Software Requirement Specification Document for Utilizing human genomes analysis to improve athletic performance

Donya Mohamed Mohy Eldien ,Mennat Allah Hisham , Mazen Mohamed,
Dannel Maged , Ibrahim Bibers
Supervised by:
Dr. Abdel Nasser Zaied
Eng.Yomna Ibrahim
Eng.Ahmed Hazem

December 29, 2020

Table 1: Document version history

Version	Date	Reason for Change
1.0	23-Dec-2020	SRS First version
1.1	28-Dec-2020	Add use Case ,Class Diagram ,similar systems

GitHub: https://github.com/MennatAllah13/Graduation

Abstract

The main idea of this project is to combine artificial intelligence and genomics to make a system using genome analysis such as DNA sequences or gene expression to identify physical points of strength. This help anyone to understand his/her traits, and how it affects his/her sports life. The system will recommend a suitable sport based on variation in genes between normal and super athletes. Besides, We will also set a group of genes responsible for causing some diseases using artificial intelligence algorithms to predict future diseases, so they can avoid risks, and increase their chances to reach the peak of the competition. Moreover, we are attempting to determine the type of muscle's tissue type whether its Constriction or extension,(fast or slow). the program classify genes' locations, and what are the functions of these genes, and its traits whether endurance, power, speed, or strength using machine learning or techniques. Our challenge is to help athletes improve their performance and the non-athletes to discover their abilities to choose the sport that will match their body features.In future work the similarity between super athletes and people helps in determining the capabilities of the future image of an athlete.

1 Introduction

1.1 Purpose of this document

The purpose of this software requirement document is to present a detailed description of utilization of human genomes analysis to discover athletes traits using machine learning tools. the main purpose of the project is to be able to classify the variation in genes for an athlete so we can recommend a specific sport depending on these traits(genes) . The earlier this analysis takes place the more progress this athlete can achieve. Also, we can predict the possible diseases that can infect an athlete.

1.2 Scope of this document

The system is developed to predict and recommend which sport is suitable for an athlete according to his genes variation as this prevents athletes from young ages to try practicing many sports and waste their time, money, and effort. Also, discover any genetic diseases that may impede that player and prevent him/her from practicing his/her favorite sport.

1.3 System Overview

The project contains two stages. the first phase is the classification where doctors insert the SNPs of the athlete or patient then get his points of strength as power, endurance, strength,and speed. The second phase is detecting possibility of a certain of disease.

1.4 System Scope

1.5 Business Context

Athletes will be able to know which kind of genes variations they have and as a result of that analysis, Doctors recommend the best sport that suits these athletes. Also from the DNA analysis doctors will be able to know if an athlete has a genetic disease that could affect his performance in a sport, So doctors will suggest the therapy for that athlete. ‘

2 Similar Systems

2.1 Academic

1. Muscle-Related Polymorphisms (MSTN rs1805086 and ACTN3 rs1815739) Are Not Associated with Exceptional Longevity in Japanese Centenarians[1]
   the researchers As life expectancy in developed countries rises, the number of elderlypeople and among them those showing exceptional longevity (EL, 100 years) steadily climbs. Amajo

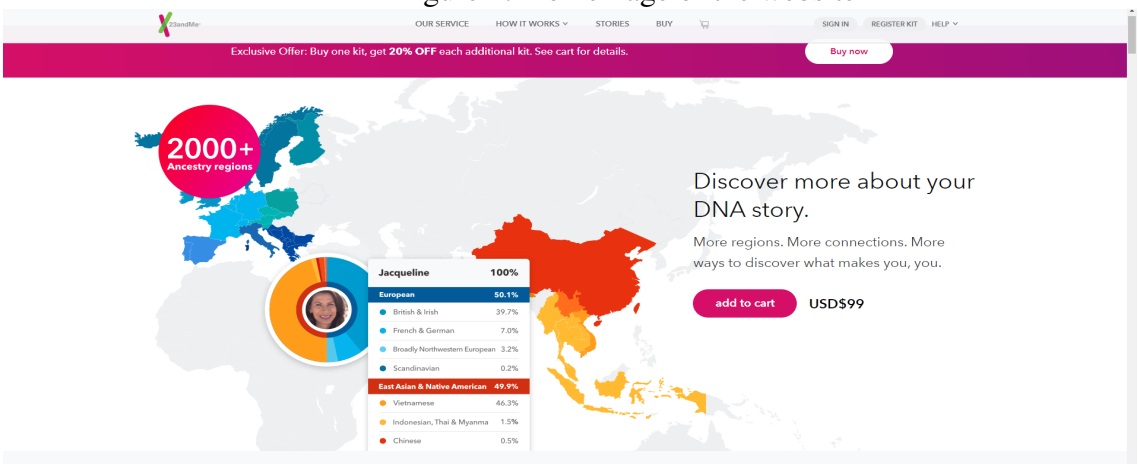
2. r problem associated with ageing is the gradual decline in those systems and organs thatdetermine physical fitness, notably the skeletal muscle tissue. Accelerated age loss of musclestrength (sarcopenia) is associated with higher mortality risk and genes potentially associated withpreservation of muscle mass and metabolic function could be potentially associated not only withhealthy ageing and disability risk.No differences were found between controls and centenarians inthe ACTN3 R577X allele/ genotype distribution. The ‘null’ XX genotype results in complete pro-tein deficiency and the Actn3-/- mouse model shows a shift in the properties of fast fibers towardsa more oxidative.

