combinations of the two, exist for monsoon forecasting. Here, a multi-model ensemble approach, combined with an artificial neural networking technique, was used to develop a soft-computing ensemble algorithm (SEA) to forecast the monthly and seasonal rainfall over the Indian subcontinent. Forecasts using January to May initial conditions along with observations during 19822014 were used to develop the model. The SEA compares well with observations.

5.Climate Change And Infectious Diseases

Changes in infectious disease transmission patterns are a likely major consequence of climate change. It is needed to learn more about the underlying complex causal relationships, and apply this information to the prediction of future impacts, using more complete, better validate.

6.Extreme Allergies and Global Warming National wildlife federation, 2010

Unchecked global warming will worsen respiratory allergies for approximately 25 million Americans. Ragweedthe primary allergen trigger of fall hay fevergrows faster, produces more pollen per plant, and has higher allergenic content under increased carbon dioxide levels. Longer growing seasons under a warmer climate allow for bigger ragweed plants that produce more pollen later into the fall. Springtime allergies

to tree pollens also could get worse. Warmer temperatures could allow significant expansion of the habitat suitable for oaks and hickories, which are two highly allergenic tree species. Changing climate conditions may even affect the amount of fungal allergens in the air. More airborne allergens could mean more asthma attacks for the approximately 10 million Americans with allergic asthma. Global warming may also exacerbate air pollution, which interacts with allergens to trigger more severe asthma attacks. Cities pose the biggest health threats for asthmatics because the urban heat island effect can exacerbate both pollen production and air pollution. These potential impacts of global warming could have a significant economic impact: allergies and asthma already cost the United States more than 32 billion annually in direct health care costs and lost productivity. Poison ivy also grows faster and is more toxic when carbon dioxide increases in the atmosphere. More than 350,000 cases of contact dermatitis from exposure to poison ivy are already reported in the United States each year. These numbers are likely to increase if poison ivy grows faster and becomes more abundant. The reactions may also become more severe because poison ivy produces a more potent form of urushiol, the allergenic substance, when carbon dioxide levels are higher. We must act now to reduce risks for allergy and asthma sufferers. An essential first step is to reduce global warming pollution to avoid the worst impacts, and enable allergy sufferers to continue enjoying the great outdoors. At the same time, states, communities, and home owners should undertake smart community planning and landscaping, with attention to allergenic plants and urban heat island effects, to limit the amount of pollen and other allergens that become airborne.

Chapter 4

PROBLEM DEFINITION AND SCOPE

4.1 PROBLEM STATEMENT

In our project we intend to analyse the data available for various patients of certain diseases which are primarily affected by change in daily routine , climate and diet . from this analysis we recognize certain pattern related to symptoms , cause and treatment of the disease . Based on this analysis we create a model to predict the risks , prevention and precautions to be taken by the patients with similar pattern . Also the prediction will be supported by the climate forecast and current changes in medical health of the patient . Based on the analysed and detected patterns , the current medical health and climate changes to be confronted by the patient we aim to give in advance prediction to the patient . We also intend to provide the precautionary methods . Along with prediction we intend to prompt the patient about taking care of his health based on his personal health status and the climate changes .

4.1.1 Goals and Objectives

1. To be able for users to make plans according health preferences.
2. To provide the user with before-hand risk analysis of specific diseases and allergies.
3. To improve general health status of the user.

4.1.2 Statement of Scope

We describe what features are in the scope of the software and what are not in scope of the software to be developed.

1. In our project techcare we are attempting to provide risk prediction of occurrence of disease and it's progress based on the machine learning model we aim to create . We want to make a system which would update the health status of a person daily and tell him to improve or change his routines . Also we want to

predict and tell the user to take precautions and identify symptoms properly . All this is related to the climate change and it's affect on daily health . We're using the climate forecast and historical data of the user to do reduction.

Chapter 5

SYSTEM ARCHITECTURE AND REQUIREMENTS

5.1 HARDWARE RESOURCES REQUIRED

1. 400MB Hard disk+1GB for Android SDK, emulator system images and caches
2. 2GB RAM minimum,4GB RAM recommended
3. Intel Processor with support for Intel VT-x, Intel EM64T(Intel 64)
4. Computer with windows Vista/7/8/10.

