UNIVERSITY PARTNER

UNIVERSITY OF WOLVERHAMPTON

HERALD COLLEGE KATHMANDU

Project and Professionalism

(6CS020)

A2: Project Report

Diabetes Detection System

Student Id :1928579

Student Name : Shreejan Shrestha

Group : C3G1

Supervisor : Mr. Hemanga Gautam

Cohort : 3

Submitted on : 13/06/2020

Word Count : 10627 (excluding TOC, abstract, acknowledgement, Table of figure, references)

Declaration Sheet

Award Title: ***BSc (Hons) Computer Science***

Declaration Sheet

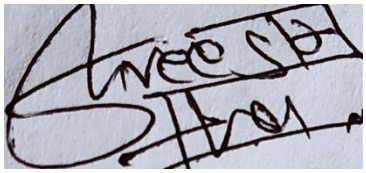
(Presented in partial fulfillment of the assessment requirements for the above award.)

This work or any part thereof has not previously been presented in any form to the University or to any other institutional body whether for assessment or for other purposes. Save for any express acknowledgements, references and/or bibliographies cited in the work. I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

It is acknowledged that the author of any project work shall own the copyright. However, by submitting such copyright work for assessment, the author grants to the University a perpetual royalty-free license to do all or any of those things referred to in section 16(I) of the Copyright Designs and Patents Act 1988. (viz. to copy work; to issue copies to the public; to perform or show or play the work in public; to broadcast the work or to make an adaptation of the work).

Student Name: **Shreejan Shrestha**

Student ID Number: **1928579**

Signature:  Date: 13/06/2020

Acknowledgement

Life is a race. Those who have determination will succeed. It is always delightful to remind the fine people of the Herald College Kathmandu for giving me the encouragement and guidance throughout the year to uphold my skills for doing final year project.

First of all, love and thanks to my parents for everything. Things that I have accomplished till now and will accomplish, there will always be the name of my both parents. My father Mr. Shree Ram Shrestha and my mother Mrs. Sajana Shrestha.

Second, with my all due respect, I would like to thank and appreciate my supervisor Mr. Hemanga Gautam and my reader Ms. Kritika Tuladhar for their continuous feedbacks and precious guidance throughout the year. Without the support of my supervisor, I would never be able to complete this report.

Third, I am very thankful that I got a life changing opportunity to study under the direction of Info Developers Team, especial thank to Mr. Subodh Shakya.

Last not the least, I am also thankful to all the websites such as research gate, google scholar, IEEE for providing all the research papers and journals that ultimately helped me to do all the necessary research and complete the report.

Abstract

Diabetes Detection: Know about your diabetes is the web application with major two purpose. First, to make the general public aware about the diabetes and teach everyone to live a healthy life and second, to provide the free platform to check the diabetes based on the individual's medical report. This web-application will be effective as diabetes is one of the dangerous and commonly found disease in many people round the world. This disease is one of the major causes for other severe illness and disease inside a person's body and can even take a person towards death. The whole motivation to build the project and write the report is to aware the people and introduce artificial intelligence as a new advanced technology in the field of web- application and to build a bridge between the people and technology.

Four different algorithms such as Decision Tree, Naïve Bayes, Linear Regression and Neural Network were tested to fit in the web-application. As the final accuracy of Neural Network with 84% accuracy was the highest among all the algorithms, it has been integrated in the web-application.

Keywords: classification algorithms, diabetes diseases, Decision Tree, Naïve Bayes, Linear Regression and Neural Network.

Table of Contents

Introduction 1
Historical Prospective of AI in health and medical: 1
Classification approach in Machine Learning: 1
Classification system and disease prediction. 2
Research methods used: 3
Overview of Diabetes: 3
Project Academic Questions: 5
Project Background: 5
Details of artifact produced: 5
Problem Statement: 6
Aims and Objectives: 6
Aims: 6
Objectives: 6
Scope and limitations: 7
Scope 7
Limitations 7
Structure of report 8
Literature Review 9
Growth of AI System 9
Increasing use of AI in medical and health sectors 9
Availability of Doctors and Chatbot. 10
Problem of Data in AI 12
Similar Systems 12
2.5.1 Knowledge Based System for Diabetes Diagnosis Using SL5 Object: 12
2.5.2 Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining. 13
2.5.3 Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists: 14
2.5.4 Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing: 15
2.5.5 Decision Support in Heart Disease Prediction System using Naive Bayes: 16
Comparative Result 17
Decision Support in Heart Disease Prediction System using Naive Bayes 17

Analysis 18
Findings: 19
Classification Algorithm 20
Neural Network 20
Main Content 23
Fact Finding Technique 23
Implementation of methodology 26
Tabular Work Break Down Structure 28
Tree-form Work-breakdown structure 29
Analysis 29
Design 32
Implementation 40
Testing 57
White-box Testing 57
Black-box Testing 60
Academic Question 61
Answering the Academic Questions 61
Conclusion 62
Critical Evaluation 63
Findings 63
Evaluation of final system: 63
Limitation of the system: 64
Future escalation: 64
References 65

List of tables

Table 1: Comparative result of similar system 17
Table 2: Tabular representation of work-break down structure 28
Table 3: Functional Requirement 30
Table 4: Non-functional Requirement 31
Table 5: Data Dictionary 43

List of images

Figure 1: Supervised Learning classification techniques 2
Figure 2 Patients waiting for doctors in Kanti Hospital (Poudel, 2019) 11
Figure 3: Expert System's Structure 13
Figure 4 The resulting Decision Tree (Habibi, et al., 2015) 14
Figure 5: Application of Naive Bayes with Jelinek smoothing. 15
Figure 6: Purposed system implementing Naive Bayes in user's data 16
Figure 7 Amount of data and performance of algorithm (Aggarwa, 2018) 19
Figure 8 Layers of Artificial Neural Network (M.Mijwel, 2018) 20
Figure 9 Architecture of two types of perceptron 21
Figure 10: Description of google form 23
Figure 11: probable percentage of patients 24
Figure 12: Update on individual's sugar level 24
Figure 13: public opinion on probably factor for diabetes 25
Figure 14: Rating for knowledge regarding diabetes 25
Figure 15: Checking the consciousness for diet 26
Figure 16 Water Fall Model for the project 27
Figure 18 Tree diagram of Work-breakdown structure 29
Figure 19 Use-case Diagram 32
Figure 20 Context DFD 33
Figure 21 Level 1 DFD 33
Figure 22: Sequence Diagram of the system 34
Figure 23 Whole System Architecture 35
Figure 24: Wireframe of Home page of a web-app 36
Figure 25: Wireframe of Diabetes and Symptom page 37
Figure 26: Wireframe of Diabetes Diet page 38
Figure 27: Wireframe of Diabetes Tester page 39
Figure 28: Wireframe of Tester form page 40
Figure 29 Importing data into notebook 41
Figure 30 Analyzing each attribute 42
Figure 31: Application of Logistic Regression 43
Figure 32: Application of Decision Tree 44
Figure 33: Naive Bayes Classifier 45
Figure 34 Architecture of neural network 46

Figure 35 ReLu activation function 47
Figure 36 Keras model. compile with optimization algorithm, loss function and metrics. 48
Figure 37 Setting epoch, batch size and dividing the data in validation split 48
Figure 38 Displaying the model final accuracy in unseen data 49
Figure 39 Confusion matrix 49
Figure 40 Architecture of Django Framework 51
Figure 41 Home Page 52
Figure 42 Diabetes Symptom Page 53
Figure 43 Diabetes Diet and Exercise Page 55
Figure 44 Check Diabetes Page 55
Figure 45 AI tester Page 56
Figure 46: Form Validation, default 0 value and default message in result display section 60
Figure 47: Filling form with correct value, displaying the value for confirmation and model's prediction 60

Introduction

Historical Prospective of AI in health and medical:

Application of computing and Artificial Intelligence in the field of health and medicine is one of the greatest achievements of human civilization. History of use of computers in the field of health and medicine has been started from 1950s (Ramesh, et al., 2004). But only from 1970s with the help of expert systems such as INTERNIST-I, MYCIN, ONCOSIN, CASNET, we can see the first application of Artificial Intelligence in health and medicine (Ishak & Siraj, 2008).

The application of artificial intelligence in health and medicine was limited to US till 1980. On 13-14 September 1985, with the motive to establish an active research community, an international conference was held in Italy. During those time, major problem to sustain artificial intelligence in the sector of health and medicine was lack of availability of data. To eliminate this problem the concept of Electronic Health Records, also known as EHR, was introduced in 1960s by Lockheed. Later, US government started using EMR after 10 years, in 70s with the Department of Veteran Affairs (Atherton, 2011).

Classification approach in Machine Learning:

Classification and Regression is the two major approaches of machine learning in supervised learning. In classification, model learns with the data which is fit into it and base on its learning from the training data it becomes able to classify new data. There are two types of classification approaches. First one is binary classification and second is multi-class classification. In binary classification, classification is made with just two output classes. For example, whether the person is suffered from diabetes or not, either a person is male or female or either the mail is spam or not. Where as in multi-class classification, the model classifies the output with more than two classes (minimum three). For example, classifying cat, dog and mouse from a group of animals, multiple disease classification, documents classification, etc. (Soofi & Awan, 2017). Classification approach used in this project is binary classification. Model will be trained with multiple algorithms like Logistic regression, Random forest, Neural Network, etc. Out of all

classifiers, one with best performance will be used in the system. Model will classify whether the patients have diabetes or not and will give the output on the basis of 1 and 0 based on the medical report given to the model as input.

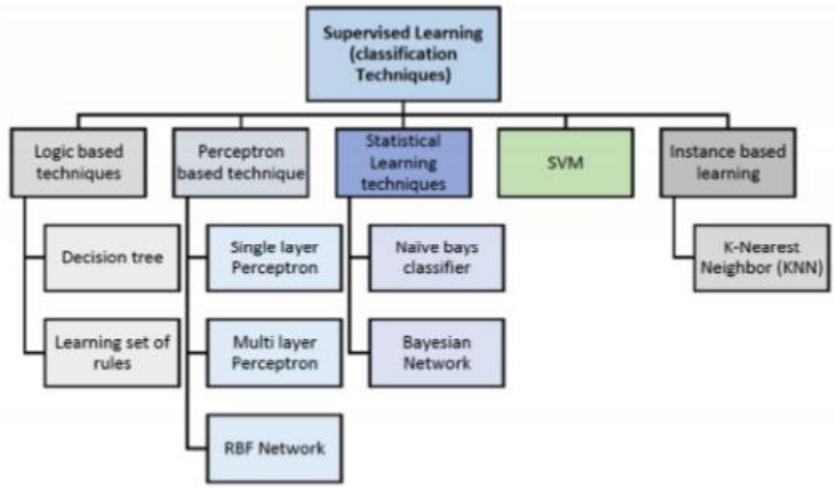


Figure 1: Supervised Learning classification techniques

Classification system and disease prediction.

In every country, many hospitals and clinics has been established either by the government or by private companies with the mind set of providing health related facilities such as counseling related to health and medicine, normal check-up to big surgeries. The major challenge for all those health service providers is providing high quality of service with minimum cost with zero error. To obtain desired results and server whole nation, implementation of artificial intelligence-based computer systems can be done, which can help in decision making process that also in minimum time period. Development of system which can detect or classify any particular disease inside a human body can be one of the greatest achievements of human civilization. Any system that makes a prediction or detection of a disease using classification algorithms of machine learning can be called as disease detection or classification system. To make a prediction or detection of various diseases, many researches has been carried out with the help of artificial intelligence and machine learning. This project, “Diabetes Detection, Know Your Diabetes”, is also one of the examples of classification system which classifies disease (C.Dharuman, et al., 2017).

Research methods used:

1. Journals and articles
2. Reports
3. Conferences papers
4. Newspapers and online news
5. Documents from various web
6. E-Books

Overview of Diabetes:

Diabetes is one of the dangerous and commonly found disease in many people round the world. Main cause of diabetes is high blood sugar level. This disease is one of the major causes for other severe illness and disease inside a person's body and can even take a person towards death. If diabetes is not cared and control, it can also cause damage to other parts of the body like kidney, heart, eye, nerves, etc.

Diabetes is one of the kinds of metabolic disease categorized by hyperglycemia. Common symptom of diabetes can be seen from following cases like weight loss, urine problem, abnormally increase in thrust, excessive appetite and problem in vision. Longstanding problems of diabetes includes Retinopathy, Nephropathy, Peripheral neuropathy, Autonomic neuropathy.

1. Retinopathy:

Retinopathy is the disease related to retina which cases serious problem to the vision of eyes.

2. Nephropathy:

Nephropathy is the disease related to kidney. Those people who has diabetes since longer period has the higher chance to suffer from this disease.

3. Peripheral neuropathy:

Peripheral neuropathy is a disease which affects in the peripheral nerves of human body. This disease can affect only single nerve to different nerves of the body at the same time.

4. Autonomic neuropathy:

When there is damage to the nerves which various day to day information to the spinal cord and brain, we can say, person autonomic neuropathy (American Diabetes Association, 2014).

There are various types of diabetes. Some of the most common types of diabetes are as follows:

1. Type 1 diabetes:

In type 1 diabetes, body does not create any insulin. In the case of type 1 diabetes, human body's immune system attacks the pancreas gland which is responsible for making insulin. Because of the absence of the insulin in the body, the level of blood rises up than normal and causes diabetes. If the person suffers from type 1 diabetes, he/she must take insulin in daily basis in order to maintain the sugar level and stay healthy.
Most of the victims of this diabetes are children and younger people, and less chances for the old age people.

2. Type 2 diabetes:

Type 2 diabetes is found in most of the diabetes patient. This disease is caused when the blood sugar level is very high than normal. When the human body does not create enough insulin or lose the ability to utilize the created insulin then the person suffers from type 2 diabetes. Unlike type 1 diabetes, type 2 diabetes can be cared and stop the development of the disease, if proper care is given in the early stages. Mostly old age people aged from 45 or above suffers from type 2 diabetes. Major causes of type 2 diabetes are overweight, heredity and unhealthy life style.

3. Gestational Diabetes:

This kind of diabetes is only found in female. If the person is pregnant, she is most likely to have gestational diabetes. Symptoms of gestational diabetes is only seen after 24^{th} week of the pregnancy. This diabetes id very dangerous even to infant that is why a mother should carefully look after her health and stay updated of you blood sugar level (Rodgers, 2017).

Project Academic Questions:

- Can diabetes be accurately predicted using Machine Learning and Deep Learning model?
- How the doctors will get benefit by this model?

Project Background:

Many people have died till date because of misclassification of the disease. Considering this situation, if a system became able to help the doctors to detect or predict a disease in accurate way, on the basis of the given symptoms, then this can be the ultimate solution of this problem. With the innovation of high-performance computing, availability of data and advancement in the algorithms, artificial intelligence are in good practice now a days in the field of disease diagnosis. “Diabetes Detection: Know about your diabetes”, will be the web-based application, which will have diabetes detection model, that helps people to check and know if they are suffering from diabetes or not. Along with the detection model, it will have basic but authentic information regarding diabetes, its symptom as well as things to eat and guidelines to follow for healthy lifestyle

Details of artifact produced:

On daily basis, we hear the news of various cases of fraud happening in many hospitals, death of the patients because of the misguidance or improper treatment from the doctors and charge of big amount of fee even for a normal follow-up. To solve these problems in case of diabetes, “Diabetes Detection: Know about your diabetes”, has been built and will help people who wants to check their diabetes’s status staying in home and know the result in advance before going to the doctors for further checkup and treatment. To build the final model for the web-application, different machine learning algorithms such as random forests, decision_tree and neural network were built and tested. As neural network gave the best accuracy among all algorithms, it has been finalized to fit into the web application. With the help of Artificial Neural Network (ANN), “Diabetes Detection: Know about your diabetes”, will detect whether the patient have a diabetes or not based on their medical report. To check the diabetes, user must have the report of their insulin level, glucose level, blood pressure, skin thickness, body mass index (BMI) and age. For female, total number of times she has become pregnant. After providing

the data, ANN will detect and show the result on the basic of 0 and 1 which is “Yes” and “No”. An Artificial Neural Network (ANN) is a data processing paradigm which is inspired with the biological nervous systems i.e. human brain working mechanism. ANN is the combination of large numbers of highly connected neurons, working in unison to solve a specific problem. In general, Artificial Neural Network are in massive practice round the globe in many sectors because of their ability to build nonlinear relationship between input variable and output variables directly from training data.

Problem Statement:

Because of unhealthy life style the number of diabetes patients are increasing rapidly in every part of the world. Even though people brag of being modernize and literate, most of people are lacking behind in term of proper guidance to prevent from diabetes. Frauding from every sector like hospitals are also increasing and takes massive fees. Because of carelessness of medical personnel human casualties are also increasing.

Aims and Objectives:

Goal of this project is to detect the diabetes, provide useful data about diabetes and study about different machine learning algorithms such as Linear Regression, Random Forest and Artificial Neural Network to find out which one is the best algorithm that can be used accurately to detect diabetes.

Aims:

- To build a web-application and prediction system which will detect whether a person is suffering from diabetes or not.
- To provide information about causes and consequences of diabetes.
- To explain the precautions to prevent from diabetes and advices and suggestions to control diabetes / sugar level if in case a person is already suffering from diabetes. People
- Aware general to stay fit and healthy.

Objectives:

- To prevent people from getting scammed by the doctors or medical staffs

- To help people to check about their diabetes/ sugar level easily with the help of simple web application.
- To provide the tips regarding physical exercises or activities to control blood sugar level.
- To provide information about hygienic food.

Scope and limitations:

Scope

Diabetes Detection System can be deployed for various purposes. Such as:

- It will be very useful to general public. Everyone can test their diabetes status at very 1st level staying at home.
- It will help medical staffs, doctors and contributes a lot in medical field.
- Doctors will also get help and one level of alert if patients have first tested with diabetes detection system.
- Time and money will be saved.

Limitations

However, systems like these also have their own major drawbacks. Such as:

- Not only this all artificial intelligence integrated system will just predict the result but will not point how and which major factor is responsible for the particular.
- It cannot be used as a final and only measure to check diabetes.

Structure of report

1. Introduction

Introduction part explains all the details from the surface level such as how the system will be built from front-end as well as backend, which algorithm will be used. This section will further explain about the problem domain and will clearly explain, how after the development of the system, the problem will get minimized, and finally, how this system will work as the best alternative as well as helping hand for medical department.

2. Literature Review

Literature Review is the core part of the report. This section focuses on history of AI in the field of medical field. This section will major focus on providing theoretical basis of the project by explaining about different algorithms which can be used in this system and analyzing similar systems.

3. Implementation of Prediction System

Execution of Detection System section will focus on the AI part of the project. This chapter will analyze the dataset that will be used to train the model and how the system will make prediction on the basic of the dataset. This section will focus on every single detail that has been implemented to build a system. It will also explain all the developmental stages of detection system in increments.