3. Genetics and the Elite Athlete: Our Understanding in 2020 [2]
   Elite Power and Sprint Perfor-mance: The ACTN3 Gene This gene encodes the structure of a sarcomeric protein found exclu-sively in type II muscle fbres (fast twitch myofbres), -actinin-3. These fibres are responsible forgenerating forces at high velocity during explosive or powerful activities. A single nucleotidepolymorphism (SNP) has been identifed which leads to a premature stop codon (X) rather than anarginine (R) at position 577. The R allele is advantageous in power sports and the RR genotype hasbeen found to be over-represented in elite power athletes [9]. On the other hand, the XX genotypeis associated with lower sprinting ability and muscle strength.Then at the end It is now clearly es-tablished that gene polymorphisms rarely act alone; the “single-gene-as-a-magic-bullet” ideologyis now largely discredited. To put it in simpler terms, just the presence .

2.2 Business Applications

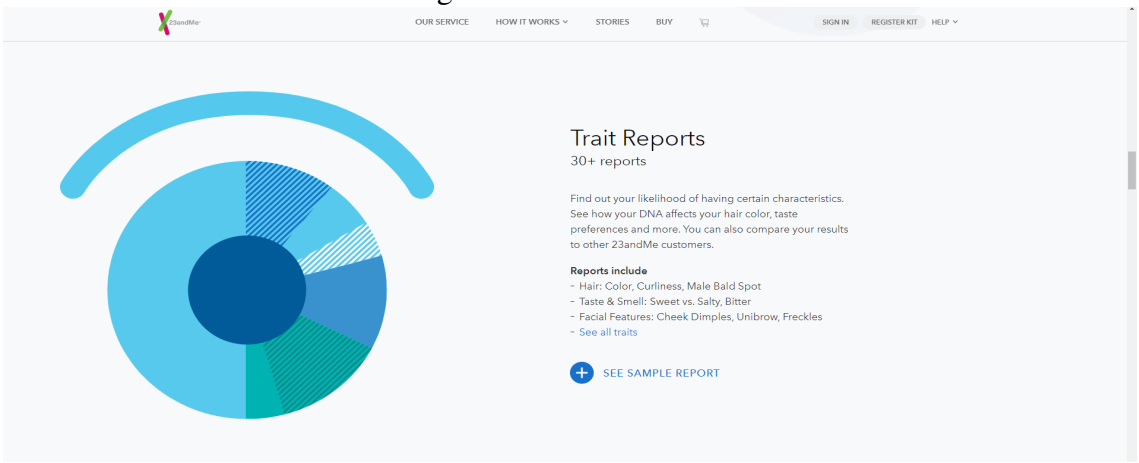
"23andMe" is a site that make a DNA test for users and it give them many results about their features , traits and family."https://www.23andme.com/en-int/"

Figure 1: Home Page of the website



23andMe's mission is to help people access, understand, and benefit from the human genome. Based in Sunnyvale, California, the company currently employs well over 500 people and ships its product to more than 50 countries worldwide. The company uses data to revolutionize health, wellness, and research. They want to improve healthcare. They want to prevent disease. And also to give individuals control over their health data. They want to dramatically accelerate the pace of research and also want to develop better drugs smarter and faster

Figure 2: one of 23andMe service



3 System Description

3.1 User Problem Statement

The main problem that faces the community is that many children play many sports without knowing if these sports would be useful for them or not or if they will show success in it. Many parents also force their children to practice and try many sports. There are a lot of mutation in genes that could change the the amino acid that could lead in a change in protein function.Also, there

are 2 types of muscle fibers: slow-twitch which is characterized as muscles with long contraction duration associated with endurance. And Fast-twitch which are characterized by fast muscle contractions of short duration. For example, a child can't practice swimming and karate/judo at the same time, Because swimming works on the extensibility of muscles while karate or judo works on the contractility of the muscles, So that child will have muscle issues in the long run and should stop practicing one of the 2 sports. Another huge problem we are facing is the diseases that can infect athletes such as hypertrophic cardiomyopathy (HCM). We have in our body active and inactive genes. Those inactive genes can be activated as a result of practicing unsuitable or specific sport.

3.2 User Objectives

The goal is to have a final product that shows the variation in genes so we can predict which trait is classified as normal or helpful mutation, and points of strength and recommend suitable sport. Also to have a final report for any possible genetic disease that could affect him/her.

3.3 User Characteristics

The expected users of the system will be admin or doctors. Therefore the users will know how to log in, sign up upload results, and make comments on them. By the time the user will adapt and interact with the system easily.