5.2 SOFTWARE RESOURCES REQUIRED

- Platform: Android SDK
- Operating System: Android
- Programming Language: python,R.
- Tools-R Studio
- Packages - keras

5.3 SYSTEM ARCHITECTURE

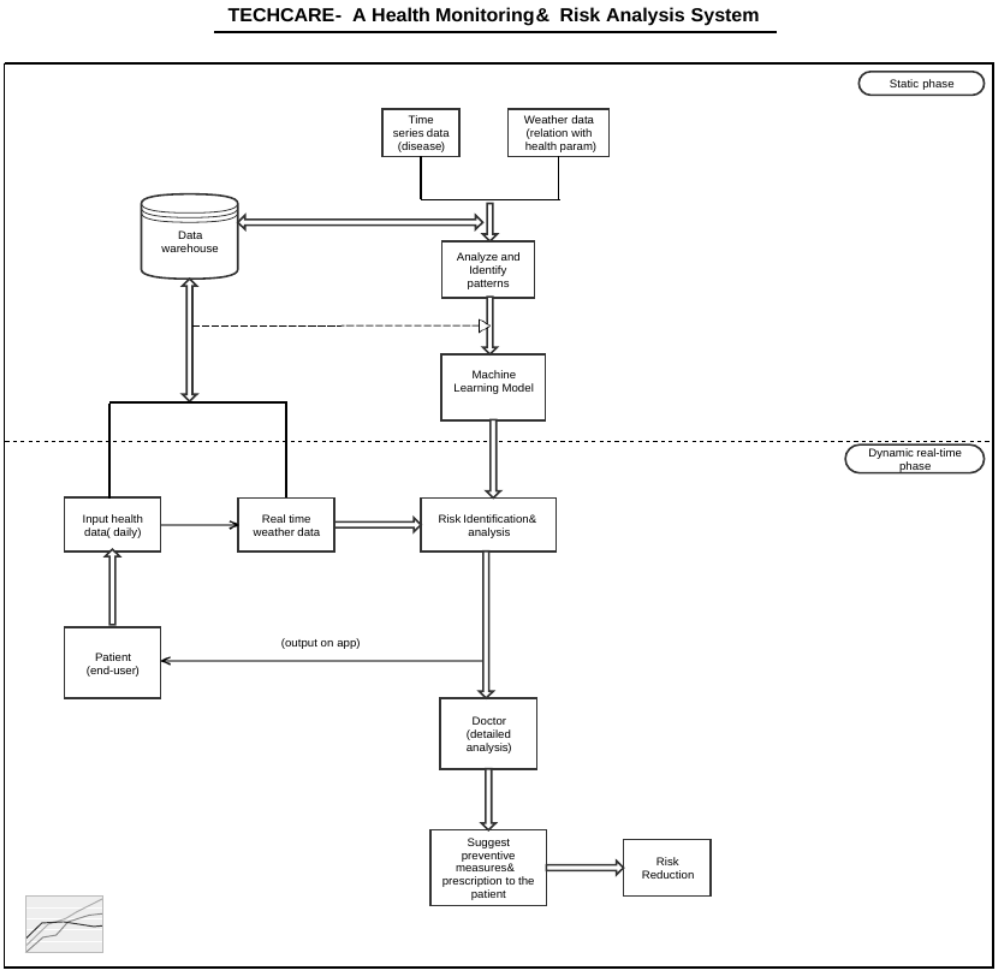

Figure 5.1: System architecture

Chapter 6

Feasibility study

6.1 Technical feasibility

6.1.1 Hardware feasibility

The hardware components are required mainly to increase the speed of training the neural network and test time implementation. GPU has also been used to train the same model and the same dataset which gives an extremely high performance of training and opens a window for datasets having large number of training examples or even those datasets having larger number of features.

6.1.2 Software feasibility

The software components used in this project are open source libraries and platforms. Python is the primary programming language used to implement the project. In order to build neural network models keras open source library has been used which provides a simpler way to code the model along with open source library for machine learning which is Tensorflow as a backend for keras implementation. Also R studio has been used to analyse the data.