4. Implementation of Web-Application

Execution of Web-Application will shop the step by step execution of web-application. Explanation about the framework used and all the pages which has been integrated in the web-application.

5. Testing

Testing of the whole system will be done. White-box testing will be carries out. Lower level testing like unit testing and integration testing will also be done in other to check the internal implementation of the system from unit by unit.

6. Conclusion

Concluding all the project and listing the further improvements that can be done to increase the benefit from the project.

7. Critical Evaluation

This section will be the critical evaluation of the final project, evaluation of the final system, findings and future escalation.

Literature Review

In this part of the project, various health related problems, and current advancements and achievements of AI for eliminating those problems has been discussed. After that, introduction to machine learning (supervised learning) and brief description of different classification algorithms has been done, that explains the technical basis of the project. At last, similar systems have been compared and analyzed. To know the status of diabetes patients (if any) and to check the general knowledge regarding diabetes, and diabetes condition of Nepal, one online based survey was conducted among random people.

Growth of AI System

Artificial Intelligence is on the biggest hype round the world. Every AI system have the great ability to adapt the situations and make decisions according to the situations. One of the biggest reasons for this tremendous growth of AI systems if the scope and opportunity they provide. It has estimated has AI technologies has the potentiality to uplift the global GDP by $15.7 trillion by 2030. Among all country, China is making the speedy process because of its large amount of investment in AI and development, around $150 billion and said to be the global AI leader by 2030. (West & Allen, 2018)

Increasing use of AI in medical and health sectors

Artificial Intelligence is making its way in almost every sector in whole world for example business, medical, national security, tracking, building smart cities, advertisement, climate and ecosystem behavior analyze, transportation and in all other basic needs of human being. Except these, there are other numerous fields where AI

has shown its impact and copying human behavior and ability is significant ways. For example, Merantix is a German company which has built a system by implementing deep learning in medical imaging which detects lymph nodes inside the human body in Computer Tomography (CT) images (Rothe, 2017). Similarly, Vara has developed an AI powered system for screening the breast cancer and help the radiologists by reducing the repetitive work load (Vara by Merantix Healthcare, n.d.) The above two mentioned AI integrated system comes under the domain of computer vision and image processing. There are many other sub-fields inside AI such as Natural Language Processing, Game Playing, Scientific Analysis, etc. which are also in the field of research and producing many remarkable AI products that helps to man-kind. With the help of Artificial Intelligence and deep learning, Quotient Health has developed a software targeting to decrease the cost of supporting EMR [Electronic Medical Report] systems. This software can optimize and standardize the traditional design of EMRs systems. Their main objective is to lower the cost and improve the health care (WelcomeAI, 2019). Ciox Health is a healthcare information management company that uses machine learning to develop and improve health information management and enhance the exchange of data and information related to health.

Their main aim is to modernize the workflow, providing the facility of access to clinical data from all over the world and improving in the accuracy of exchange of health related information and related data (Ciox, 2014). At the university of Chicago, mammograms of about 22,000 women were taken to implement automated screening of mammograms, which is also known as computer-aided diagnosis (CAD). After the test, system identified 52% that has a sing of breast cancers a year before they were diagnosed as a breast cancer (Ganesan, et al., 2013).

Availability of Doctors and Chatbot.

Artificial Intelligence is not only limited to health care. Other than just diagnosing the different diseases, detecting or predicting diseases, artificial intelligence can also be effectively used for optimizing the clinical processes. According to WHO, about 57 countries are facing problem of lack of health assistants such as physicians, doctors, nurses and midwives. These problems are directly linked with poor health status and health care facilities that many countries are facing. According to the report produced by

WHO, there is shortage of round seven million of doctors and other health personals worldwide. (Scheffler, et al., 2008). If we watch the data of India, as the population is increasing day by day and year by year, they are facing the problem of availability of doctors. There is shortage of projected 6,00,000 doctors and 2 million nurses in India. In India, there is only one doctor for every 10,189 people and situation will get worse as the population has been increasing (The Economic Times, 2019). Whereas if we look the data of Nepal, even Nepal is also facing similar kind of problem. Comparing the total number of available doctors to the population density of the nation, the density is 7 doctors per 10,000 population (World HEALTH ORGANIZATION, 2020).

Figure 2 Patients waiting for doctors in Kanti Hospital (Poudel, 2019)

AI health assistance can be of great use. They can be used for helping the doctors and clinical staffs, saving their precious time and making doctors and nurses ease to stay focused in other critical cases. Similar AI health assistance Chatbot can be found in the form of “MANDY”. Mandy was developed to assist the healthcare staffs. This chatbot makes an interaction with the patients and record their complaints in the form of natural language and lastly submit the report to the doctors for detail investigation and analysis

(Ni, et al., 2017). Chatbot like this and similar other use Deep learning and its sub-field Natural Language Processing to interact with the patients and doctors and generate basic suggestion and recommendations.

Problem of Data in AI

To build a system and to train the system's model, which will later detect or predict a disease based on truth and fact, needs adequate amount of data. Almost all health related, AI integrated system face the problem of availability of data. Because of privacy issues, data are not made available for the research and development purposes. Due to the reason of lack of open data source, development of these systems gets difficult. In context of Nepal, till now, many sources of data have not made digitalized. It still follows traditional methods like preparing a report in paper. Though there is vast amount of data generation every day related to medical field, they are not made open to researchers, students to make use of it and build insights from it. Because of scarcity of data, it makes difficult to apply machine learning algorithms, and formulate concepts and ideas from data in order to build decision support tools.

Similar Systems

Various algorithms, use of expert systems and various methods have been discovered and implemented for classifying and detecting various kinds of diseases. Not only for diabetes, implementation of AI can be found for diagnosing various other diseases. Some of the similar systems in the domain of health and medical have been discussed below.

2.5.1 Knowledge Based System for Diabetes Diagnosis Using SL5 Object:

Expert system is a computer programming system which uses artificial intelligence tools, technique and technologies to make a replication of behavior of humans or any organization who is expert in particular filed. Knowledge based system is one of the main components of expert system. This is a type of system which contains vital information about the problem domain and often represents as facts and rules (Dheir, et al., 2019). SL5 stands for "Simpler Level 5 Object". SL5 is both rule-based language for specifying expert systems and a rule-based system for performing expert systems which is written in the Sl5 object language (Naser, 2015).

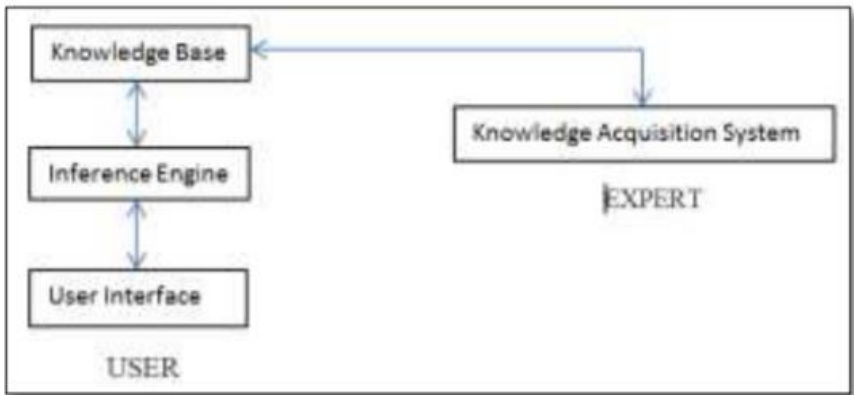


Figure 3: Expert System's Structure

This system detect whether a person have a diabetes or not, on the basis of eleven symptoms of diabetes such as frequent urination, excessive thirst, increased hunger, weight loss, tiredness, lack of interest and concentration, tingling sensation or numbness in the hands or feet, blurred vision, frequent infections, slow-healing wounds and vomiting and stomach pain. This system is based on classification problem and helps to diagnosis diabetes by querying the person, asking yes or no questions regarding the symptoms felt by the patient (Mettleq, et al., 2019).

2.5.2 Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining.

This is a prediction model which uses decision tree and J48 algorithm to predict the type two diabetes. This model has been developed in WEKA software. Decision tree is one of the extremely used and powerful machine learning algorithms, which is widely used in many medical fields for classification problem related with the prediction and diagnosis cases, such as diabetes. To train and test the model, all required data were collected from the database of a web-based health center diabetes control system in Tabriz, Iran. Data set includes the features such as age, sex, systolic and diastolic blood pressure, family history of diabetes, and body mass index (BMI). 22,398 records have been used for data mining process and training the model. As the model was using decision tree, age was positioned on top as a root node of the tree as the information gain of the age factor was higher. For the validation process, the 10-fold Cross Validation method has been used. This article has also recommended "The Pima Indians dataset", to train and test the model. The model using decision tree and J48 algorithm has got

0.78 accuracy in normal dataset where using The Pima Indian dataset, it has got 72.75 accuracy (Habibi, et al., 2015)

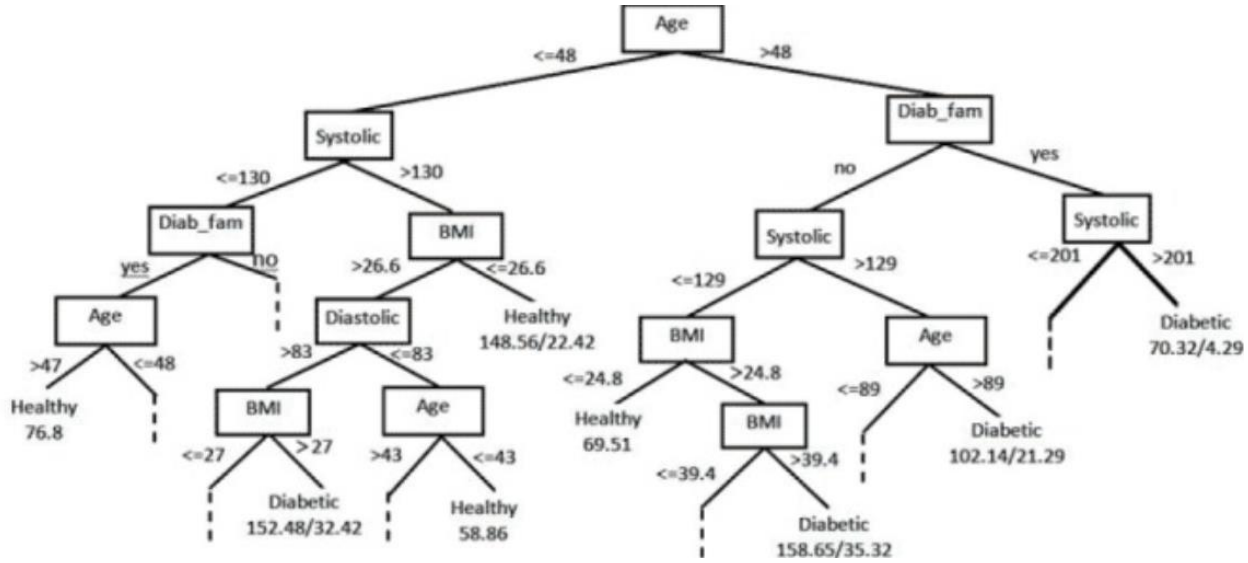


Figure 4 The resulting Decision Tree (Habibi, et al., 2015)

2.5.3 Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists:

Similar system which uses Artificial Intelligence to diagnose a disease is “Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists”. Breast cancer is one of the common but most dangerous cancer found in women. Despite the use of many modern technologies and applying many therapies, it is still the major cause of increasing cancer-related mortality. The main aim of this system is to deploy artificial intelligence in mammography for detecting breast cancer. This system has used convolutional neural network, feature classifiers and image analysis algorithm to spot classification and soft tissue lesions in two different modules. Digital mammography images of many vendors can also be used in the system and make use of both mediolateral oblique and cranio-caudal to view each breast. This system has used 9000 data of mammograms with cancer and 180000 data of mammograms without abnormalities. The AI system is independently tested with the data which has very been used in the process of developing and training the mode. This system has secured 0.80 of ROC curve and 61.4% higher AUC than that of radiologists. (Rodriguez-Ruiz, et al., 2019)

2.5.4 Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing:

Ms. Rupali R. Patil has built a Decision System in Heart Disease Prediction System (HDPS) with the help of Naïve Bayes algorithm. It is one of the famous data mining tool and modeling technique. Data mining is one of the important steps which helps to extract hidden and implicit data, which has never been recognized by the experts. Data mining assists in finding new and more précised pattern in the data which can be used in future to make smarter decisions. This system uses medical report of the patients and predict the chances of suffering from heart disease. It has been implemented in MATLAB and uses age, gander, BP, blood sugar level, pain in the chest report and ECG graph. Based on the trained model with similar cases, model predicts the probability of disease.

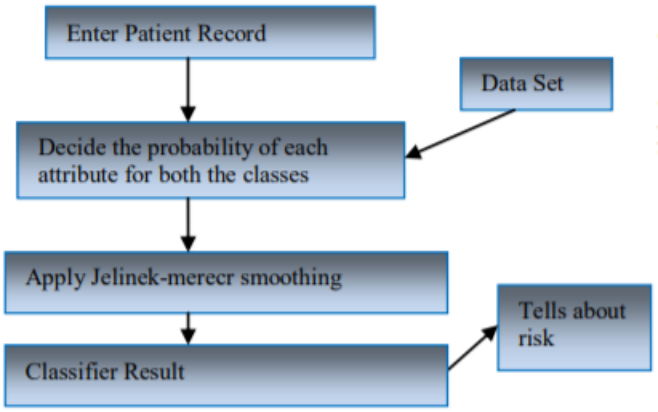


Figure 5: Application of Naive Bayes with Jelinek smoothing.

Naïve Bayes is the fundamental for every machine learning and data mining process. This algorithm seems more effective when the inputs data has higher dimension or lager number of parameters. Because of its ability to perform well with higher dimensions, it can obtain good result than any other big classification algorithms. Jelinek-Mercer smoothing is a tool that can be used to capture significant patterns inside the data, avoiding noises at the same time. This model has got an accuracy of 78% by using Naïve Bayes and has got an accuracy of 86%, 8% increased, by using Naïve Bayes with smoothing technique. Medical students and nurses can be taught and train to diagnose the heart disease patients by using this system (R.Patil, 2014).

2.5.5 Decision Support in Heart Disease Prediction System using Naive Bayes:

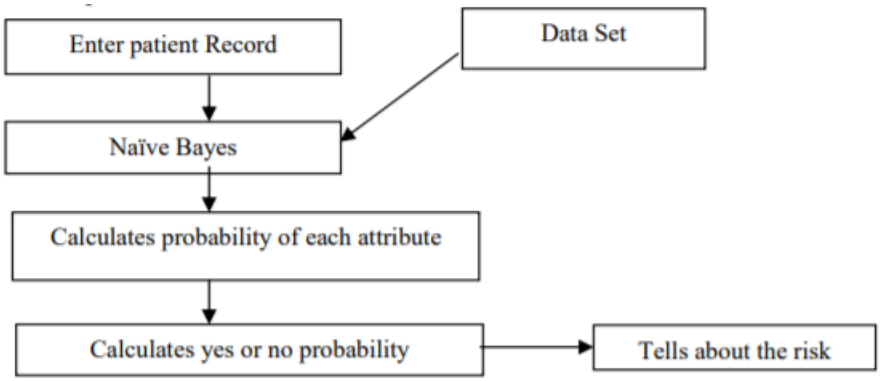


Figure 6: Purposed system implementing Naive Bayes in user's data

Mrs. G. Subbalakshmi, Mr. K. Ramesh, Mr. M. Chinna Rao has also contributed in developing the AI system which will predict the heart disease with the help of Naïve Bayes algorithm. Decision Support in Heart Disease Prediction System is a web-based questionnaire application which also uses similar data mining tools and technique as above-mentioned model but this system is lacked of smoothing technique. Evaluation of performance and accuracy of the model has not been mentioned in the article (G.Subbalakshmi, et al., 2011).

Comparative Result

Table 1. Comparative result of similar system:

S. N.	Name of application	year	Problem/ Disease	Tools and Technique used	Performance/Accuracy
1	Knowledge Based System for Diabetes Diagnosis Using SL5 Object:	2019	Diabetes	Simpler Level 5 Object (SL5) an expert system	-
2	Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining.	2015	Type 2 diabetes melitus	Decision Tree and J48 algorithm	0.78
3	Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists	2019	Breast Cancer	convolutional neural network, feature classifiers and image analysis algorithm	0.80 ROC
4	Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing	2014	Heart Disease	Naive Bayes and Jelinek-mercer smoothing:	75% in Naïve Bayes 86% using smoothing
5	Decision Support in Heart Disease Prediction System using Naive Bayes	2011	Heart Disease	Naïve Bayes	-

Table 1: Comparative result of similar system

Analysis

How is my project different from the above listed systems?

In above mentioned five systems, first two systems are predicting diabetes. In those two systems, one of the systems have used Simpler Level 5 Object (SL5), an expert system and second have used decision tree and J48 algorithm to predict diabetes. Whereas, among remaining three other systems, third system have used convolutional neural network (CNN) to predict breast cancer. Fourth and fifth system have used Naïve Bayes and Jelinek-mercer smoothing and Naïve Bayes respectively. Each system has used different types of classification algorithms. If we focus only in diabetes prediction model, none of two have used Artificial Neural Network to classify the diabetes and one among two has been found using and recommending the same dataset which has been used in this project i.e. The Pima Indian Dataset with different concept and algorithms. Both the authors have used few similar but different dataset and approach to classify the diabetes.

System based on expert system have fit in 11 attributes as the symptoms of diabetes such as frequent urination, excessive thirst, increased hunger, weight loss, tiredness, lack of interest and concentration, tingling sensation or numbness in the hands or feet, blurred vision, frequent infections, slow-healing wounds and vomiting and stomach pain whereas

System using decision tree have used 6 attributes from Tabriz, Iran's data set which includes the features such as age, sex, systolic and diastolic blood pressure, family history of diabetes, and body mass index (BMI). As both the systems are classification base model with high dimensions using ANN would be the better option and have the higher chances for higher accuracy. Neural network has the capability to handle the high dimensionality problems using multiple effective frameworks like keras, tensorflow, etc.

But the major drawback of neural network is, it requires large amount of dataset to train the model on. As technologies are being updated on the daily basis and because advancement of operating systems and processing power, availability of data and stronger algorithms, AI researchers prefers neural networks more than classical algorithms.

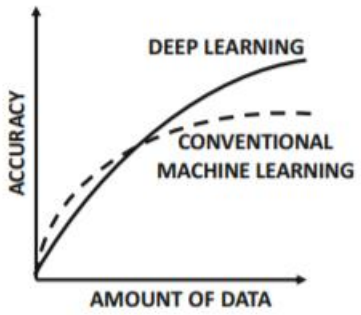


Figure 7 Amount of data and performance of algorithm (Aggarwa, 2018)

With the intention to fill this gap by increasing accuracy and performance, this project has implemented neural network with The Pima Indian Dataset to predict the diabetes with 84% accuracy. The dataset has 8 important attributes on the basis of which first layer of assurance can be taken regarding the status of diabetes inside our body.