3.4 System Context

4 Functional Requirements

4.1 System Functions

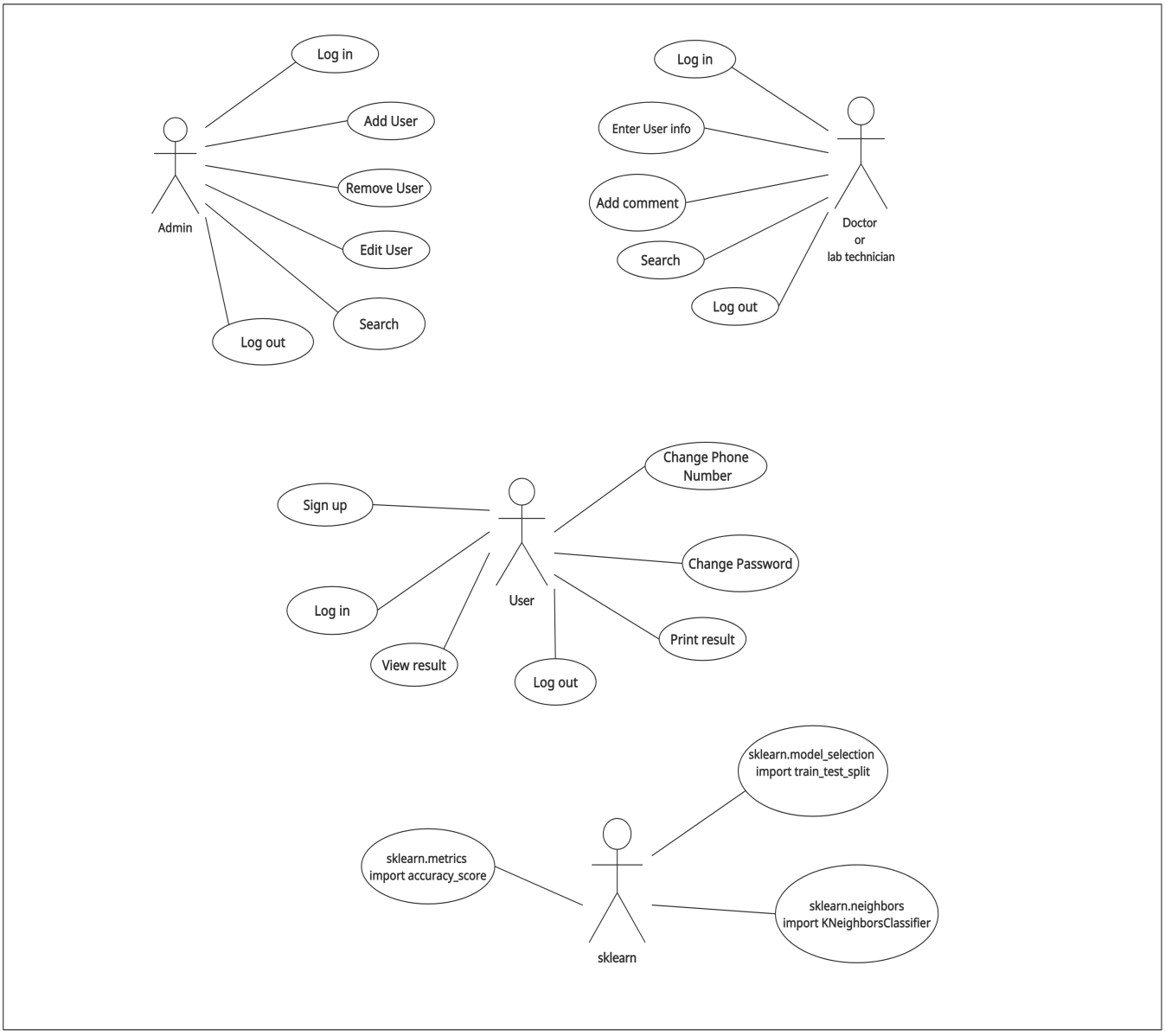


Figure 3: Use Case

1. The customer have to register first , so his information will be saved in the data base.
2. The customer can edit his phone number or username. The admin can add or create accounts to doctors or other admins.
3. The customer can access his results directly.
4. The admin shall update security rules by removing or editing accounts
5. The doctor shall be able to upload the DNA test.
6. The doctor shall be able to add comments on the final result.
7. The customer can download his report as a PDF file
8. The doctor and the admin can search about users by name.
9. The user shall be able to log out from his account

4.2 Detailed Functional Specification

Show the details of all functions shown in section 4.1. Describe each function in the following structure.

Table 2: Sign Up Function Description

Name	User Register
Code	FR01
Priority	Extreme
Critical	The user has to sign up to save his info in the database.
Description	This function takes the information from the user and insert it in the database
Input	Username , E-mail , Phone number, Password.
Output	Boolean (true for success, false otherwise)
Pre-condition	None
Post-condition	User information saved in the database
Dependency	None
Risk	The user may not enter the needed information, or enter it wrongly, so it must be validated.

Table 3: User log in Function Description

Name	User login
Code	FR02
Priority	Extreme
Critical	The user has to log in to see or download his result.
Description	This function takes the information from the user and check if it is in the data base
Input	E-mail , Password.
Output	Boolean (true if found, false otherwise)
Pre-condition	He has to sign up first.
Post-condition	The User access his profile.
Dependency	FR01
Risk	The user may login without signing up or enter the info wrongly , so it must be validated.

Table 4: Uploading Data Function Description

Name	Uploading file
Code	FR03
Priority	Extreme
Critical	The user (Doctor or lab technician) has to upload the DNA file.
Description	This function let the doctor upload the DNA file (csv file) and the system starts processing.
Input	CSV file
Output	If successful, the system starts processing the data. If not, the system asks the user to re-check the input.
Pre-condition	The user has to be logged in as a doctor.
Post-condition	The DNA results will be uploaded and the system starts processing
Dependency	FR02
Risk	The user may upload wrong file.