6.2 Economical feasibility

The software and the hardware part of the project implementation doesn't require any cost so far. Due to use of free and open source libraries for deep learning, optimization and preprocessing for the software part which doesn't incur any cost for training the model.

6.3 Schedule feasibility

The solution would be built in an estimated time of 7-8 months.The model building,training and implementation would be divided into sequential phases as each component in the project is functionally dependent on the previous one. Time constraints

have been imposed for every phase and especially in the phase of increasing accuracy and fine tuning the parameters of the model and making it more robust.

6.4 Operational feasibility

The system would be used for patients to self test their health and get predictions of diseases based on the changing weather.

Chapter 7

Design

7.1 UML modeling

The Unified Modelling Language (UML) is a general-purpose, developmental, modelling language in the field of software engineering, that is intended to provide a standard way to visualize the design of a system.UML is a standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems. UML was created by Object Management Group and UML 1.0 specification draft was pro- posed to the OMG in January 1997.

7.1.1 Goals of UML:

- Provide users with a ready-to-use, expressive visual modeling language so they can develop and exchange meaningful models.
- Provide extensibility and specialization mechanisms to extend the core concepts.
- Be independent of particular programming languages and development processes.Integrate best practices
- Provide a formal basis for understanding the modelling language.
- Encourage the growth of the OO tools market.

7.1.2 UML diagrams for the project:

Use Case diagram

Use case diagrams are usually referred to as behaviour diagrams used to describe a set of actions (use cases) that some system or systems (subject) should or can perform in collaboration with one or more external users of the system (actors).

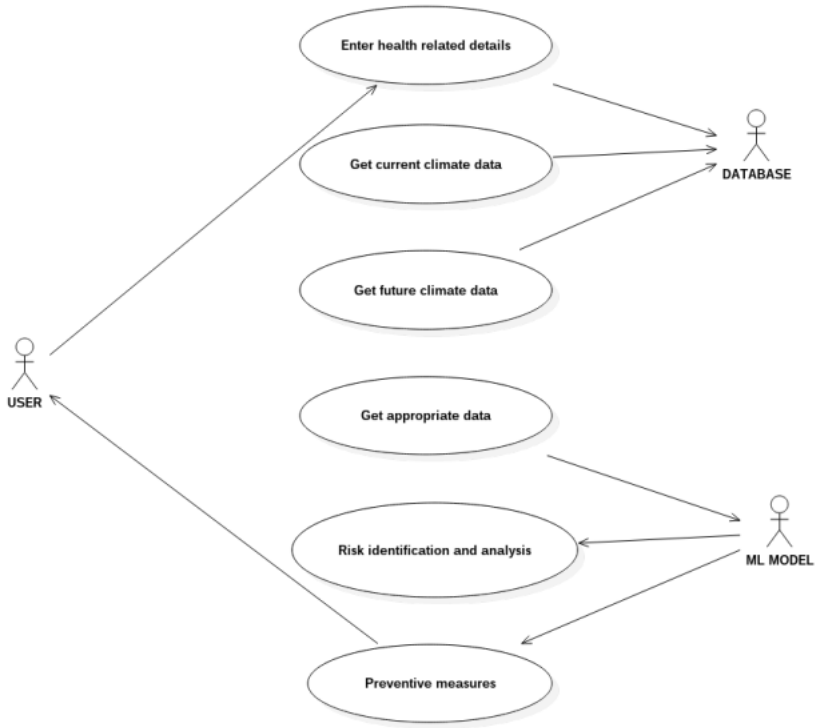

Figure 7.1: Use Case diagram

Activity diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams are intended to model both computational and organizational processes (i.e. workflows).

Figure 7.2: Activity Diagram

Component diagram

Component diagram is a special kind of diagram in UML. The purpose is also different from all other diagrams discussed so far. It does not describe the functionality of the system but it describes the components used to make those functionalities.