In above mentioned systems, many are just theory base and have not been implemented in real world to test and serve real life patients. Only one system has been implemented in web application. This system provides the platform to connect every people with testing tool of diabetes.

Findings:

After analyzing all the related systems and their concept of implementing AI in the field of medical, I have gained few insights:

1. For classification there are many algorithms and every author have used different approach to classify a disease.
2. As medical field is the critical sector where one mistake can cost someone's life, accuracy of the model is the major concern. Instead of using only one algorithm and dataset, using multiple algorithms can help to obtain better accuracy and classify disease.
3. For classifying the diabetes, age and body fat are the major attributes.

Classification Algorithm

Neural Network

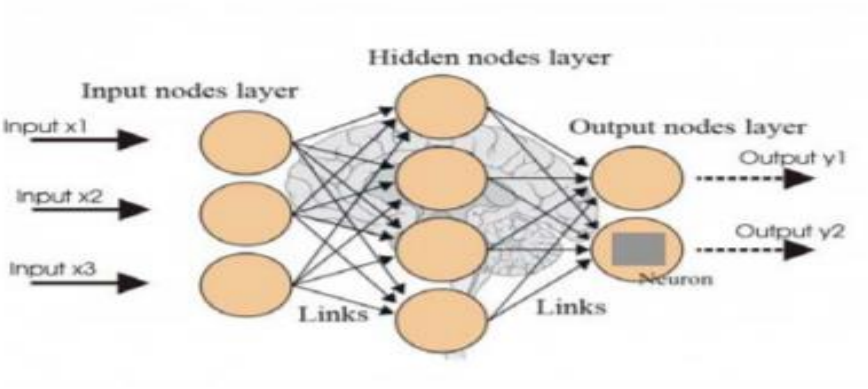


Figure 8 Layers of Artificial Neural Network (M.Mijwel, 2018)

Neural networks have been developed to make a similar system that can replicate the human brain to perform the machine learning tasks. Main aim of neural network is to give birth to the artificial intelligence by developing the superior machines whose architecture simulates the calculations made by human nervous systems. Development of neural networks were started immediately after the arrival of computers in 1950s/60s. Perceptron algorithm developed by Rosenbaltt is the benchmark and was the foundation of neural network which created the interest and directed the new vision among the researchers for artificial intelligence. At the initial period, building the deep learning model was a bit disappointing as the rigorous nature of neural network that needed huge data and computation power was creating the hurdle. But, at the end of the century, availabilities of data and growing power of computers made the development of neural network as expected.

On the basis of supervise machine learning, neural network can be divided into 2 parts they are Classification and Regression. The major difference in these two approaches are the selection of an activation function in final layer and using the loss function as per the problem.

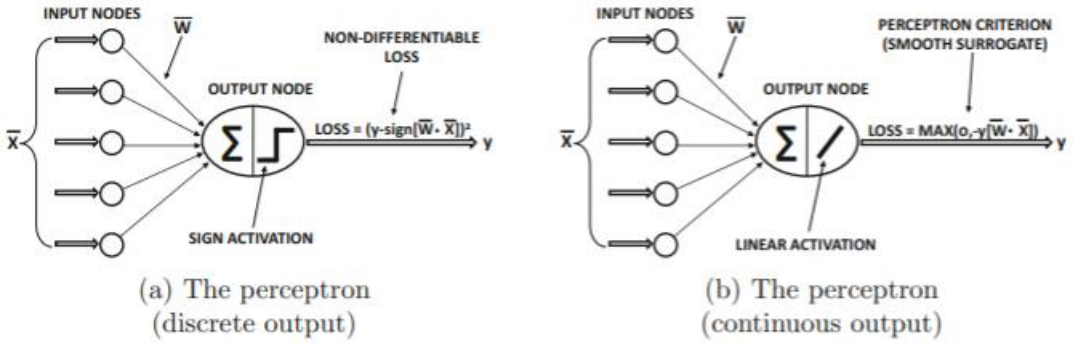


Figure 9 Architecture of two types of perceptron

For classification the output will be on the basic of yes or no, or class base basis whereas regression's output will be continuous value like weather prediction, stock market prediction. Diving deeper into classification, classification has two types:

1. Binary classification: output layer will have at least two classes and result will be on the either 0 or 1.
2. Multi-class classification: output layer will have more than 2 classes and result will be based on the higher probability of a particular class (Aggarwal, 2018).

Advantages of Neural Network

1. Can work with incomplete knowledge:

   After the model is trained, it can predict the output even with the incomplete data and accuracy of the output hinges on the significance of the missing value.

2. Ability of high fault tolerance

   Fraud in one or multiple cells of neural network does not stop it from producing outputs. This ability makes the neural network fault tolerance.

3. Capability of parallel processing

   Because of having numerical strength, neural network has the ability to perform more than one task at a time.

4. Skill to teach machine learning

   ANN has the ability to learn the small events from big set of data.

Disadvantages of Neural Network

1. Dependence in hardware

Neural Network is hardware dependence and processors that can-do parallel processing is required.

2. Inexplicable behavior of the model

Neural network works similar to the working mechanism of human brains. Because of this reason, if the error occurs from inside the hidden layers than it does not provide any clue what is happening inside the network.

3. Training time might get longer

Though Neural Network has the best ability to handle the large amount of data to train the model, depending on the processing power, it can even take hours, day to train the model.

4. Hard to visualize the problems

As the Neural Network works fully on numeric system, the problems are also introduced in the form of numeric. Depending on the experience of the user, this problem will get difficult.

In this project, binary classification base neural network model has been implemented which will predict the diabetes on the basis of 0 or 1.

Main Content

Fact Finding Technique

As one of the fact-finding techniques, google online survey was conducted with fifty-five people of different family to know their knowledge and awareness regarding diabetes and its consequences. Questions like, “Are there anyone in your family suffering from diabetes? “and 5 other high-level questions were asked and found out that, the numbers of diabetes patient (at least one) in each family are more than family that has no diabetes patients. As the survey was targeted to the local society of Nepal, after the survey it was found that, 51% of family do not check their diabetes (sugar level) timely and are not conscious about their diet. All the people do not have proper knowledge regarding the major factors that causes diabetes and most of them think heredity, age and eating too much sweet only causes diabetes. Among the fifty-five people who took a survey was asked personal about what can be the possible factor for a woman to suffer from diabetes and unfortunately, some of woman do not know about Gestational Diabetes, which happens only to female during and after pregnancy. As diabetes is killing many people round the globe, awareness about the preventive and precaution measures must be spread strictly.

Below are the screenshots of the survey that was organized through online for 1 week.

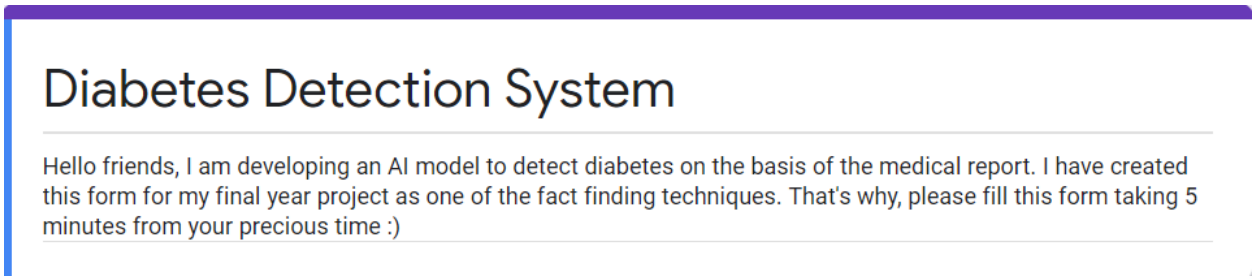


Figure 10: Description of google form

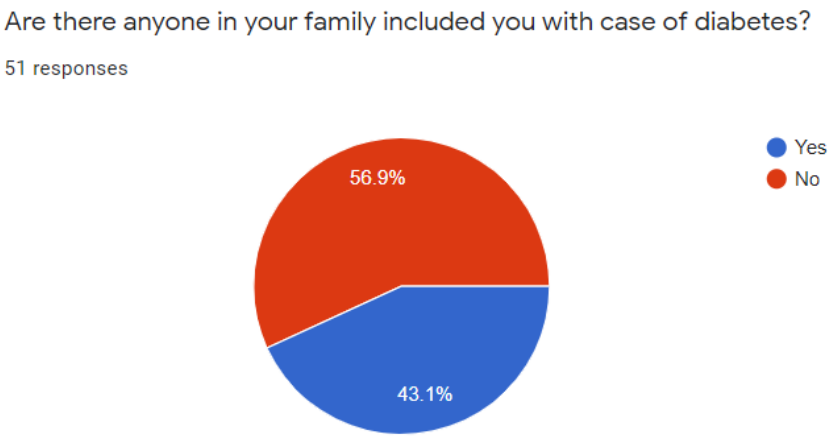


Figure 11: probable percentage of patients

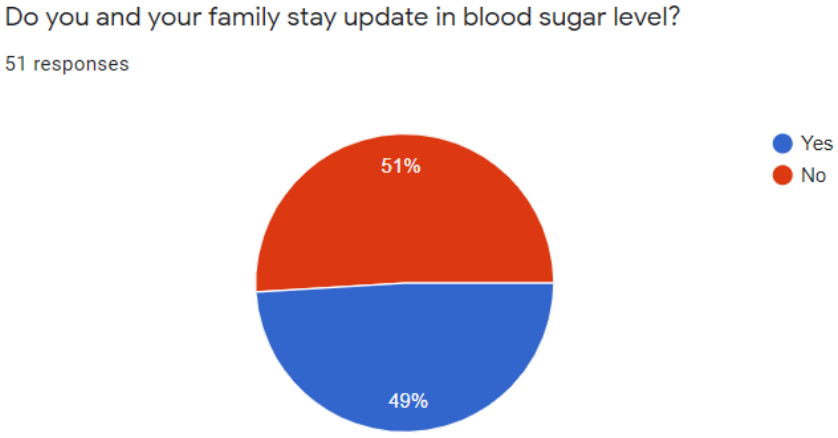


Figure 12: Update on individual's sugar level

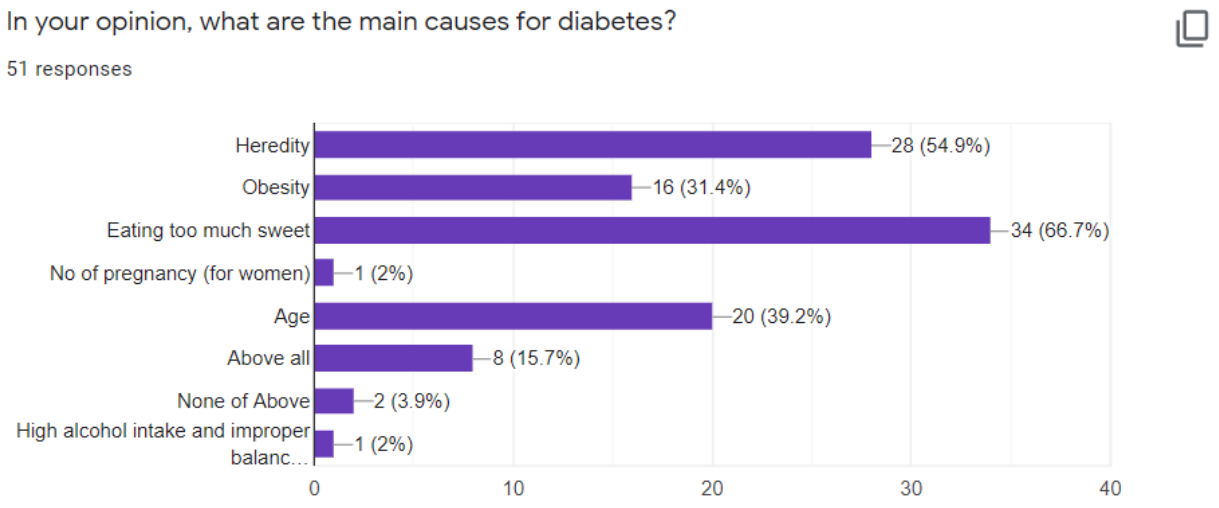


Figure 13: public opinion on probably factor for diabetes

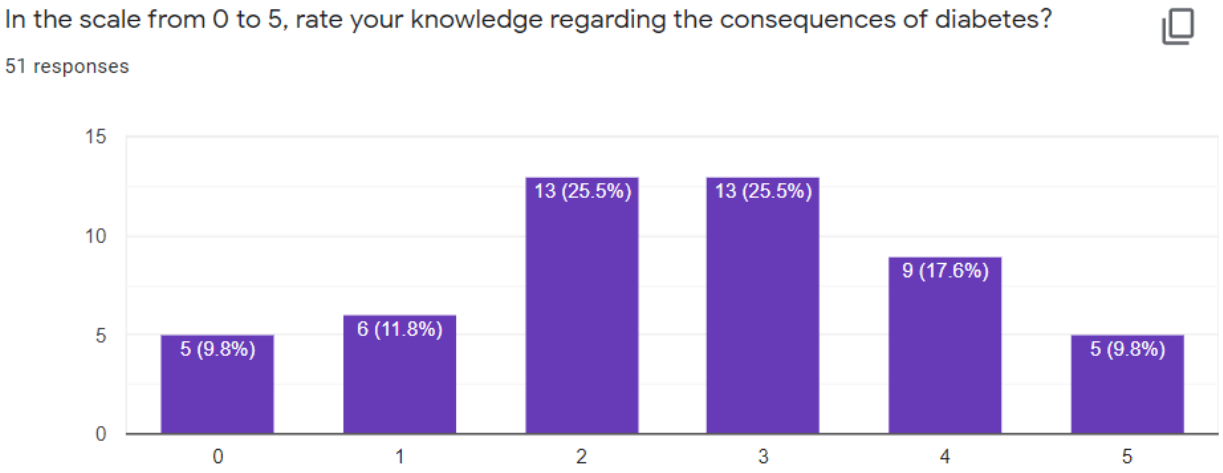


Figure 14: Rating for knowledge regarding diabetes

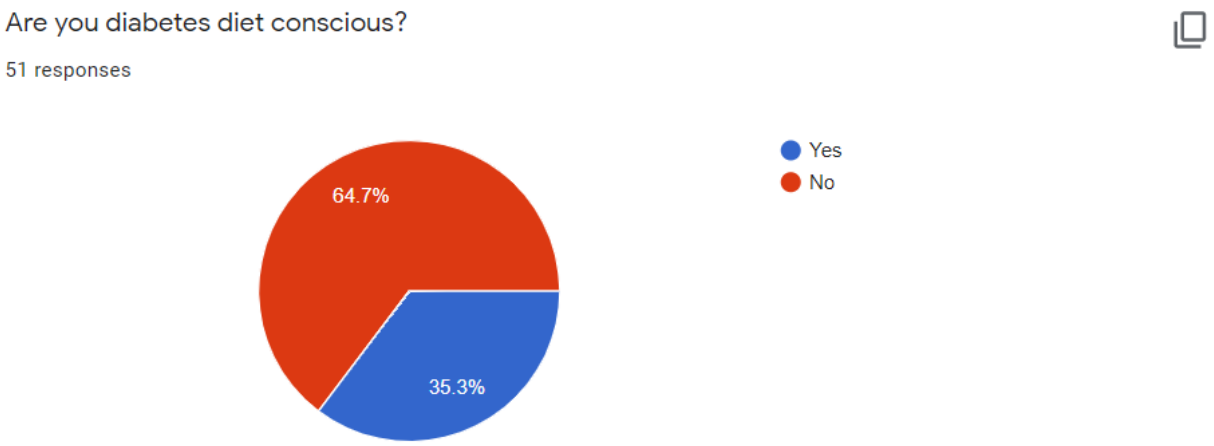


Figure 15: Checking the consciousness for diet

Implementation of methodology

A Software Development methodology refers to the pre-defined structure that are implemented to plan, manage and control the procedure of developing or engineering an information system or software. There are multiple of Software Development Life Cycle (SDLC) followed by many engineering fields. Every SDLC have a sequence of segments which must be followed and carry out throughout the development process in order to maintain flow of the project and deliver a complete project. Methodology that is chosen for this project is Waterfall model. Development of an artifacts is not an easy task. As adequate research was done for whole system development from selection of framework, language, AI algorithms, this system was fixed to be done with the methodology of waterfall. Waterfall is the type of methodology in which all steps has to carried out one after another in a sequence but one can shift to next step only after completing the current step completely and after going to next step, returning to previous step is prohibited.

Different stages like analysis, design, implementation, testing was carried out being inside the periphery of waterfall. Maintenance stage will be carried only after the deployment and feedback from users and requirement of the system. Various algorithms were tested and best one has been integrated in Django. Django takes the medical report and gives the output from the saved model (Bassil, 2012).