Table 5: view results Function Description

Name	View result
Code	FR04
Priority	Extreme
Critical	The user sees the results of the DNA test.
Description	The system takes the sample and starts processing the data, then shows the result to the user.
Input	none
Output	The user's points of strength or the possibility of some heart diseases
Pre-condition	The DNA samples have to be uploaded to the system
Post-condition	none
Dependency	FR03
Risk	none

Table 6: print results Function Description

Name	Print result
Code	FR05
Priority	Low
Critical	The system prints the result in a PDF document
Description	The system takes the sample and starts processing the data then shows the result to the user.
Input	none
Output	a PDF file of the results
Pre-condition	the user viewed his results
Post-condition	The user can download his result as a PDF file
Dependency	FR03
Risk	none

Table 7: Add User Function Description

Name	Add Comment
Code	FR06
Priority	high
Critical	Adding a new user by the admin
Description	This function is to add a user in database by the Admin
Input	User information
Output	Boolean (true for success, false otherwise)
Pre-condition	validate data entered
Post-condition	User information saved in database
Dependency	None
Risk	The admin may not enter the needed info , or enter it wrongly

Table 8: Add User Function Description

Name	Add User
Code	FR06
Priority	Extreme
Critical	Adding a new user by the admin
Description	This function is to add a user in database by the Admin
Input	User information
Output	Boolean (true for success, false otherwise)
Pre-condition	validate data entered
Post-condition	User information saved in database
Dependency	None
Risk	The admin may not enter the needed info , or enter it wrongly

Table 9: Edit User Function Description

Name	Edit User
Code	FR07
Priority	high
Critical	edit user info by the admin
Description	This function enables the admin to edit user information
Input	Username , edited information
Output	Boolean (true for success, false otherwise)
Pre-condition	validate data entered
Post-condition	User information will be edited
Dependency	None
Risk	None

Table 10: remove User Function Description

Name	Delete User
Code	FR08
Priority	high
Critical	edit user info by the admin
Description	This function enables the admin to delete user
Input	Username
Output	Boolean (true for success, false otherwise)
Pre-condition	User is found
Post-condition	User will be deleted
Dependency	None
Risk	None

Table 11: encrypt password Function Description

Name	Encrypt Password
Code	FR09
Priority	Extreme
Critical	encrypt password so the admin will be unable to know it
Description	This function used to encrypt the password to keep it private.
Input	User Password
Output	Encrypted Password
Pre-condition	user enters a password
Post-condition	The password will be added in Data base encrypted
Dependency	FR01
Risk	None

Table 12: Decrypt password Function Description

Name	Decrypt Password
Code	FR10
Priority	high
Critical	return the password to the original form
Description	This function used to return the password to its original form
Input	Encrypted Password
Output	User Password
Pre-condition	Decrypted password
Post-condition	Original Password
Dependency	FR01
Risk	None

Table 13: Change password Function Description

Name	Change Password
Code	FR11
Priority	low
Critical	The user changes the old password to a new one
Description	This function used to change the password of the user
Input	old Password , new Password
Output	Boolean (true if changed , false otherwise)
Pre-condition	the old password is correct
Post-condition	the Password changed in the data base
Dependency	FR01
Risk	None

Table 14: Change password Function Description

Name	Change Phone number
Code	FR12
Priority	low
Critical	The user changes the old phone number to a new one
Description	This function used to change the phone number of the user
Input	old phone number , new phone number
Output	Boolean (true if changed , false otherwise)
Pre-condition	the old phone number is correct
Post-condition	the phone number changed in the data base
Dependency	FR01
Risk	None

Table 15: Search Function Description

Name	Search
Code	FR13
Priority	medium
Critical	The user (doctor or admin) search for another user the old password to a new one
Description	This function search for the entered username in the database
Input	Username
Output	the user profile if found
Pre-condition	the user searching for has an account already
Post-condition	none
Dependency	FR02
Risk	the user name is entered wrong

Table 16: Add comment Function Description

Name	add comment
Code	FR14
Priority	medium
Critical	The user (doctor or lab technician) can add comment on the result.
Description	This function let the doctor comment on the final output and add this comment in the report.
Input	text
Output	Boolean (true if added , false otherwise)
Pre-condition	the user searching for has an account already
Post-condition	none
Dependency	FR02
Risk	none

Table 17: User log out Function Description

Name	User logout
Code	FR15
Priority	high
Critical	The user can logout from his account
Description	This function enables the user to log out
Input	None
Output	Boolean (true if logged out, false otherwise)
Pre-condition	The user have to be logged in
Post-condition	The User is logged out
Dependency	FR02
Risk	none

5 Interface Requirements

The system is designed simply to be easily used by doctors and lab technicians who have lacked computer skills. Mainly Our users are of three types, admins who can add, edit and remove users directly from the database and the doctors or lab technicians who enter or upload the DNA samples and can add comments on the final results, and finally the customer who can access his result from our website. Our plan is to update our User Interface and make a website connected with our database.