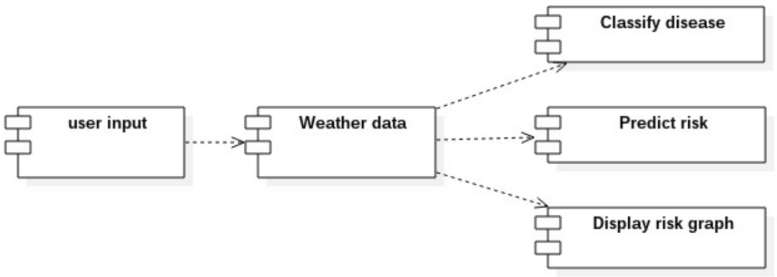

Figure 7.3: Component Diagram

Deployment diagram

Deployment diagram is a structure diagram which shows architecture of the system as deployment (distribution) of software artifacts to deployment targets. Artifacts represent concrete elements in the physical world that are the result of a development process.

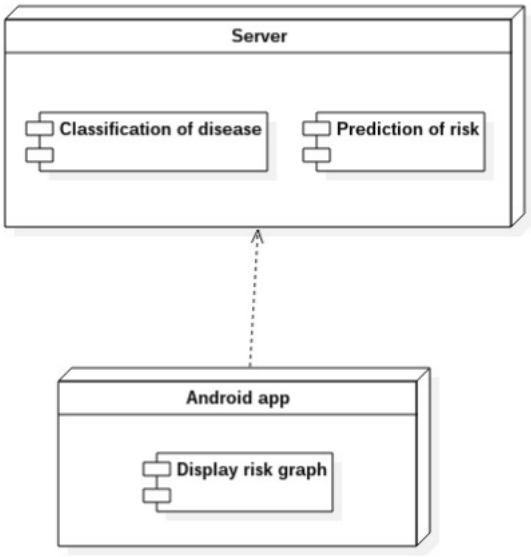

Figure 7.4: Deployment Diagram

Sequence diagram

A sequence diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart. A sequence diagram shows object interactions arranged in time sequence.