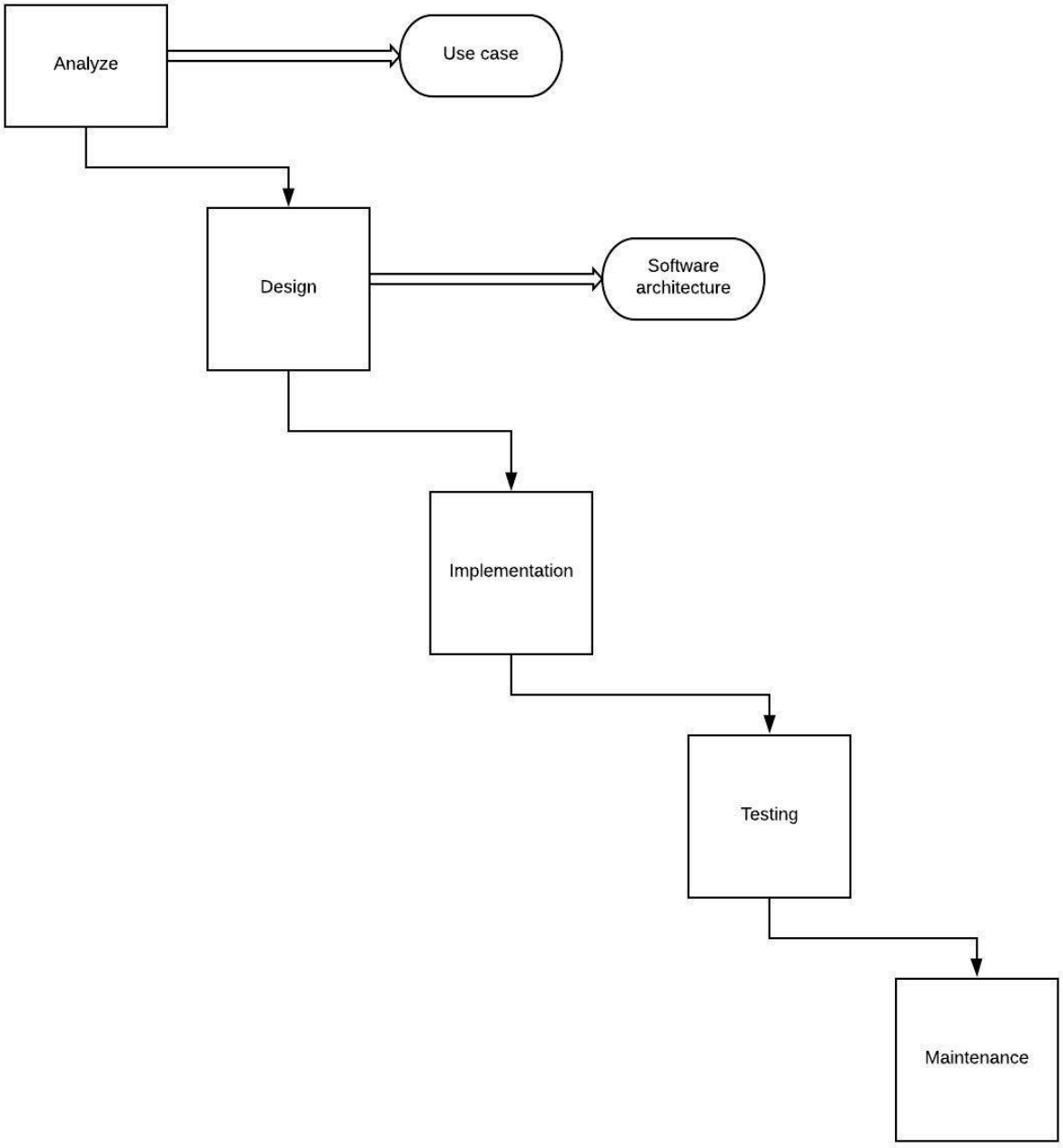


Figure 16 Water Fall Model for the project

Tabular Work Break Down Structure

All stages are explained in detail

Level 1	Level 2	Level 3
1) Diabetes Detection System	1.1) Analysis	1.1.1) Scope Identification 1.1.2) Feasibility Study 1.1.3) Resource Planning 1.1.4) SRS Document
	1.2) Design	1.2.1) System architecture 1.2.2) Use-case Diagram 1.2.3) DFD
	1.3) Implementation	1.3.1) Developing the working model 1.3.2) Developing the Django framework 1.3.3) System Development
	1.4) Testing	1.4.1) White Box Testing 1.4.2) Black Box Testing
	1.5) Deploying	1.5.1) Evaluation of Final Report

Table 2: Tabular representation of work-break down structure

Tree-form Work-breakdown structure

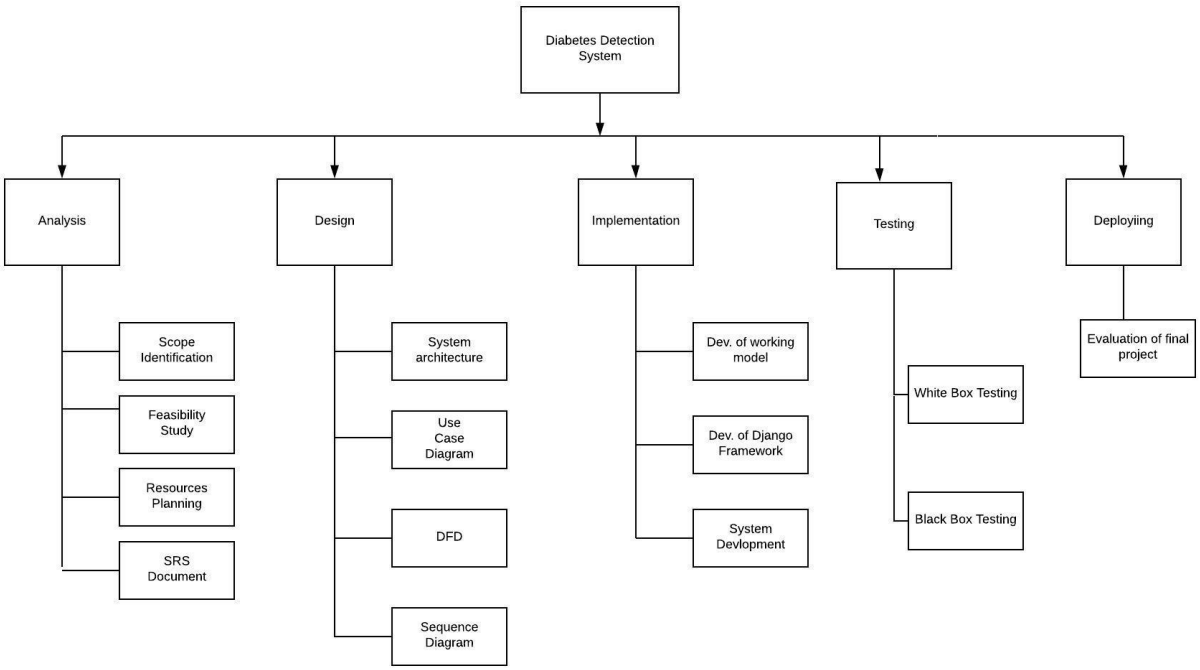


Figure 17 Tree diagram of Work-breakdown structure

Analysis

Requirement analysis, feasibility study and brainstorming were done for the research required, each step to carry out to complete the project. Scope of the project was well defined. Feasibility study was done on technical feasibility, resources feasibility, schedule feasibility. Analysis of the data was done of each attribute.

Functional Requirement

Functional Requirement defines the functional performance about the system which the system must be able to do when deployed. It shows all the component that a system should facilitate its users with the essential functionality.

Key No	Task	Description	Functional dependencies
1	Upload report	User will upload the medical report In the form to check the diabetes status	No
2	Classify the diabetes	System will classify the disease	1
3	Get the result	System will give the result after analyzing	1,2

Table 3: Functional Requirement

Non- Functional Requirement

Non-functional Requirement can also be called as the secondary needs of the system. Non-functional requirement shows how good the system is on the basis of various factors like security, usability, speed, reliability. Non-functional requirement is also known as quality requirement (Dabbagh1, et al., 2015).

Key No	Task	Description
1	Security	Security is always the major issue of every system. When security related question for user's report safety comes, the short and sweet answer is, this system does not store user's report in any format. User's data is just used by the model once to make a prediction.
2	Response	System response is appreciable. In every test it has responded with in less than 5 seconds.
3	UI/UX Design	Simple and easy to use.
4	System Availability	System will be available 24*7

Table 4: Non-functional Requirement

Design

System designing is essential to make a visualization of the working mechanism of the system. Gantt chart, System architecture diagram has been presented in this section of report.

Use-case Diagram

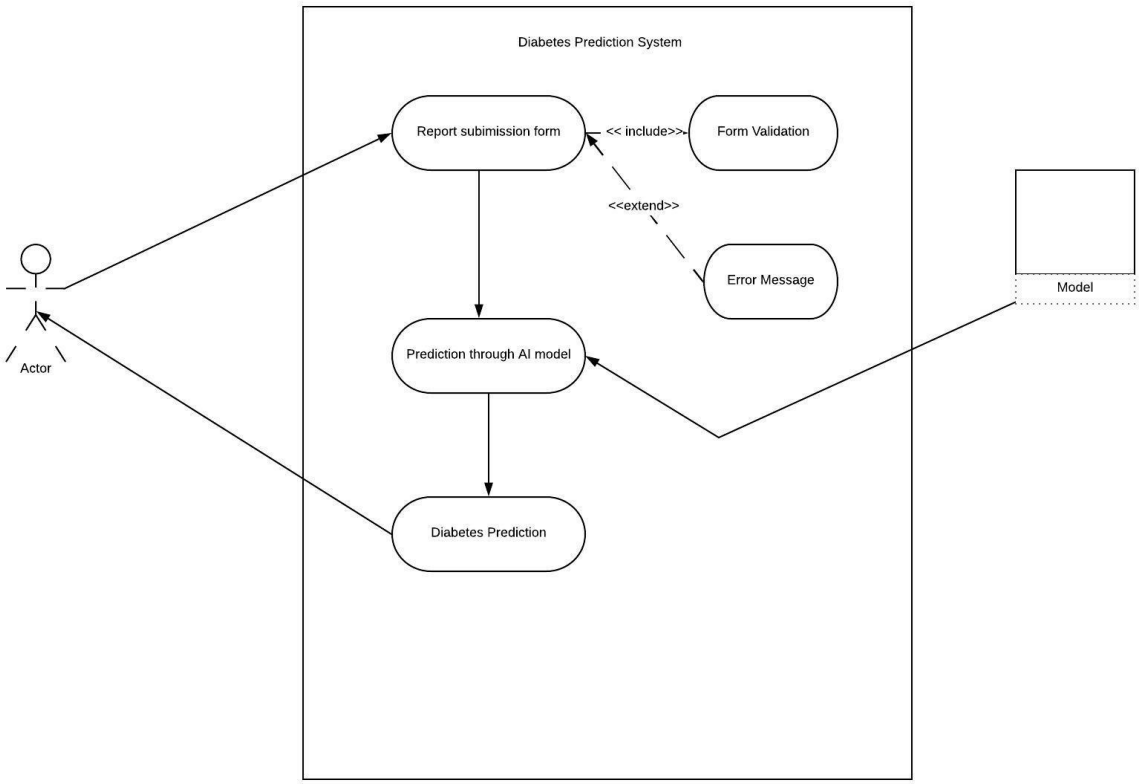


Figure 18 Use-case Diagram

Use-Case diagram of above figure explains the interaction of user with the system. As this is a complete open and free web-application for all types of user, user authentication has not been included in the system. User has to provide their medical report to the AI model through the form and after submitting the report, model will predict whether the report has the sign of diabetes or not. After the model gives the output, the result will get displayed on the web-application.

Data-flow Diagram

Context / Level 0 DFD

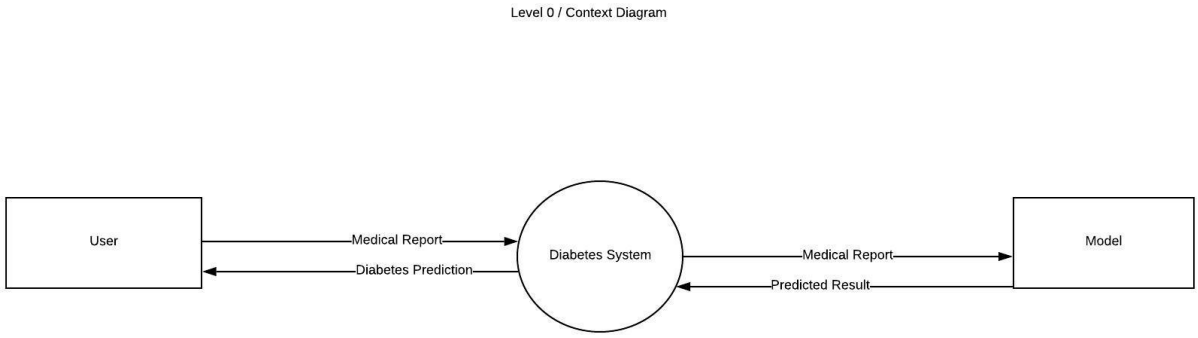


Figure 19 Context DFD

Level 1 DFD

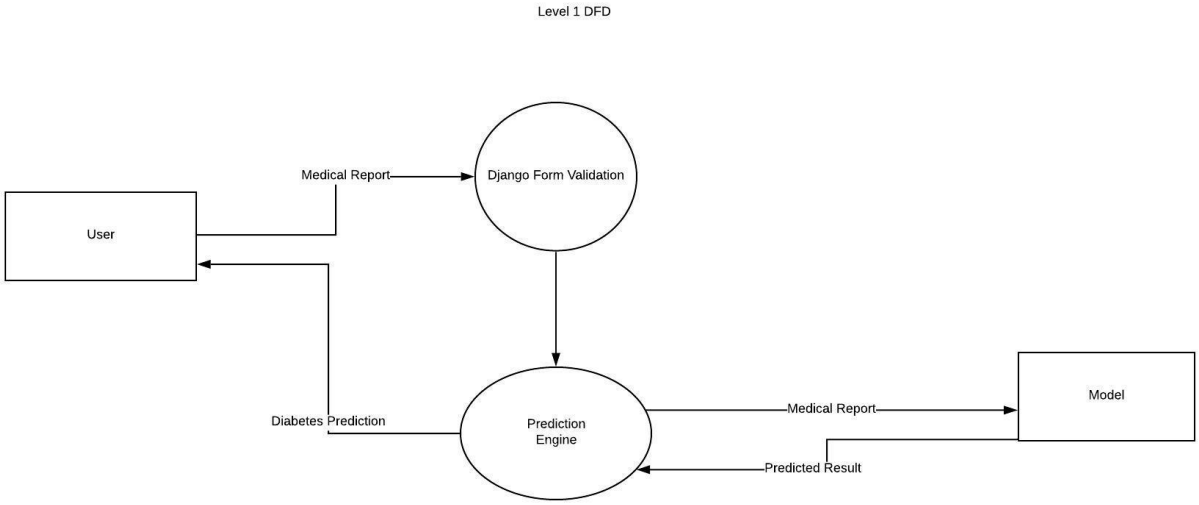


Figure 20 Level 1 DFD

Sequence Diagram

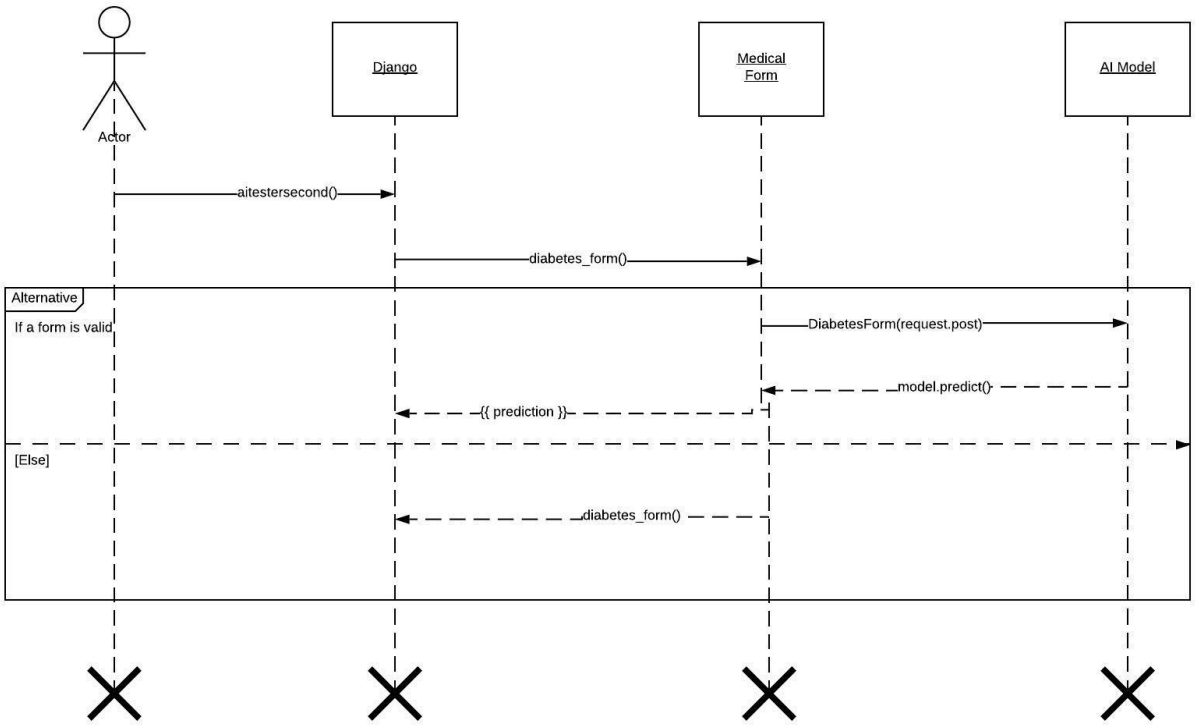


Figure 21: Sequence Diagram of the system

System Architecture

System architecture is based on Information Architecture.

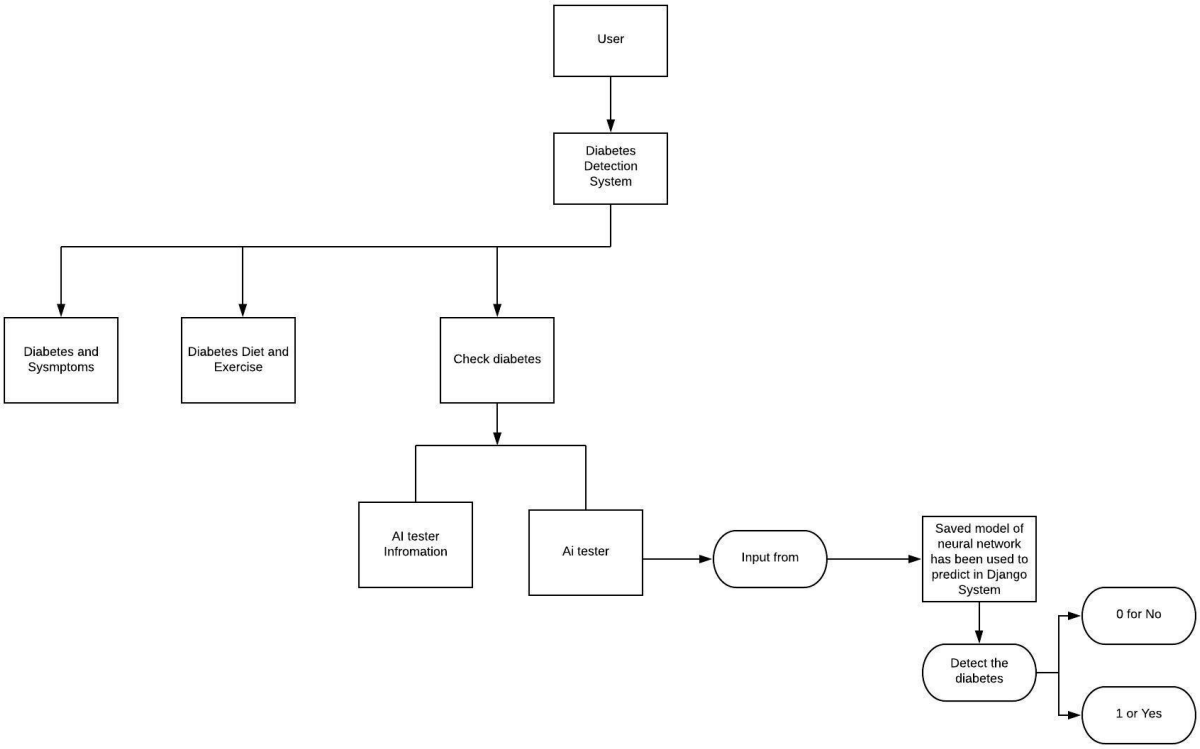


Figure 22 Whole System Architecture

The overall architecture of the system has been shown in the above diagram. This project has followed Information Architecture. User will access the web-application. In the web-app, there are three main pages with one home page. One web-page provides the information regarding diabetes and its symptoms, second web-page provides the information regarding diabetes's diet and the exercise and final web-page has sub web-page. Check diabetes will give the information about the AI system and required medical report to test in the AI model. AI tester will direct the user to form for testing the diabetes and after the submitting the form, the AI model will predict the diabetes and gives the output (Dilon & Don, 2005).