5.0.1 GUI

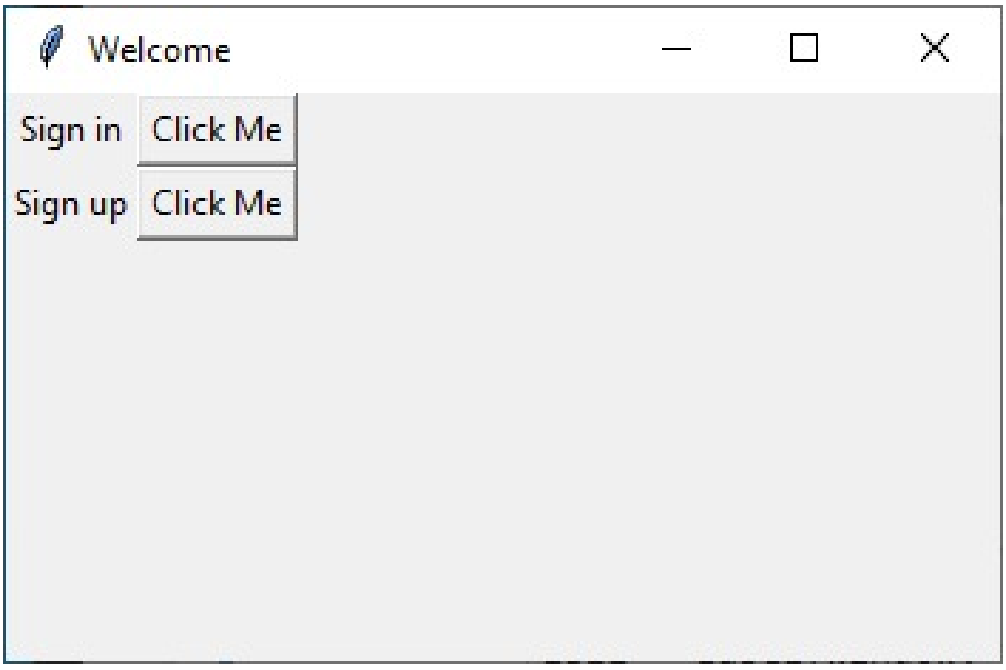


Figure 4: This is Our home page

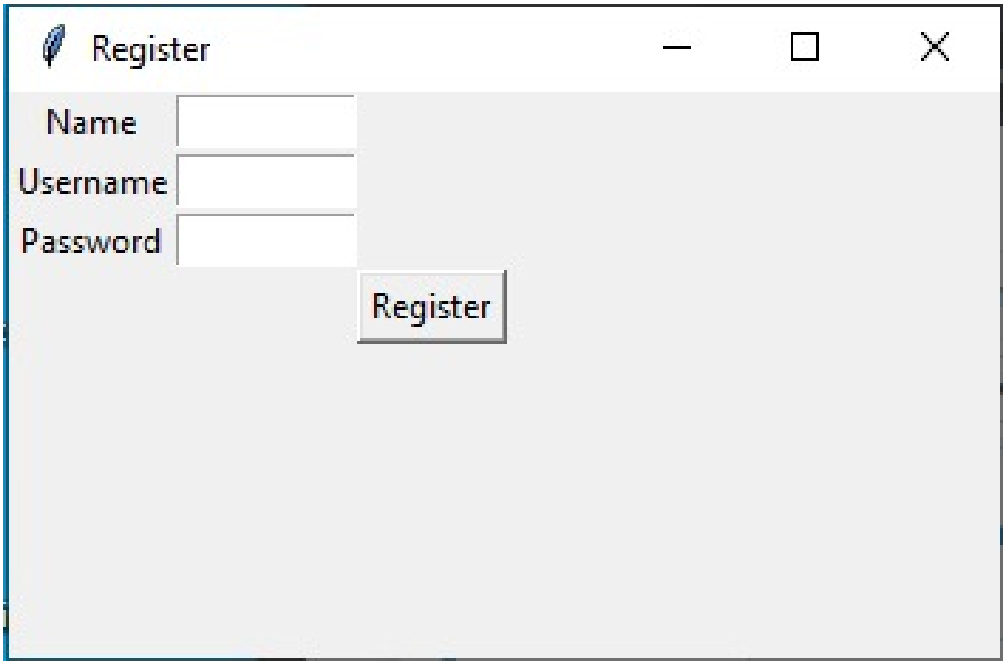


Figure 5: Register page

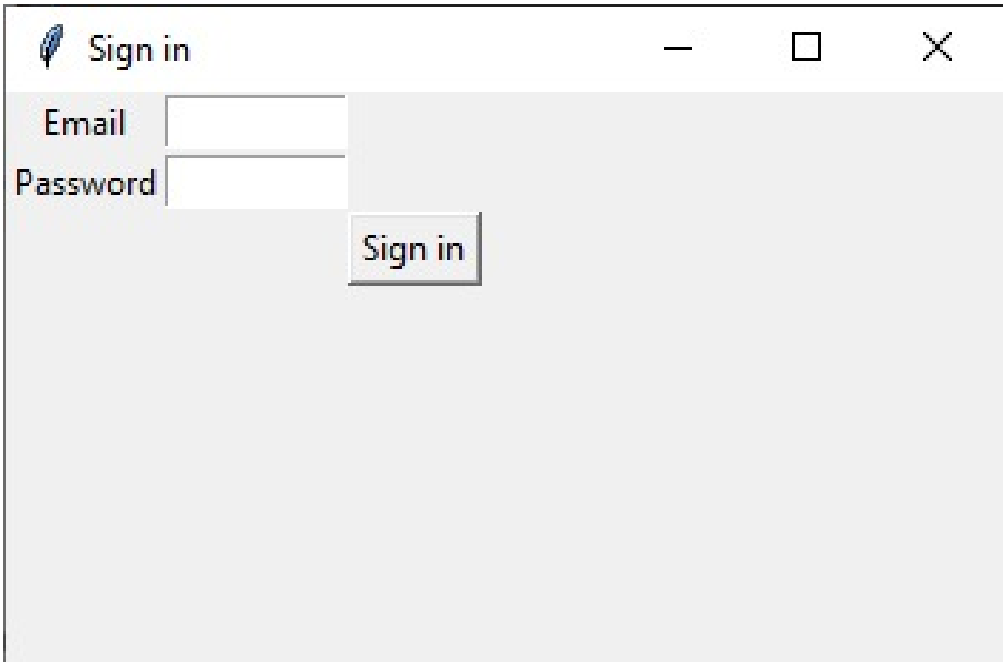


Figure 6: login page

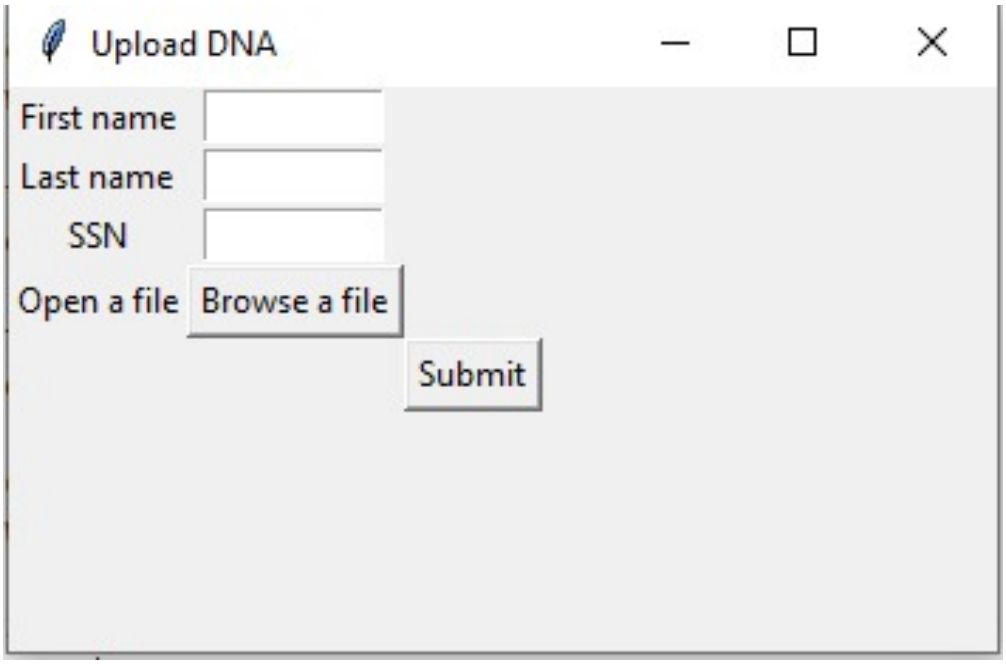


Figure 7: Doctor uploading DNA file

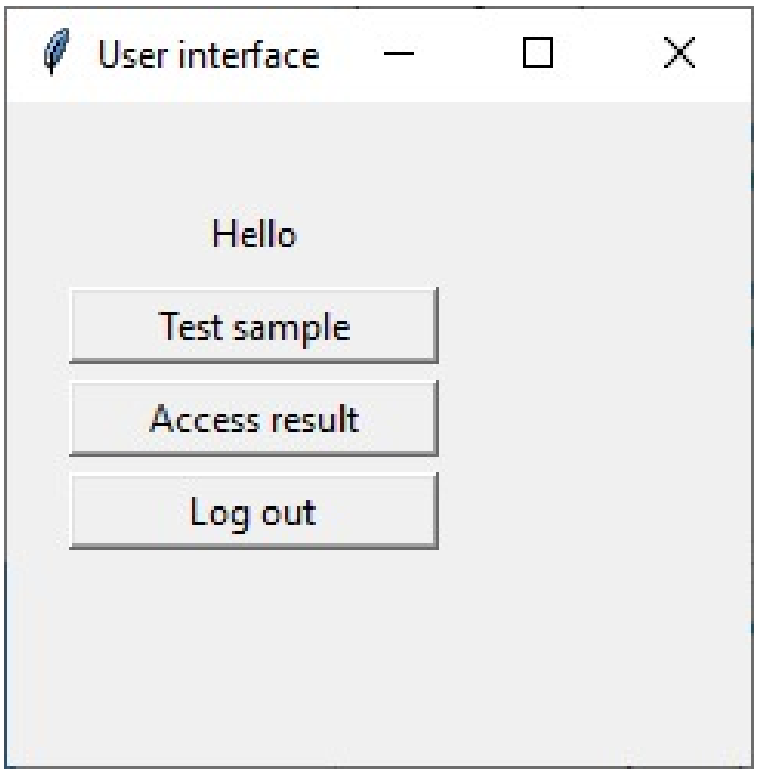


Figure 8: This is the User interface

5.0.2 CLI

none

5.1 Hardware Interfaces

We have no hardware interface

5.2 Communications Interfaces

We have no Communications interface

5.3 API

1. Pandas
2. sklearn
3. tkinte

6 Design Constraints

6.1 Standards Compliance

As most doctors and lab technicians aren't computer friends, we are willing to make the system easy to use with no complications.

6.2 Hardware Limitations

There are no hardware limitations yet

6.3 Other Constraints as appropriate

7 Non-functional Requirements

7.1 Security

Security is very important for the system so no one has the access to the athlete's data unless he has an account and his account is allowed to access any data.

7.2 Reliability

The system is reliable and ready to handle any issues. The time needed to analysis an athlete on the system has an medium speed to check where the DNA data is large.

7.3 Maintainability

the code is easy and simple and can be maintained in anytime.

7.4 Portability

the system will be available in any windows on pc since the code is written by python.

7.5 Extensibility

We are looking forward to add the Nutrition analysis for an athlete to recommend for him the best diet that suits his/her body.