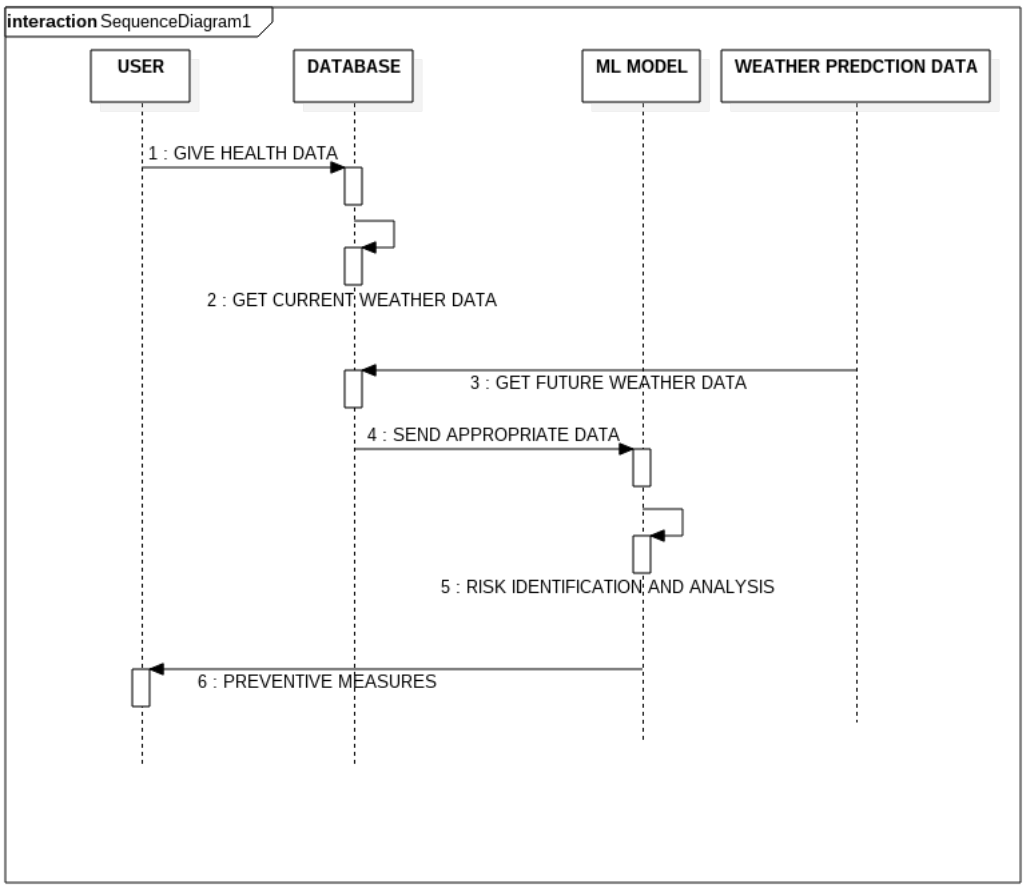

Figure 7.5: Sequence Diagram

Interaction diagram

Interaction diagrams are models that describe how a group of objects collaborate in some behavior - typically a single use-case. The diagrams show a number of example objects and the messages that are passed between these objects within the use-case.

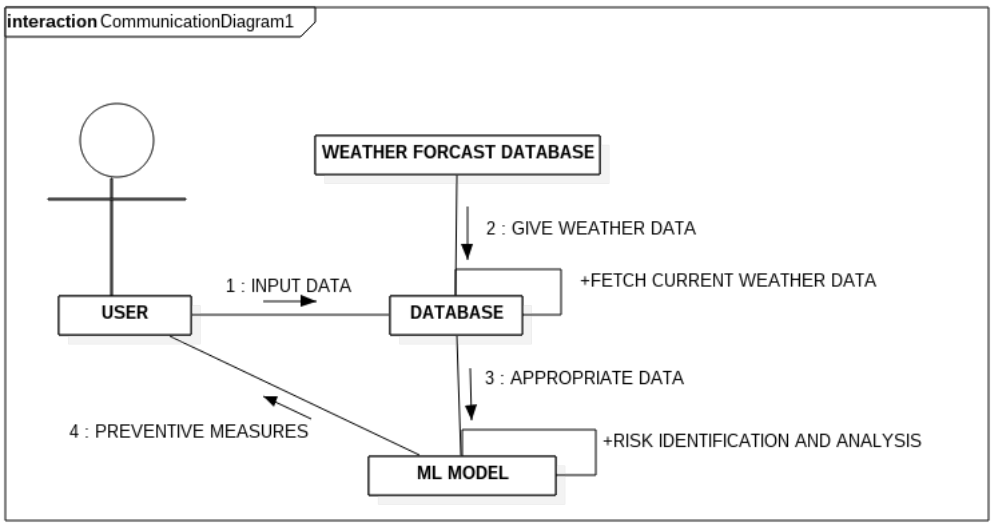

Figure 7.6: Interaction Diagram

Statechart diagram

A state diagram, also called a state machine diagram or statechart diagram, is an illustration of the states an object can attain as well as the transitions between those states in the Unified Modeling Language (UML).

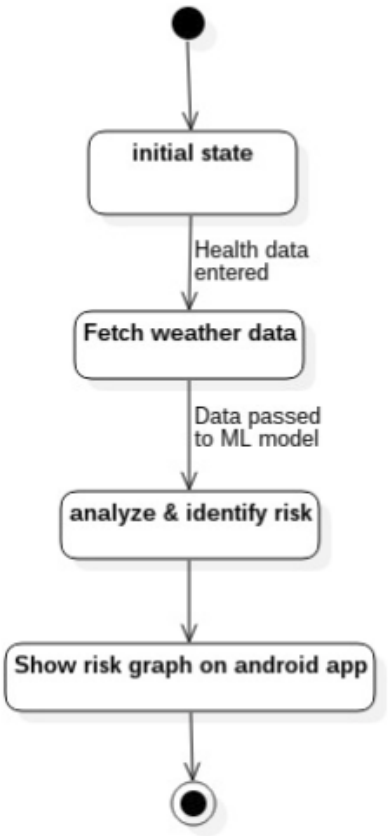

Figure 7.7: Statechart Diagram