User Interface (design document)

Wireframe

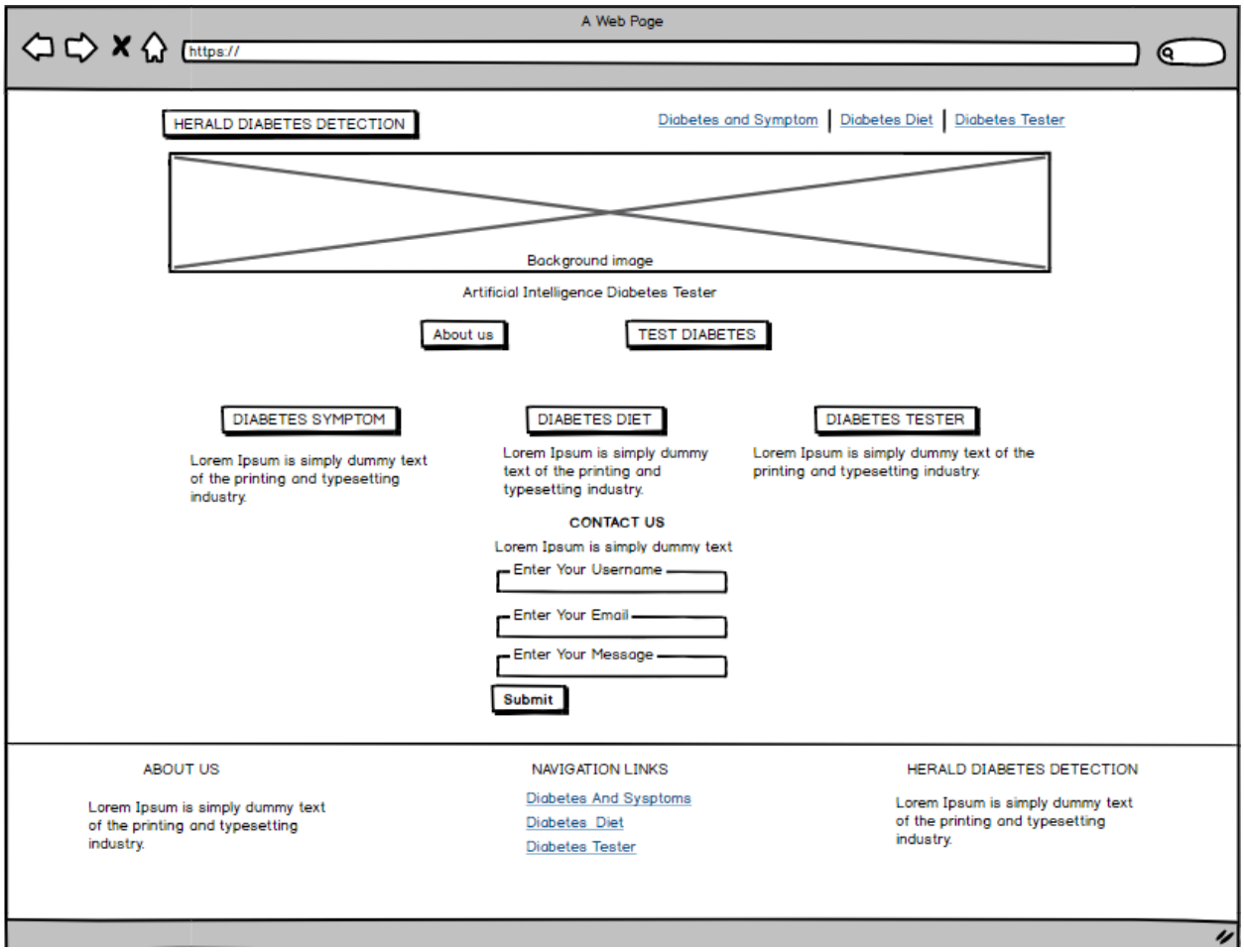


Figure 23: Wireframe of Home page of a web-app

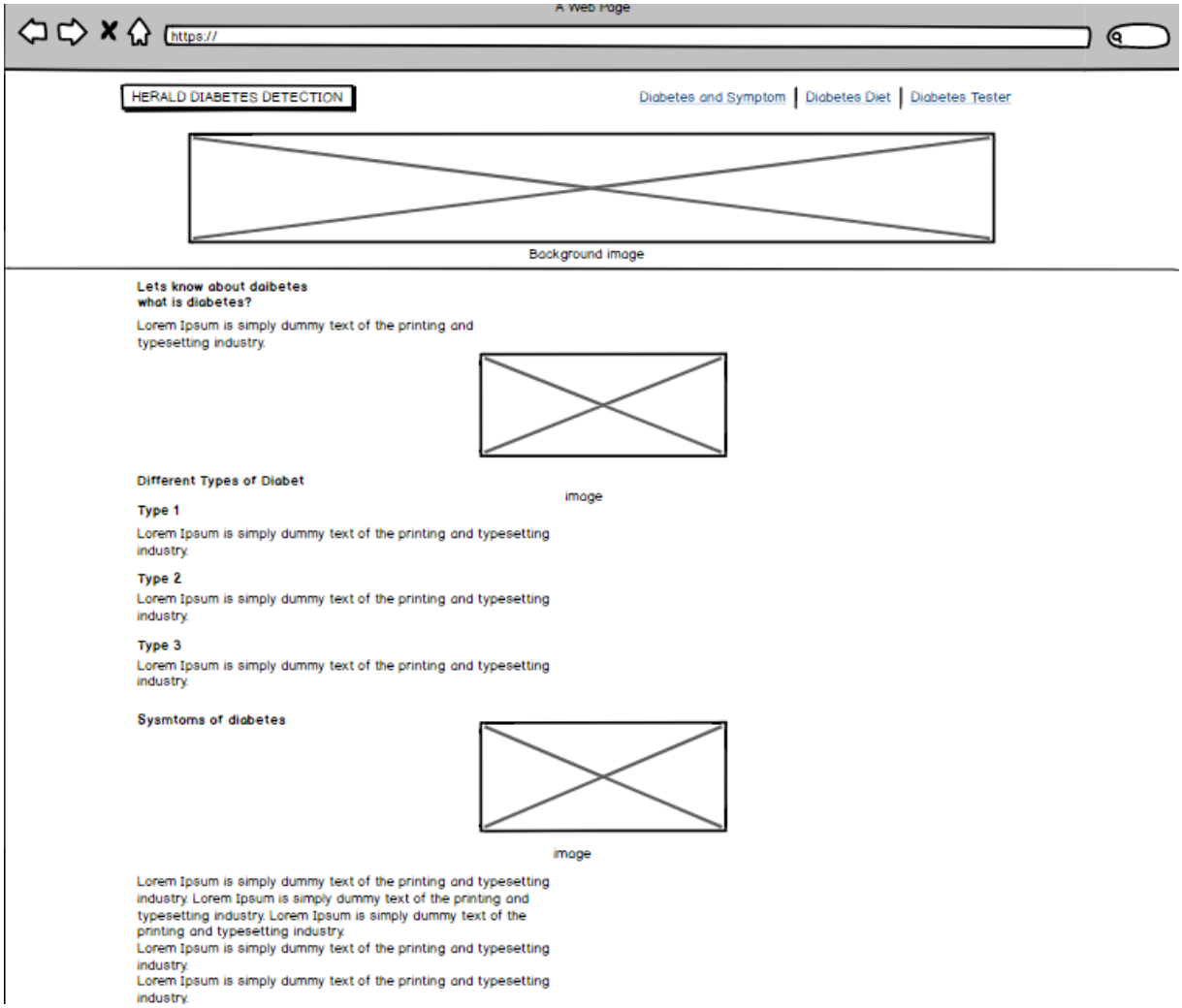


Figure 24: Wireframe of Diabetes and Symptom page

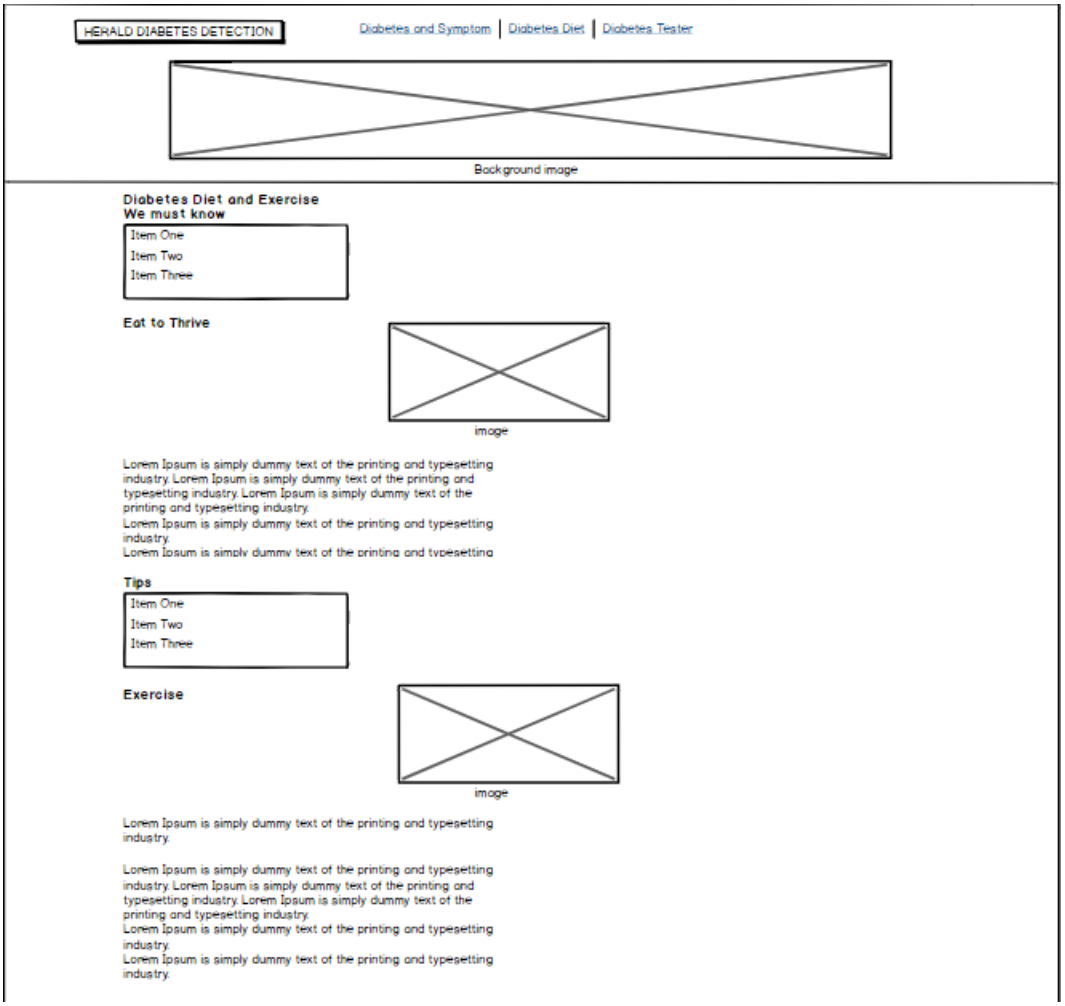


Figure 25: Wireframe of Diabetes Diet page

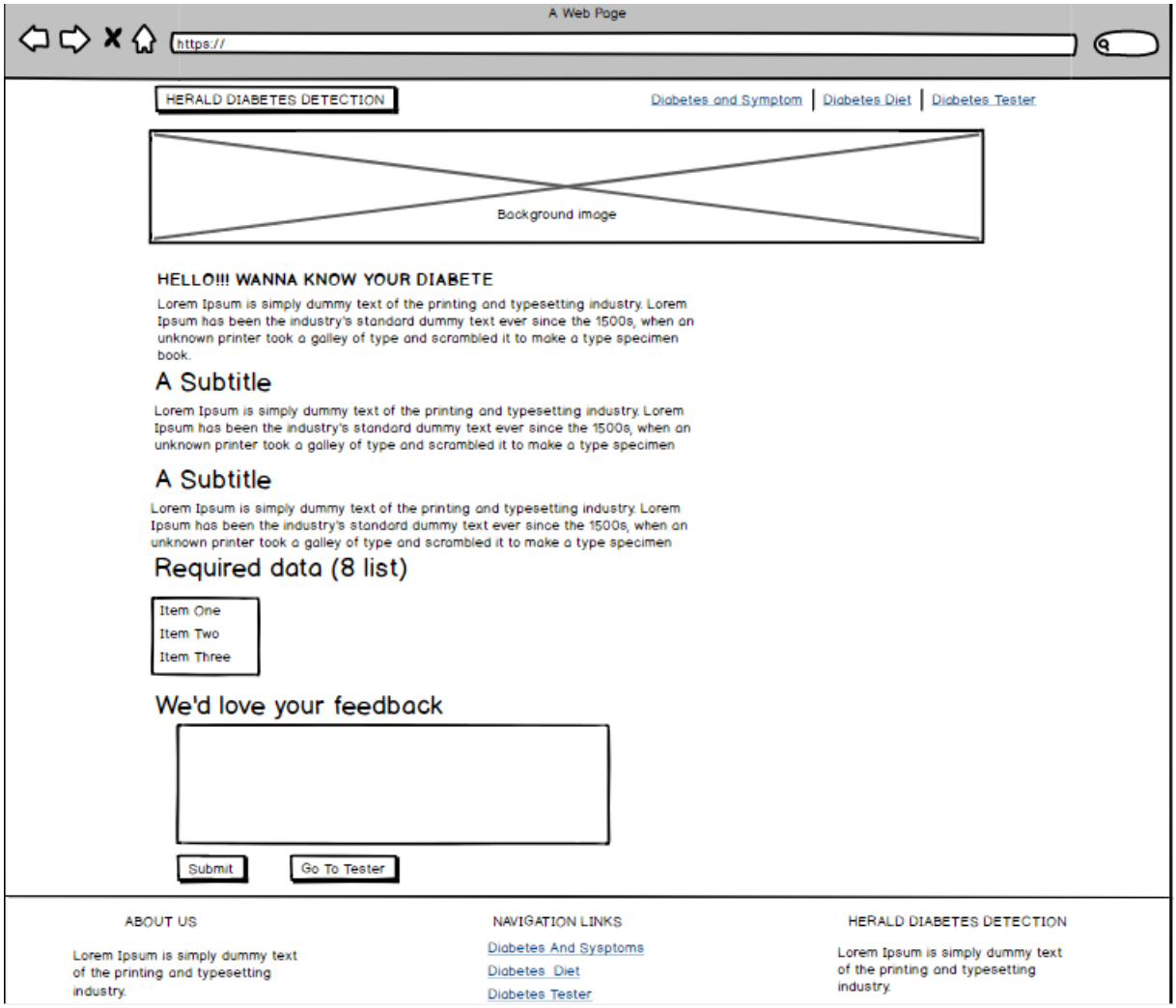


Figure 26: Wireframe of Diabetes Tester page

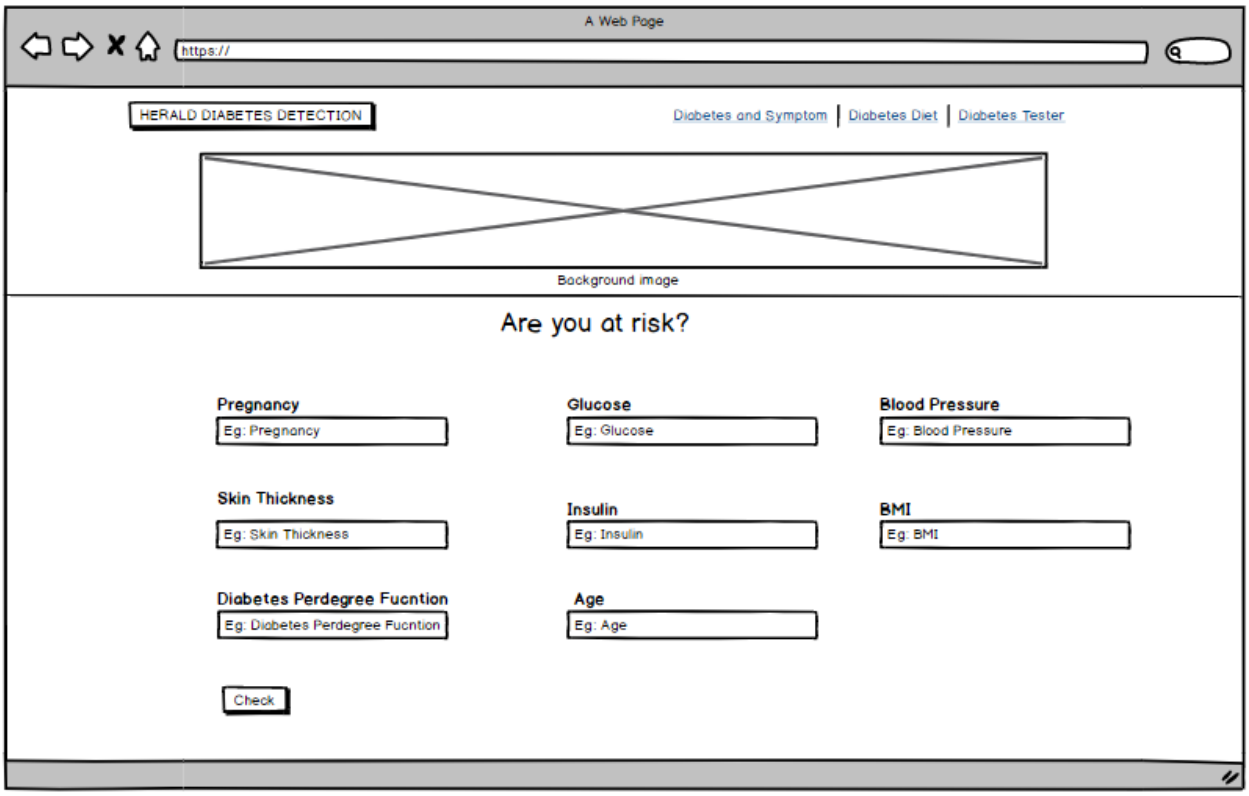


Figure 27: Wireframe of Tester form page

Implementation

Implementation of working model

Selection of programming language for developing whole system

To develop whole system, python language has been chosen. Some of the benefits with python are:

1. Readable
2. Mostly used for data science and machine learning
3. Presence of many built-in libraries
4. Worldwide developers support
5. It has fully supported IDE developed only for python.
6. Best framework support for developing web-application

Database Design

Data collections: Data is the core part of every AI system. We cannot imagine AI without data. For this project dataset has been collected from the Kaggle's Pima Indians Diabetes Dataset. This dataset has been collected and prepared by National Institute of Diabetes and Digestive and Kidney Disease and has total of 770 real time diabetes medical report. The dataset was originally in the form of csv and has been processed in the same file format throughout the development of the system. The only aim of this dataset is to predict if a person has diabetes or not based on some diagnostic calculations. This dataset comes under supervised learning and based on binary classification. The dataset has total of 9 columns with 8 attributes to measure the medical condition for diabetes and 1 is the output column.

After downloading the dataset, data cleaning and preprocessing set has been carried out to check if there are any missing value, or false or dirty data in the dataset. At first, dataset was imported into Jupyter notebook using pandas read_csv() function. .head () function displays the data present in the dataset.

```
In [1]: import numpy as np
        import pandas as pd
        dataset = pd.read_csv('D:/college/herald/5th sem/fyp/diabetes.csv')
        dataset.head(10)
```

Out[1]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

Figure 28 Importing data into notebook

After successful import of the data to pandas .info () function has been used to get the overall information of all the individual attributes inside the dataset.

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
Pregnancies                 768 non-null int64
Glucose                     768 non-null int64
BloodPressure               768 non-null int64
SkinThickness               768 non-null int64
Insulin                     768 non-null int64
BMI                         768 non-null float64
DiabetesPedigreeFunction    768 non-null float64
Age                         768 non-null int64
Outcome                     768 non-null int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

Figure 29 Analyzing each attribute

Form above figure, we can that out of 9 columns, BMI and DiabetesPedigreeFunction have the float64 as a datatype whereas remaining 7 have the datatype of int64. There is no any missing value in the dataset, as it has shown in above figure with the 'non-null' value with each attribute.

Brief description of each attribute has been listed below.

S.N.	Attributes Used	Datatype	Attributes Description
1	Pregnancy	Int64	For female, pregnancy plays the vital role for diabetes. This attribute takes the number of times the person has become pregnant. For male, the value will be 0.
2	Glucose	Int64	An oral glucose tolerance test of plasma glucose
3	BloodPressure	Int64	If the blood pressure is higher than normal then there will be higher probability of diabetes. This attribute measures the level of Blood Pressure.
4	SkinThickness	Int64	The obesity is one of the major factors which makes any human unhealthy and prone to diseases.
5	Insulin	Int64	
6	BMI	Float64	Body Mass Index measures the excess mass present in the body.
7	DiabetesPedigreeFunction	Float64	This attribute helps to calculate the prediabetes stages and chances of diabetes.

8	Age	Int64	Age is also the major factor for diabetes. Type 1 diabetes affects the younger people whereas type 2 diabetes affects elderly aged people.
9	Outcome	Int64	Output will have either 1 or 0 int value.

Table 5: Data Dictionary

Explanation of model architecture

As per the proposal, to build a best working model, four different algorithms were implemented and their accuracy were compared. Machine learning model such as logistic regression, decision tree, artificial neural network and naïve bayes algorithms were implemented in Jupyter notebook and confusion matrix was analyzed.

Explanation of each algorithms:

Logistic Regression:

Logistic regression is supervised machine learning algorithm used in classification-based problems. In logistic regression is mostly used for binary class classification i.e. output variable has usually two variables. Logistic regression is also known as logistic model ad logit model. Logistic regression is also used for analyzing the relationship in between multiple independent variables as well as categorical dependent attributes. Logistic regression gives an output on the basis of probabilistic values. Using logistic regression in The Pima Indians Dataset, it got an accuracy of 87% (J Korean Acad Nurs, 2013).

```
In [18]: from sklearn.linear_model import LogisticRegression

In [19]: dataLR = LogisticRegression()
         dataLR.fit(Xtrain, Ytrain)

         /home/shree/.local/lib/python2.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning: Default solver will be chan
         ged to 'lbfgs' in 0.22. Specify a solver to silence this warning.
           FutureWarning)

Out[19]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, max_iter=100, multi_class='warn',
                   n_jobs=None, penalty='l2', random_state=None, solver='warn',
                   tol=0.0001, verbose=0, warm_start=False)

In [20]: accuracy = dataLR.score(Xtest,Ytest)
         print("accuracy = ", accuracy * 100, "%")

         ('accuracy = ', 87.01298701298701, '%')
```

Figure 30: Application of Logistic Regression

Decision Tree:

Second algorithm that was applied to analyze the accuracy in the diabetes dataset was decision tree. Decision tree is like a tree structure, that cab be visualize in the form of flowchart. It uses multiple of nodes and internodes to classify and to predict the classes. All the nodes that are inside the tree are considered as an internal node and are test cases. Those internal nodes separate all the instances with various features. After applying decision in the dataset, it got an accuracy of 76 % (Iyer, et al., 2015).

```
In [25]: from sklearn.tree import DecisionTreeClassifier
         tree = DecisionTreeClassifier(max_depth=4,random_state=0)

In [26]: tree.fit(Xtrain,Ytrain)

Out[26]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=4,
                     max_features=None, max_leaf_nodes=None,
                     min_impurity_decrease=0.0, min_impurity_split=None,
                     min_samples_leaf=1, min_samples_split=2,
                     min_weight_fraction_leaf=0.0, presort=False, random_state=0,
                     splitter='best')

In [27]: print("Accuracy on training set: {:.3f}".format(tree.score(Xtrain,Ytrain)))
         print("Accuracy on test set: {:.3f}".format(tree.score(Xtest,Ytest)))

         Accuracy on training set: 0.790
         Accuracy on test set: 0.766
```

Figure 31: Application of Decision Tree

Naïve Bayes:

Another machine learning algorithm that was tested was Naïve Bayes. Naïve Bayes is a mathematical classifier that makes an assumption of having no dependency between variables. The aim of Naïve Bayes is to make an attempt to maximize the following probability in defining the class. Applying the dataset in Naïve Bayes got an accuracy of 80.52, which can be considered as a descent accuracy as well (Patel, et al., 2013).

```
from sklearn.naive_bayes import GaussianNB

datanb = GaussianNB()
datanb.fit(Xtrain,Ytrain.ravel())
GaussianNB(priors=None)
```

```
GaussianNB(priors=None, var_smoothing=1e-09)
```

```
nb_predict= datanb.predict(Xtest)
print(nb_predict)
```

```
[1 0 0 1 0 0 1 1 1 0 1 1 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1
 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 0 1 1 1 1 0 0 0 0 0 0 1
 1 0 1]
```

```
from sklearn import metrics
print("Accuracy: {0:.4f}".format(metrics.accuracy_score(Ytest,nb_predict)))
```

```
Accuracy: 0.8052
```

Figure 32: Naive Bayes Classifier

As neural network got the higher accuracy than any other algorithms, neural network was further tested and implemented in pycharm IDE to integrate with Django framework.

Architecture of neural network:

```
In [19]:
import keras
from keras.models import Sequential
from keras.layers import Dense
from keras.optimizers import Adam

Using TensorFlow backend.

In [20]:
model = Sequential([
    Dense(16, input_dim=8, activation='relu'),
    Dense(16,activation='relu'),
    Dense(1,activation='sigmoid'),
])

In [21]:
# model = Sequential()
# model.add(Dense(18, input_shape=(8,), activation='relu'))
# model.add(Dense(32,activation='relu'))
# model.add(Dense(4, activation='relu'))
# model.add(Dense(1,activation='sigmoid'))

In [22]:
model.summary()

Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 16)                144
_________________________________________________________________
dense_2 (Dense)              (None, 16)                272
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 17
=================================================================
Total params: 433
Trainable params: 433
Non-trainable params: 0
_________________________________________________________________
```

Figure 33 Architecture of neural network

From fig:14, we can see, Keras model has been used to build a neural network model. Keras is the open source machine learning library which is TensorFlow's high level API and use backend of TensorFlow. Keras was developed to provide an easy and clean platform to develop a deep learning models because of this reason keras is the choice of all the beginners as well as intermediate level developers.

Why Keras??

1. Reliable, modest and extensible API.
2. Easy to apply and attain output without any complex installments.
3. User friendly framework that can run in both GPU and CPU.
4. It has good support from the large community all-round the globe.
5. It is built in python language which make it readable.
6. It can support any complex and deep neural network like CNN, RNN as well (Chollet, 2019).

Selection of Dense Layers:

Dense layer is the fully connected layer in which every output from each neurons of each layers are connected to input of the neurons of next layer. The dense layer takes the input and return the output in the following ways: Output= activation (dot (input, kernel) +bias)

Selection of Sequential Model:

Sequential model is made up of 5 predefined layers which needed to be done in a sequence. It is easy to understand and has the capability to perform in all types of neural network. Layers of sequential model are:

1. Defining mode architecture

Architecture of the model has been defined by using 3 dense layers. First dense has used 16 neurons, input_dim has been set to 8 as the input parameter of the model is 8 and for an activation function, ReLu has been used.

ReLu is always used as a compulsory activation function in hidden layers. The full form of ReLu is rectified linear units. It has the threshold at 0 i.e.

f(x) = max (0, x), which means if the value is less than or equal to 0, it will make it 0 and constant for value above 0.

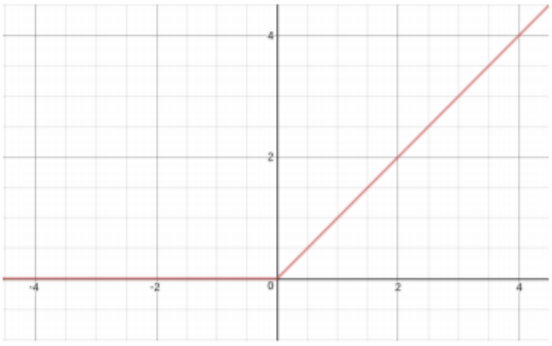


Figure 34 ReLu activation function

Second dense has also used same activation function and neurons. In final layer, only one neuron has been passed and for activation function sigmoid has been

used as the binary classification. Sigmoid gives the result on the basic of 0 and 1 for two output classes.

2. Compile

```
In [23]:  1  model.compile(Adam(),loss='binary_crossentropy',metrics=['accuracy'])
```

Figure 35 Keras model. compile with optimization algorithm, loss function and metrics.

Model. compile helps to select the optimization and loss function which is the most important steps for calculating the error and optimizing the model by updating the loss value. This process of updating the loss value by continuously updating and minimizing the loss is known as backpropagation.

Optimizer is one of the important arguments along with selecting loss function and metrics which helps to optimize the model in the process of back-propagation. Adam is also known as an adaptive learning rate optimization algorithm that has been developed to train and optimize deep neural network (Bushaev, 2018).

3. Fit

```
In [24]:  1  trainfit=model.fit(Xtrain,Ytrain, epochs=60, batch_size=50, verbose=2, validation_split=0.1)
WARNING:tensorflow:From /home/shree/.local/lib/python2.7/site-packages/keras/backend/tensorflow_backend.py:422:
The name tf.global_variables is deprecated. Please use tf.compat.v1.global_variables instead.

Train on 621 samples, validate on 70 samples
Epoch 1/60
 - 1s - loss: 0.7134 - accuracy: 0.4847 - val_loss: 0.6951 - val_accuracy: 0.5286
Epoch 2/60
 - 0s - loss: 0.6849 - accuracy: 0.5942 - val_loss: 0.6692 - val_accuracy: 0.6714
Epoch 3/60
 - 0s - loss: 0.6607 - accuracy: 0.6715 - val_loss: 0.6455 - val_accuracy: 0.7286
Epoch 4/60
 - 0s - loss: 0.6382 - accuracy: 0.6973 - val_loss: 0.6226 - val_accuracy: 0.7429
Epoch 5/60
 - 0s - loss: 0.6171 - accuracy: 0.7118 - val_loss: 0.6012 - val_accuracy: 0.7571
Epoch 6/60
 - 0s - loss: 0.5965 - accuracy: 0.7166 - val_loss: 0.5813 - val_accuracy: 0.7571
Epoch 7/60
 - 0s - loss: 0.5760 - accuracy: 0.7311 - val_loss: 0.5612 - val_accuracy: 0.7429
Epoch 8/60
 - 0s - loss: 0.5566 - accuracy: 0.7343 - val_loss: 0.5421 - val_accuracy: 0.7429
```

Figure 36 Setting epoch, batch size and dividing the data in validation split

Model.fit is all about training the model. It takes labelled data of Xtrain and Xtest as an input and train the model for given epochs in batch size. For this model epochs have been set to 60 and batch size has been set to 50 which means it will take 50 training sample from a training dataset at a time and will run 60 times. Validation_split has been set to 0.1 which means 10% of

data from training sample will get divided and stored as a validation set which is one of the forms of unseen data for a model.

4. Evaluate

```
In [25]:
1 _, test_accuracy=model.evaluate(Xtest,Ytest,batch_size=5,verbose=2)
2 print("accuracy =", test_accuracy)

('accuracy =', 0.8051947951316833)
```

Figure 37 Displaying the model final accuracy in unseen data

This layer helps to calculate the accuracy after the model has been trained. For now, model. Evaluate has been implemented in test dataset with batch_size 5 and verbose 2.

5. Predict

As model.evaluate has used test data set, model.predict will be used directly after deploying the system.

Evaluation of final accuracy of the model:

```
In [27]:
1 C_matrix = confusion_matrix(Ytest, predictions)
2 C_matrix

Out[27]: array([[43,  8],
                [ 7, 19]])

In [28]:
1 print(classification_report(Ytest, predictions))

              precision    recall  f1-score   support

           0       0.86      0.84      0.85        51
           1       0.70      0.73      0.72        26

   micro avg       0.81      0.81      0.81        77
   macro avg       0.78      0.79      0.78        77
weighted avg       0.81      0.81      0.81        77
```

Figure 38 Confusion matrix

From the above confusion matrix, we can see out of 77 data used for testing the model, 52 data have been predicted correctly by the model. In 52 data 43 have been predicted as True Negative and 19 have been predicted as True Positive. Unfortunately, model has wrongly predicted 15 test data. 8 data have been predicted as False Positive whereas 7 data have been predicted as False Negative.

Classification report helps us to analyze the accuracy on the basic of precision, recall and f1-score. As this model is of health domain, the case will directly get case sensitive. And we can see the accuracy in term of all three components of analyzing the accuracy is in good range.

Selection of framework for Web-application

As, for this project, python programming language has been chosen for developing the model, Django is the best option to build a web-application. Some of the reasons for choosing Django as a framework are:

1. Django has many pre build functions and methods.
2. Larger community support
3. Security is high
4. Can be scaled as per the need and requirement

Implementation of Django as a Web-application:

Django is the popular high-level framework which uses python as a backend for developing the web-application. Django has been developed to support rapid development of the web-application with clean and user-friendly practical design. Whenever we develop web-application, concept of MVC architecture comes automatically. The full form of MVC is Model, View, and Controller. Django follow similar but slightly different pattern, which is known as MVT. MVT stands for Model-View-Template. The major difference in MVC and MVT is, following the MVT pattern, Django looks after the Controller part by itself so that the developer does not have invest the time in coding to make an interaction between The View and The Model of the system. The T in MVT is the template which is a HTML file that comes with Django Template Language (DTL) (tutorialspoint, 2015).

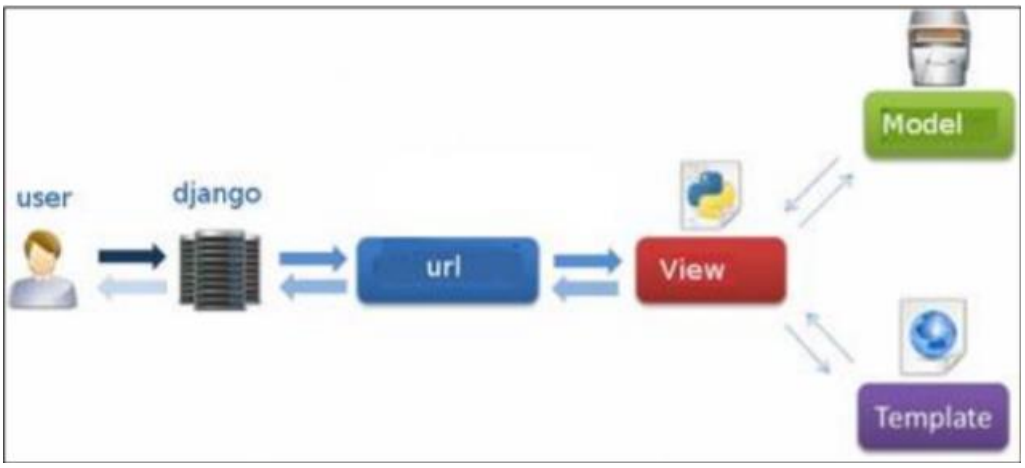


Figure 39 Architecture of Django Framework

To run the web-application in Django server, python manage.py runserver command has to be used in a terminal being inside the python virtual environment.

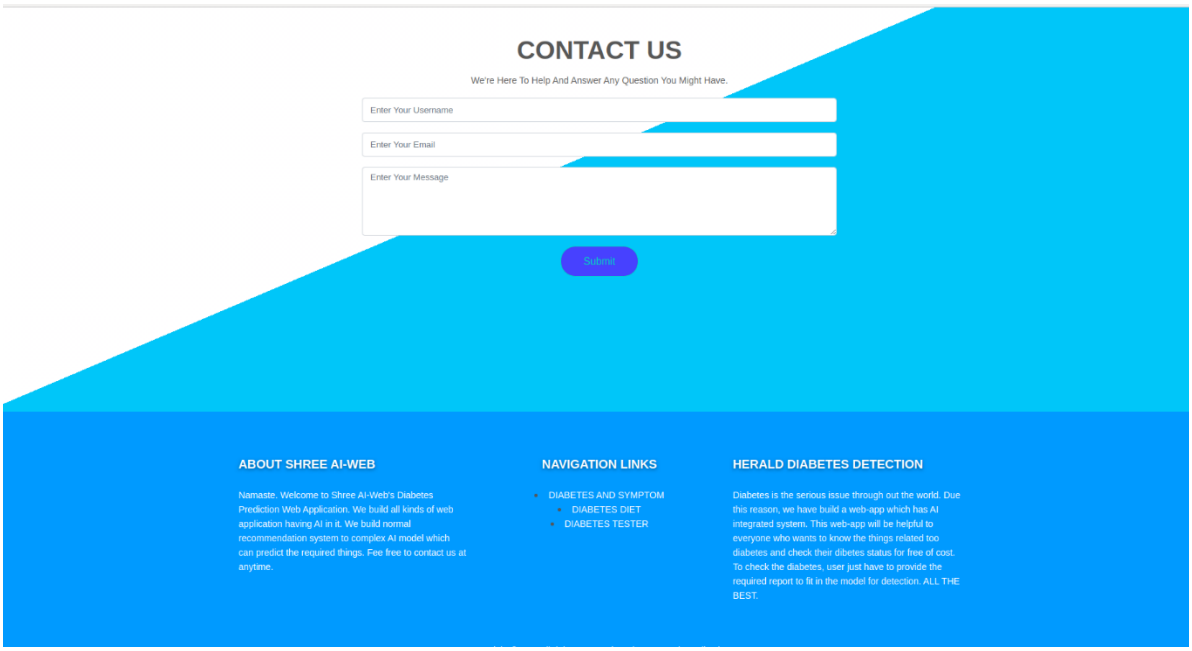


Figure 40 Home Page

Above page is the index/ home page of the web-application which will get displayed at the first.

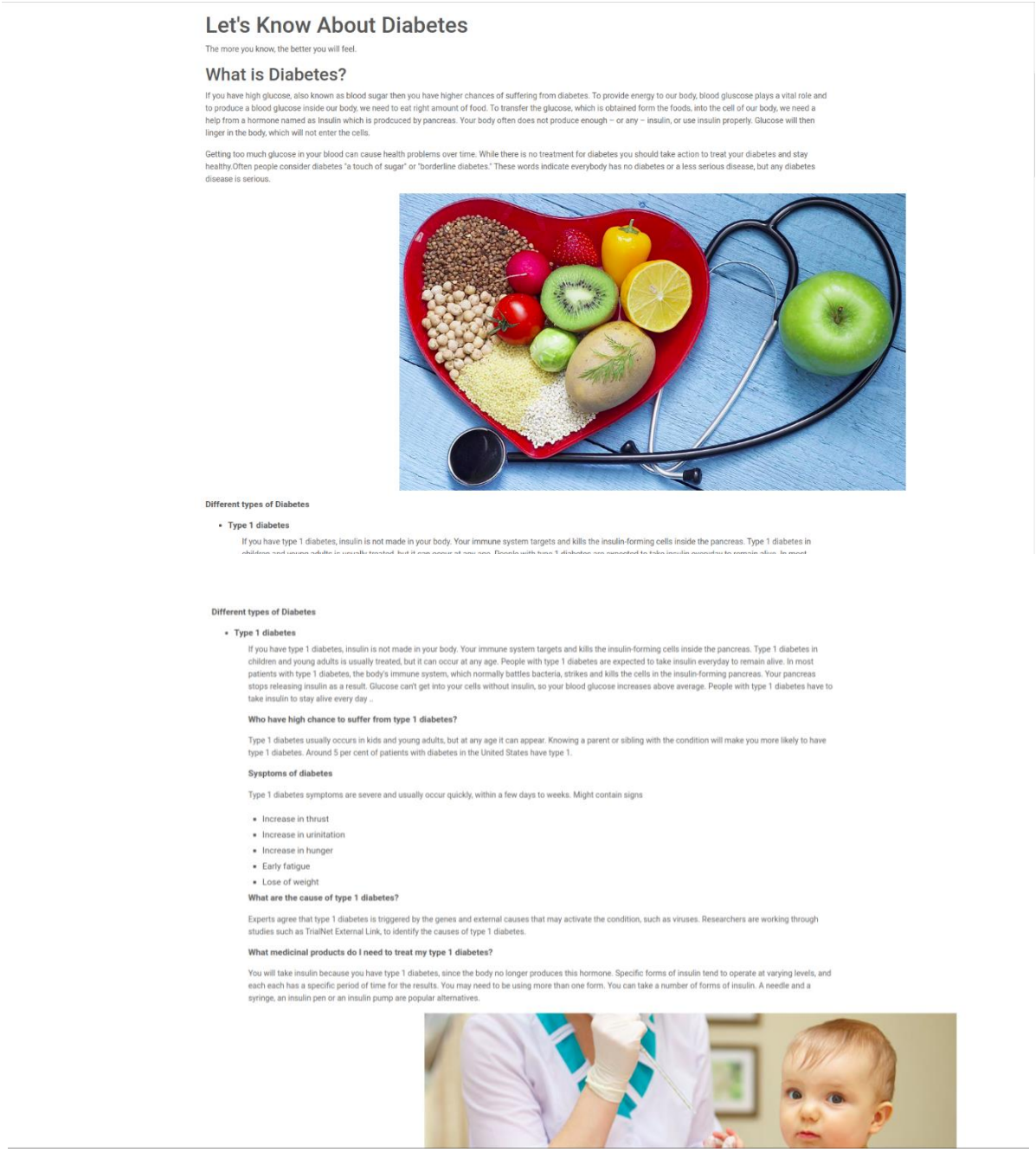
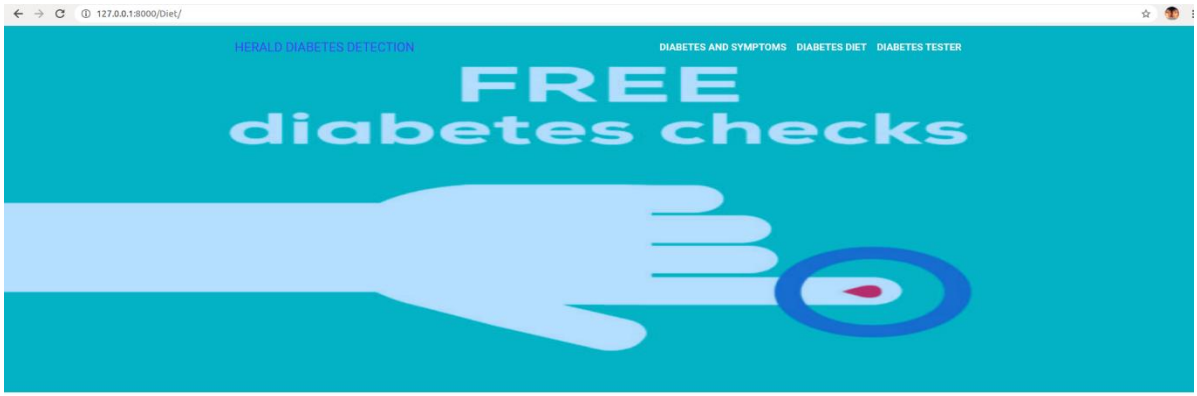


Figure 41 Diabetes Symptom Page

This page will provide all the fundamental information related to diabetes, types of diabetes and cause of each types of diabetes. As the main aim of this web-application is to provide an open platform for everyone to check the diabetes for free of cost, web-app does not have the user authentication section. Anyone can come up with their medical report and check their sugar level status with ease.



Diabetes Diet & Physical Activity

WE MUST KNOW

- What are the foods and drink can I have if I have diabetes?
- Which foods should I avoid or limit if I have diabetes?
- List of physical activities i must include if I have diabetes

Eat to thrive

Nutrition and physical activity are important parts of a healthy lifestyle when you have diabetes. Along with other benefits, following a healthy meal plan and being active can help you keep your blood glucose level, also called blood sugar, in your target range. To manage your blood glucose, you need to balance what you eat and drink with physical activity and diabetes medicine, if you take any. What you choose to eat, how much you eat, and when you eat are all important in keeping your blood glucose level in the range that your health care team recommends. To manage your blood glucose, you need to balance what you eat and drink with physical activity and diabetes medicine, if you take any. What you choose to eat, how much you eat, and when you eat are all important in keeping your blood glucose level in the range that your health care team recommends.

Becoming more active and making changes in what you eat and drink can seem challenging at first. You may find it easier to start with small changes and get help from your family, friends, and health care team.

Eating hygienic foods and doing physically actives in good amount can help you

- keep your BP, glucose level and cholesterol in health rate
- maintains your weight according to your age and height
- controls and avoid high diabetes risk
- keep your mental health sound and feels good
- provide energy for whole day

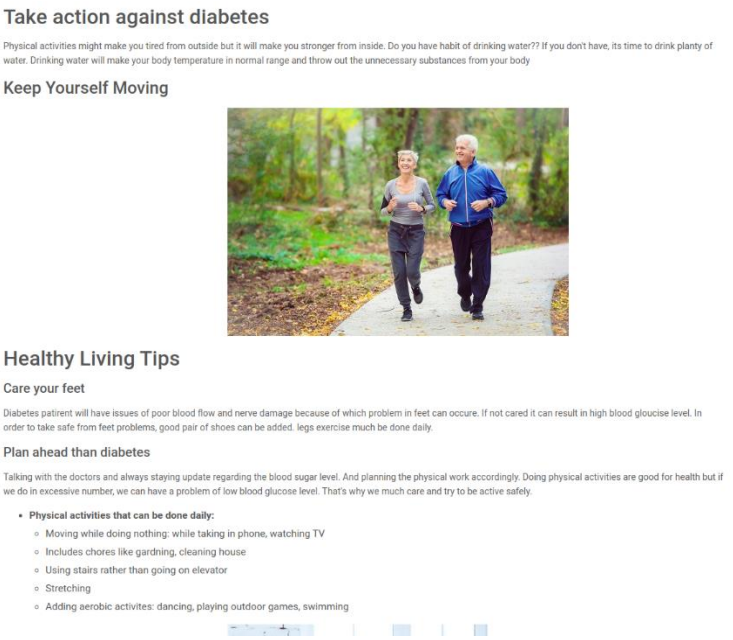


Figure 42 Diabetes Diet and Exercise Page

This page will provide brief information to maintain the healthy lifestyle and basic information regarding the diet to eat and to avoid, exercises to focus if anyone has the symptoms or are suffering from the diabetes.

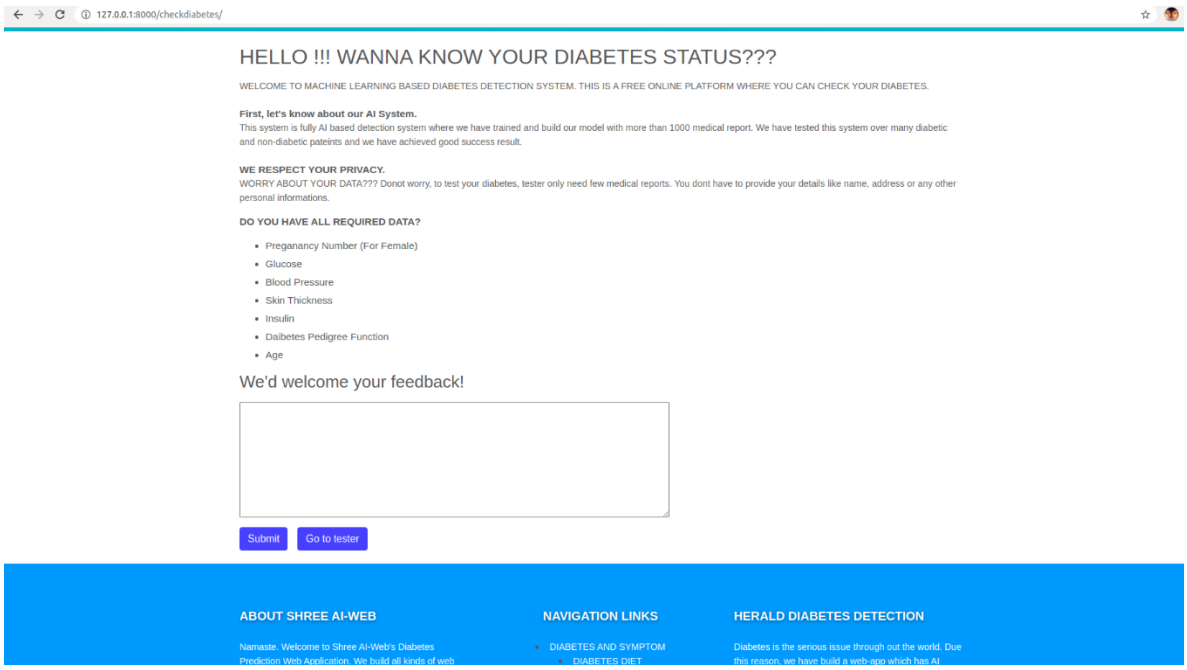


Figure 43 Check Diabetes Page

This is the core section of the web-application which gives the access to the model to test the medical report for diabetes. Basic information has been given regarding the AI model and the list of medical report needed to test the diabetes. Which will be helpful for the user who has visit the web-app for the first time. If the user has all the medical report ready, he/she can proceed to the test page which directly links with the model.

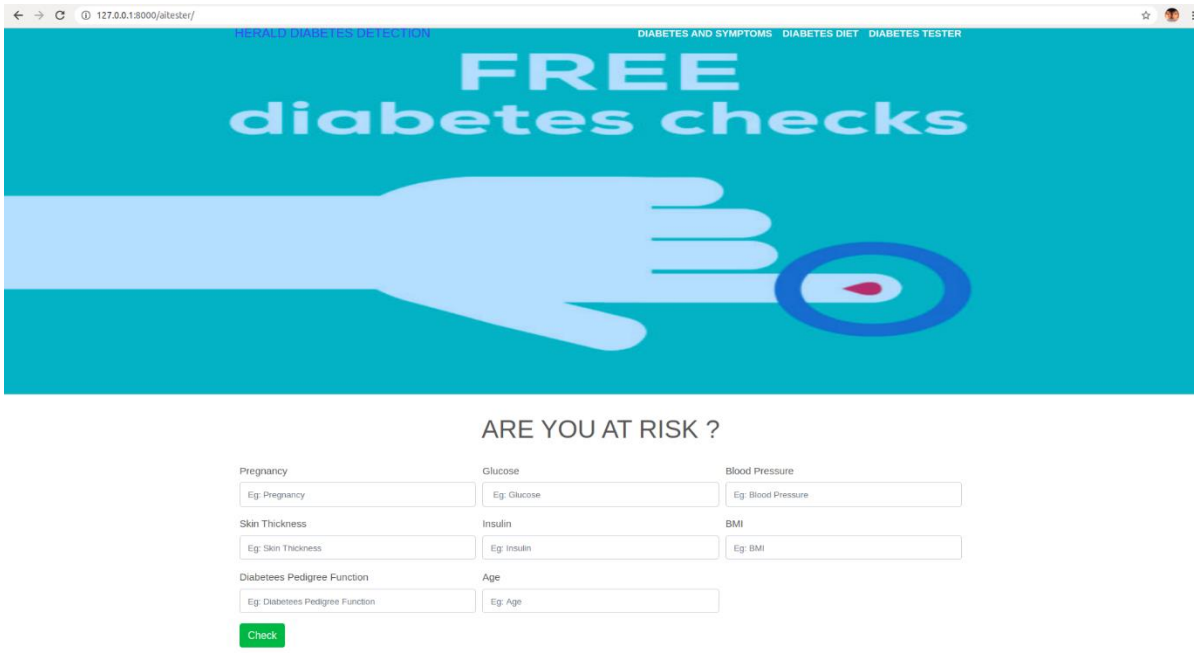


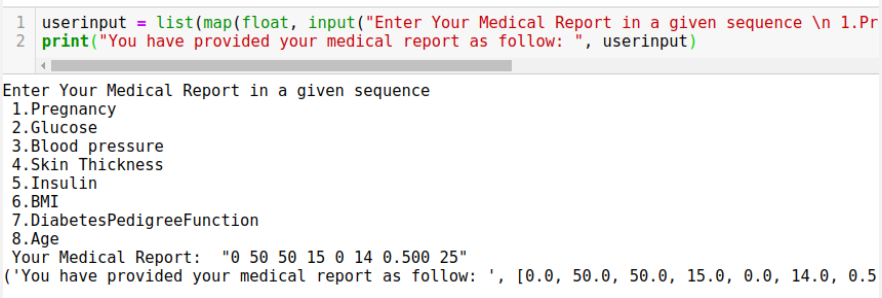
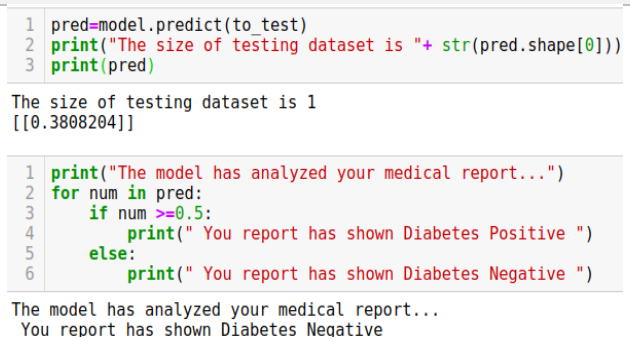
Figure 44 AI tester Page

This is the fourth page which is linked with the machine learning model directly. From the above image of the page, we can see, this page has one form with 8 sections. All sections are the attributes of the model to make a prediction. Pregnancy can be made 0 for every male user and for those female users who has not been pregnant till data. After filling the form and submitting the report will go the model. Model will analyze the report and make a prediction whether the medical report has the sign of diabetes or not. Pop-up message will be displayed, displaying the result in term of Yes or No. ‘Yes’ for presence or sign of diabetes and ‘No’ for no sign of diabetes.

Testing

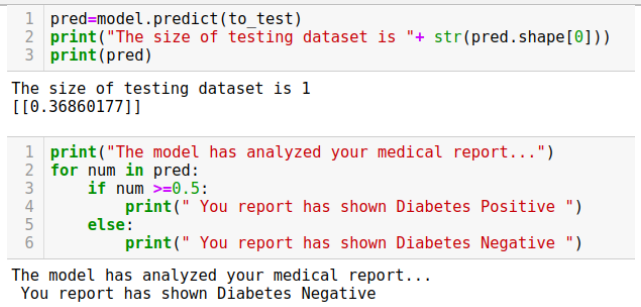
White-box Testing

Testing has been done with 3 test cases. To get the real-time diabetes data from user, python's input function has been used. Split() function to split the string value on the basis of space. Map () function has been used to typecast the input into float and at last, the final set of data has been stored in list to feed to model. Briefing and justification of each sample has been done below in tabular form as part of white box testing.

<table>
<tr><td>Test Case No.</td><td colspan="3">1</td></tr>
<tr><td>Test Case Objective:</td><td colspan="3">Predict Diabetes Negative</td></tr>
<tr><td>Test Data</td><td colspan="2">Data
 
</td><td>Expected Result
Diabetes Negative</td></tr>
<tr><td>Actual Result</td><td colspan="3"> 
</td></tr>
<tr><td>Test Result</td><td>Pass</td><td colspan="2">(Diabetes Negative)</td></tr>
<tr><td>Justification</td><td colspan="3">Medical report of all variables with low value was passed to the model to check whether it will predict diabetes as negative or not. Testing was successful with correct prediction.</td></tr>
</table>

<table>
<tr><td>Test Case No.</td><td colspan="2">2</td></tr>
<tr><td>Test Case Objective</td><td colspan="2">Predict Diabetes Negative</td></tr>
<tr><td>Test Data</td><td>Data

Enter Your Medical Report in a given sequence
1.Pregnancy
2.Glucose
3.Blood pressure
4.Skin Thickness
5.Insulin
6.BMI
7.DiabetesPedigreeFunction
8.Age
Your Medical Report: "2 89 66 23 94 28.1 0.167 21"
('You have provided your medical report as follow: ', [2.0, 89.0, 66.0, 23.0, 94.0, 28.1, 0.167, 2</td><td>Expected Result

Diabetes Negative</td></tr>
<tr><td>Actual Result</td><td colspan="2"> 
</td></tr>
<tr><td>Test Result</td><td>Pass</td><td>Diabetes Negative</td></tr>
<tr><td>Justification</td><td colspan="2">Taking the reference from original data that has diabetes negative case, above data was prepared with an intension to get diabetes negative as a prediction. Testing was successful with correct prediction.</td></tr>
</table>

<table>
<tr><td>Test Case No.</td><td colspan="3">3</td></tr>
<tr><td>Test Case Objective :</td><td colspan="3">Predict Diabetes Positive</td></tr>
<tr><td>Test Data</td><td colspan="2">Data</td><td>Expected Result</td></tr>
<tr><td></td><td colspan="2">

```
userinput = list(map(float, input("Enter Your Medical Report in a given sequence \n 1.Pregnancy
print("You have provided your medical report as follow: ", userinput)
```

```
Enter Your Medical Report in a given sequence
 1.Pregnancy
 2.Glucose
 3.Blood pressure
 4.Skin Thickness
 5.Insulin
 6.BMI
 7.DiabetesPedigreeFunction
 8.Age
 Your Medical Report:  "6 148 72 35 0 33.6 0.627 50"
('You have provided your medical report as follow: ', [6.0, 148.0, 72.0, 35.0, 0.0, 33.6, 0.627, 50.
```

</td><td>Diabetes Positive</td></tr>
<tr><td>Actual Result</td><td colspan="3">

```
pred=model.predict(to_test)
print("The size of testing dataset is "+ str(pred.shape[0]))
print(pred)
```

```
The size of testing dataset is 1
[[0.33723274]]
```

```
print("The model has analyzed your medical report...")
for num in pred:
    if num >=0.5:
        print(" You report has shown Diabetes Positive ")
    else:
        print(" You report has shown Diabetes Negative ")
```

```
The model has analyzed your medical report...
 You report has shown Diabetes Negative
```

</td></tr>
<tr><td>Test Result</td><td>Fail</td><td colspan="2">Diabetes Negative</td></tr>
<tr><td>Justification</td><td colspan="3">Higher values were chosen for diabetes positive result but, model made a false prediction. Possible reasons can be in the original dataset, there was the dominance of negative data over positive:
1. Because of this, the model was trained with negative cases more and has overfit with negative data.
2. In original dataset, insulin value has not been prioritized enough but insulin plays the vital role in maintaining the sugar level in human body.
To solve this issue:
1. Object serialization / pickle can be used.
2. Data up sampling can be used.</td></tr>
</table>

Out of three instances, two predictions were correct whereas 1 was incorrect.

Black-box Testing

For Black-box testing of AI-model following testing has been done.

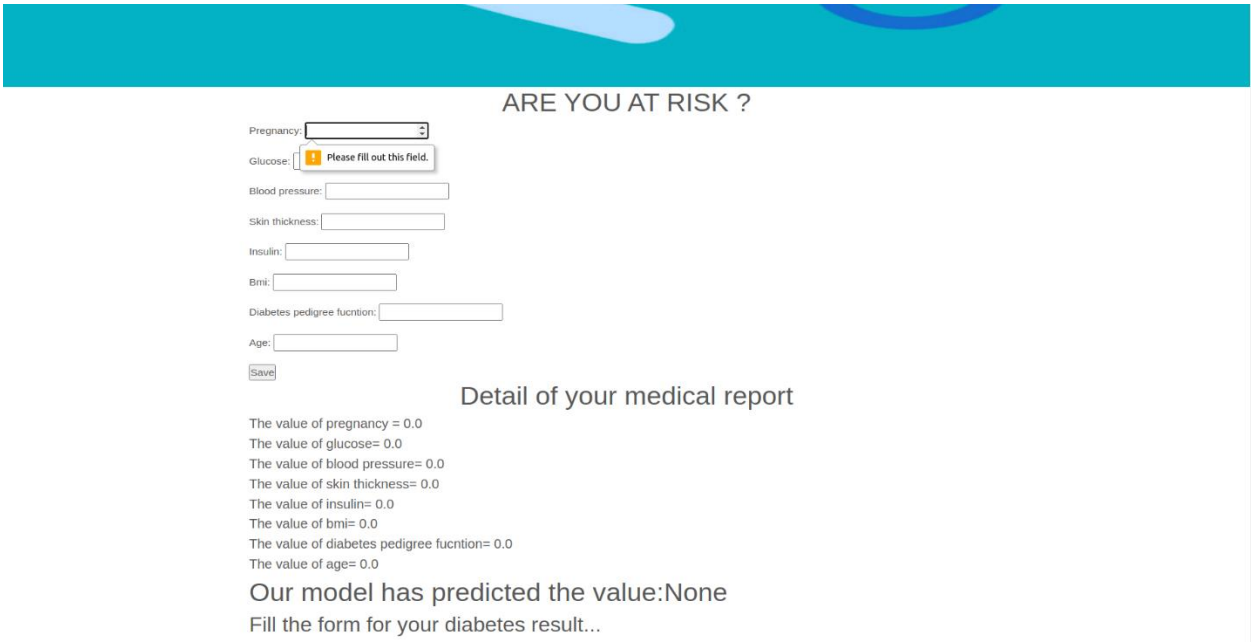


Figure 45: Form Validation, default 0 value and default message in result display section

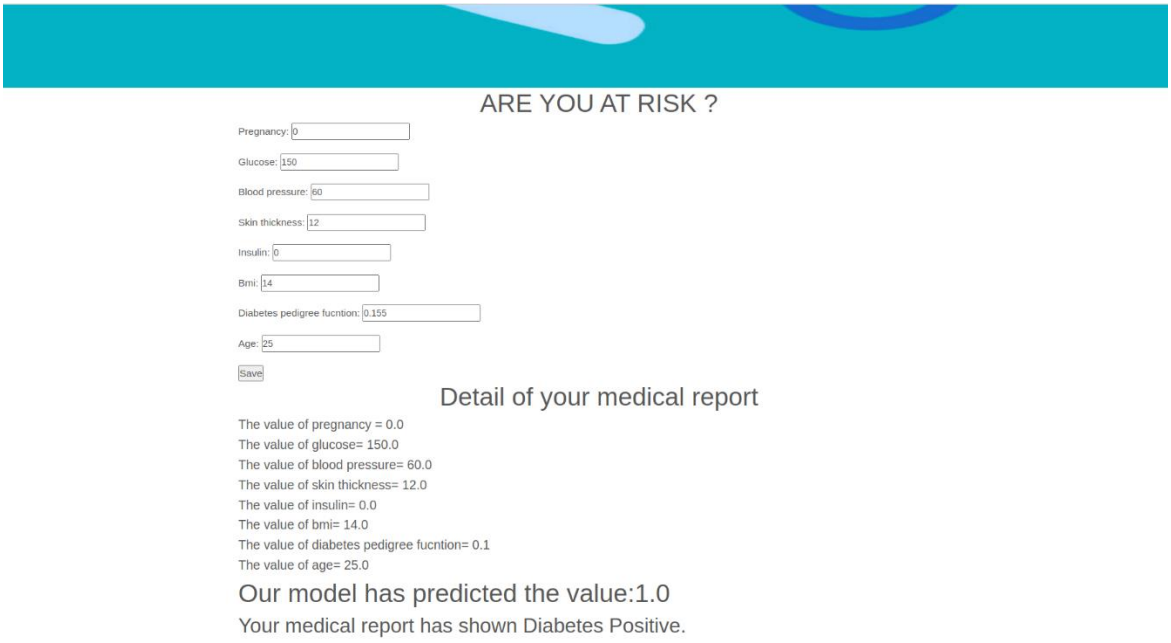


Figure 47: Filling form with correct value, displaying the value for confirmation and model's prediction

Academic Question

Answering the Academic Questions

- Can diabetes be accurately predicted using Machine Learning and Deep Learning model?

Yes, diabetes can be predicted accurately using an AI model. Health has also been the hot topic for artificial intelligence. Either it is AI or any other field, experts of the domain will always look at least once that if the field he/she is working can serve human-kind or not. Many researches have been carried since past 3 decades and many AI scientists and researchers are still studying and researching for better result. Not only diabetes, every disease can be predicted and cured with the assistance of AI and machine learning algorithms but if and only if the model has been taught and guide in the right manner. Similar to the newly born baby. If the data has a fault or false information, no machine or human can judge anything good or bad, or right or wrong. Focusing on this project model, there are still many places to improve but still it can detect the diabetes right with 84% accuracy. It can be improved by adding more data and making the architecture of the model big.

- How the doctors will get benefit by this model?

Population of whole world is increasing rapidly. But if we analyze the ratio of number of doctors to the population of any country, number of doctors are not increasing in the same way as the population. Because of this, doctors are having hard time to watch every patient. Human being is prone to errors. He/she gets tired, frustrated and feel monotonous. Because of the pressure everyone makes a mistake. But health is such a sector where one cannot afford to make a mistake. One mistake can cost someone's life. Analyzing this situation, if there will be a system which can work like a human being and have an ability to make a decision like a human than it will be a great help for every doctor. Model like this can be used as a first chase checkup which can assist the doctors in preparing the report and go for further diagnosis as per needed.

Conclusion

At the beginning of the project, the aim was to find out the real-time diabetes dataset and to analyze the hidden patterns using various classification algorithms like Decision Tree, Naïve Bayes, Logistic Regression and Neural Network. Data was taken from The Pima Indian Diabetes Dataset with 9 attributes and 769 total records. After building the model with the same dataset and four different machine learning's classification algorithms and comparing their final performance on the basis of accuracy, Neural Network got the highest accuracy of 84% and outperformed other algorithms. Coming to this stage, the goal of building the AI model to predict the diabetes was developed. After the development of AI model, all focus was on building the web-application to provide the UI for the system, so that the general public can use it and check their diabetes/ sugar level. Python's Django framework was selected to build a web-application. Focusing on the academic question, which has asked, "can diabetes be accurately predicted using Machine Learning and Deep Learning model?", after the final development of the model with the accuracy of 84% in recall and 85% in f1 score, we can say, yes, diabetes can be predicted with the help of machine learning algorithm.

At last, after implementing four different algorithms, all algorithms have their own way of implementing the dataset and working mechanism in order to achieve the best accuracy and performance. If we train the model with cleaned and normalized dataset, the model will learn without having any problem and will not provide misleading results.

Critical Evaluation

Findings

After researching and studying multiple of journals, I found that most of the journals have suggested and used traditional machine learning classifiers such as Decision Tree, Naïve Bayes, etc. Use of Artificial Neural Network for classifying diabetes is rear, but after researching about the neural network and other algorithms, I find out that, if the data size increases and becomes too large, the traditional machine learning algorithms will not have the similar consistent like it used to have during small – medium dataset in term of performance and accuracy. In fact, Neural Network was built to solve this issue and works best with large amount of data. After months of research and implementing all the research, I find out, data is the heart of AI model and the performance of every AI model depends on the quality of data. Strength of the system is to predict the diabetes with real-time medical report with an accuracy of 84%. This accuracy can be improved in Neural Network by tuning the hyper parameter such as learning rate, epochs size, batch size and adjusting number of layers and neurons, choice of activation functions as well as improving the dataset.

Evaluation of final system:

As the data is the lifeline for AI, quality and quantity of the dataset can be improved to increase accuracy. More source of data needs to be explored for better dataset. Inside the model, pattern of training can also be modified from train_test_split to k-fold cross validation so that even if the dataset is less, the purposed method will regulate the dataset for k times. Being in the medical domain, every case is linked to health and are sensitive. Model's performance on the basis of 'recall' can be improved for better accuracy in sensitive case classification.

Limitation of the system:

- This system is limited to classify whether the report has sign of diabetes or not but will not separate the type of diabetes.
- As per the scope of the project, system will classify diabetes but due to absence of authentication, admin cannot perform digital marketing.

Future escalation:

System will be able to,

- classify the type of diabetes.
- more attributes will be added in the dataset.
- will ask the users for a permission to save their medical report so that the data can be utilized for improving the performance of the model.
- will have another AI system to track report, which will forward an email of suggestions and diet plans for those whose report was diabetes positive while testing in the system.
- web-application will have another page for showcasing the success classification and improvement of user.

References

Aggarwa, C. C., 2018. *Neural Network and Deep learning.* 1st ed. Yorktown Heights, NY, USA: Springer.

Aggarwal, C. C., 2018. *Neural Network and Deep Learning.* 1st ed. Yorktown Heights, NY, USA: Springer.

American Diabetes Association, 2014. Diabetes Care. *Diagnosis and Classification of Diabetes Mellitus,* Volume 37.

Atherton, J., 2011. HISTORY OF MEDICINE. *Development of the Electronic Health Record ,* Volume 13, pp. 186-189.

Bassil, Y., 2012. A Simulation Model for the Waterfall Software Development Life Cycle. *International Journal of Engineering & Technology (iJET),* 2(5), pp. 2049-3444.

Bushaev, V., 2018. *towards data science.* [Online] Available at: https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c
[Accessed 20 05 2020].

C.Dharuman, P.Venkatesan & Sharmila, S., 2017. International Journal of Pure and Applied Mathematics. *Disease Classification Using Machine Learning Algorithm- A Comparitive Study,* 114(06), pp. 1-10.

Chollet, F., 2019. Keras tutorials point. *Keras,* pp. 1-17.

Ciox, 2014. *Ciox.* [Online] Available at: https://www.cioxhealth.com/about-us
[Accessed 14 04 2020].

Dabbagh1, M., Lee, S. P. & Parizi, R. M., 2015. Functional and non-functional requirements prioritization: empirical evaluation of IPA, AHP-based, and HAM-based approaches. *Soft Computing.*

Dheir, I. M., Mettleq, A. S. A., Elsharif, A. A. & Mohammed N. Abu Al-qumboz, S. S., 2019. Knowledge Based System for Diabetes Diagnosis Using SL5. *International Journal of Academic Pedagogical Research (IJAPR),* 3(4), pp. 1-10.

G.Subbalakshmi, Ramesh, K. & Rao, M. C., 2011. Indian Journal of Computer Science and Engineering (IJCSE). *Decision Support in Heart Disease Prediction System using Niave Bayes,* Volume 2, pp. 170-176.

Ganesan, K., Acharya, U. R. & Chua, C. K., 2013. IEEE REVIEWS IN BIOMEDICAL ENGINEERING. *Computer-Aided Breast Cancer Detection Using mammograms,* Volume 6, p. 77.

Habibi, S., Ahmadi, M. & Alizadeh, S., 2015. Global Book of Health and Science. *Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining,* 5(7).

Habibi, S., Ahmadi, M. & Alizadeh, S., 2015. Global Journal of Health Science. *Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining,* 5(7), pp. 304-310.

Ishak, W. H. W. & Siraj, F., 2008. *ARTIFICIAL INTELLIGENCE IN MEDICAL APPLICATION: AN EXPLORATION.*

Iyer, A., Jeyalatha, S. & Sumbaly, R., 2015. DIAGNOSIS OF DIABETES USING CLASSIFICATION MINING TECHNIQUES. *International Journal of Data Mining & Knowledge Management Process (IJDKP) ,* 5(1), pp. 4-5.

J Korean Acad Nurs, 2013. An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. 43(2), p. 155.

M.Mijwel, M., 2018. Artificial Neural Network Advantages and Disadvantages. *Artificial Neural Network Advantages and Disadvantages.*

Mettleq, A. S. A., Dheir, I. M., Elsharif, A. A. & Mohammed N. Abu Al-qumboz, S. S., 2019. Knowledge Based System for Diabetes Diagnosis Using SL5. *International Journal of Academic Pedagogical Research (IJAPR),* 4(3), pp. 1-10.

Naser, P. S., 2015. SL5 OBJECT: SIMPLER LEVEL 5 OBJECT. *International Jornal of Soft Computing Mathematics and Control (IJSCMC),* 4(4).

Ni, L., Lu, C., Liu, N. & Liu, J., 2017. Knowledge and Systems Sciences. *MANDY: Towards A Smart Primary Care Chatboc Application,* Volume 780.

Patel, S. B., Yadav, P. K. & P.Shukla, D. D., 2013. Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques. *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS),* 4(2), pp. 61-64.

Poudel, A., 2019. *THE KATHMANDU POST.* [Online] Available at: https://kathmandupost.com/national/2019/04/01/hospitals-countrywide-ailing-due-to-lack-of-doctors-and-nurses
[Accessed 21 04 2020].

R.Patil, M., 2014. International Journal of Advanced Research in Computer and Communication Engineering. *Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing,* 3(5), pp. 6787-6789.

Ramesh, A., Kambhampati, C. & Monson, J., 2004. *Artificial intelligence in medicine,* 86(5), pp. 334-338.

Rodgers, G. P., 2017. *National Institute of Diabetes and Digestive and Kidney Diseases.* [Online]
Available at: https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes
[Accessed 25 03 2020].

Rodriguez-Ruiz, A. et al., 2019. JNCI J NATI Cancer Institute. *Stand-Alone Aetificial Intelligence for Breast Cancer,* 11(9).

Rothe, R., 2017. *medium.com.* [Online]
Available at: https://medium.com/merantix
[Accessed 14 04 2020].

Scheffler, R. M., Liu, J. X. & Kinfu, Y., 2008. Bulletin of the World Health Organization. *Forecasting the global shortage of physicians: an economic- and needs-based approach,* Volume 86, pp. 497-576.

Soofi, A. A. & Awan, A., 2017. Journal of Basic & Applied Sciences. *Classification Techniques in Machine Learning: Applications and Issues ,* Volume 13, pp. 459-465.

The Economic Times, 2019. *THE ECONOMICS TIMES.* [Online]
Available at: https://economictimes.indiatimes.com/industry/healthcare/biotech/healthcare/india-facing-shortage-of-600000-doctors-2-million-nurses-study/articleshow/68875822.cms
[Accessed 21 04 2020].

tutorialspoint, 2015. django. *tutorialspoint SIMPLYEASYLEARING,* pp. 1-2.

Vara by Merantix Healthcare, n.d. *Vara.* [Online]
Available at: https://www.varahealthcare.com/
[Accessed 14 04 2020].

WelcomeAI, 2019. *Welcome AI.* [Online]
Available at: https://www.welcome.ai/quotient-health
[Accessed 14 04 2020].

West, D. M. & Allen, J. R., 2018. *Bookings.* [Online]
Available at: https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/
[Accessed 14 04 2020].

World HEALTH ORGANIZATION, 2020. *global health workforce alliance.* [Online]
Available at: https://www.who.int/workforcealliance/countries/npl/en/
[Accessed 21 04 2020].