8 Data Design

8.1 Data Description

8.2 Database design description

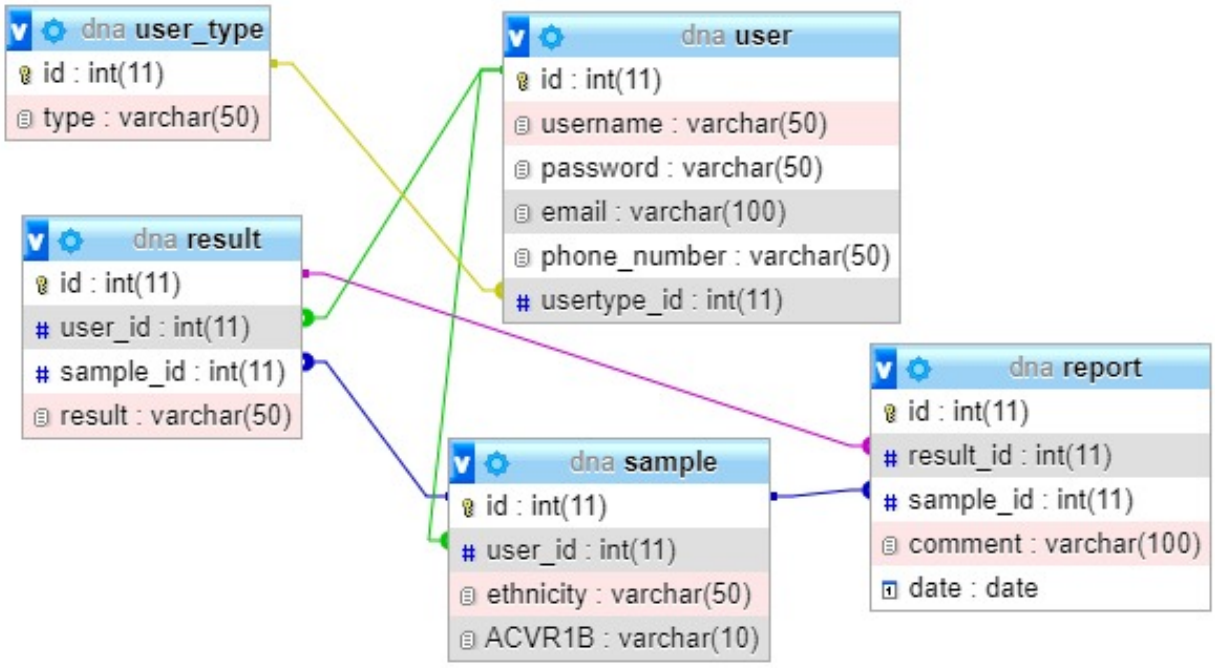


Figure 9: Database schema diagram

9 Preliminary Object-Oriented Domain Analysis

9.1 Inheritance Relationships

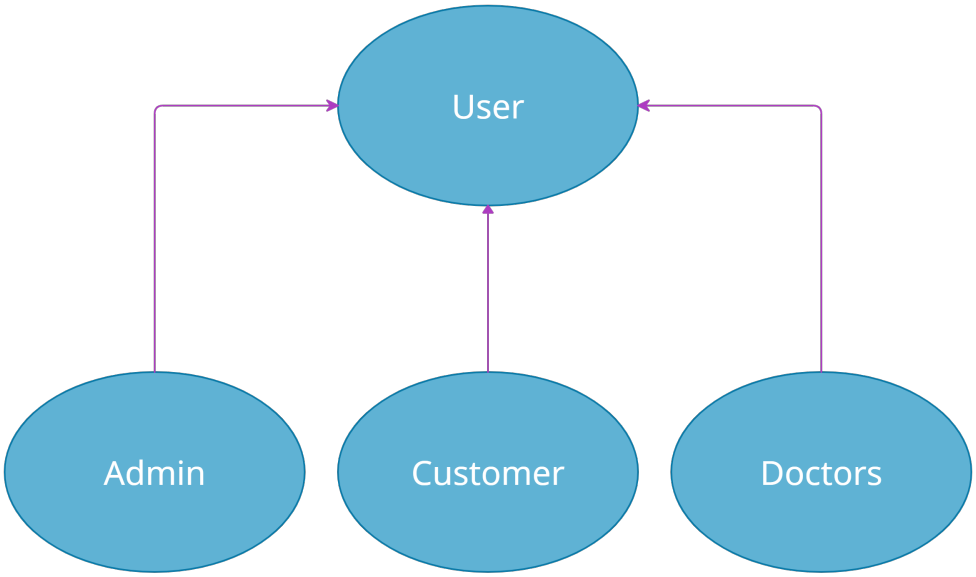


Figure 10: Database schema diagram

9.2 Class descriptions

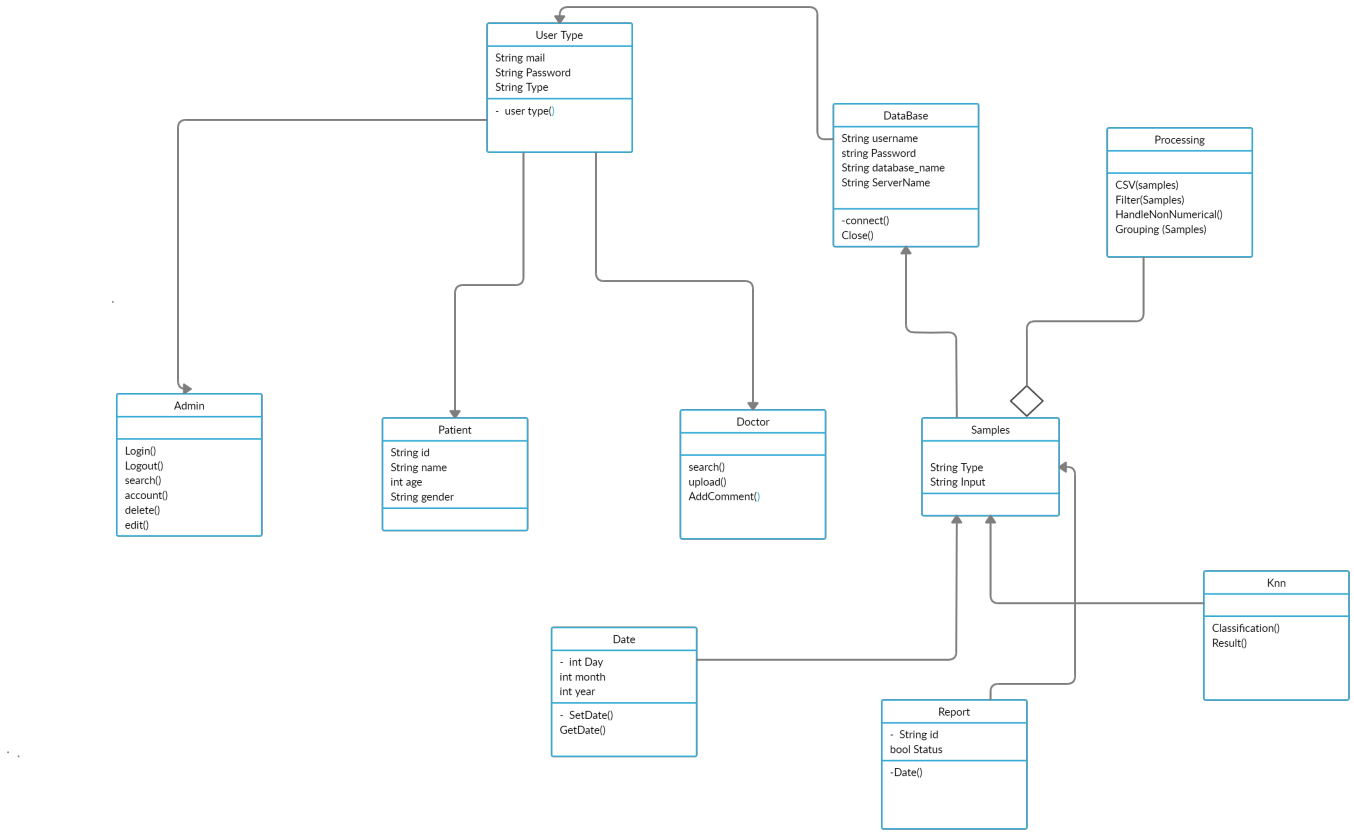


Figure 11: Class Diagram

9.2.1 Admin

1. Class Name: Admin
2. Super Classes: UserType
3. Sub Classes:N/A
4. Purpose: this class holds all functionalities for Admin
5. Collaborations: N/A
6. Attributes: N/A
7. Operations: add,delete,edit doctor and patients

9.2.2 Doctor

1. Class Name: Doctors
2. Super Classes: UserType
3. Sub Classes:N/A
4. Purpose: this class holds all functionalities for Doctor
5. Collaborations: N/A
6. Attributes: N/A
7. Operations: add comment on the analysis,search,upload analysis

9.2.3 Patient

1. Class Name: Patient
2.
3. Super Classes: UserType
4. Sub Classes:N/A
5. Purpose: this class is the holds all information about a any patient.
6. Collaborations: N/A
7. Attributes: id,name,age,gender
8. Operations: read analysis

9.2.4 Samples

1. Class Name: Samples
2. Super Classes: N/A
3. Sub Classes:N/A
4. Purpose: this class is the holds all information about a DNA Sample.
5. Collaborations: N/A
6. Attributes: type,input
7. Operations: N/A

9.2.5 Report

1. Class Name: Report
2. Super Classes: N/A
3. Sub Classes:N/A
4. Purpose: this class is the holds all information about DNA sample report.
5. Collaborations: N/A
6. Attributes: id,status
7. Operations: date

9.2.6 DataBase

1. Class Name: Database
2. Super Classes: N/A
3. Sub Classes:N/A
4. Purpose: this class is the holds all information about the DataBase.
5. Collaborations: N/A
6. Attributes: username,password,databasename ,serverName
7. Operations: connect,close

9.2.7 Processing

1. Class Name: Processing
2. Super Classes: N/A
3. Sub Classes:N/A
4. Purpose: this class is the holds all information about the processing operations.
5. Collaborations: N/A
6. Attributes: N/A
7. Operations: CSV,filter,remove,grouping Samples.

10 Operational Scenarios

As a customer, firstly he has to sign up to create an account and save his data in the database. If he signed up already, he has to enter his email and password correctly to access his profile, if not the system will notify him to re-enter his info. After that, he can log into the website and access his result, change his password or his phone number. As an admin, he will already have an account, so he will log in directly into the system. He also can add, delete, or edit users can search about any user. As doctors or lab technicians, they are added to the system by the admin. They can log in directly, but if they aren't added by the admin, they can't access their profile. They also upload the DNA file to the system to start processing, they can add comments on the final result.

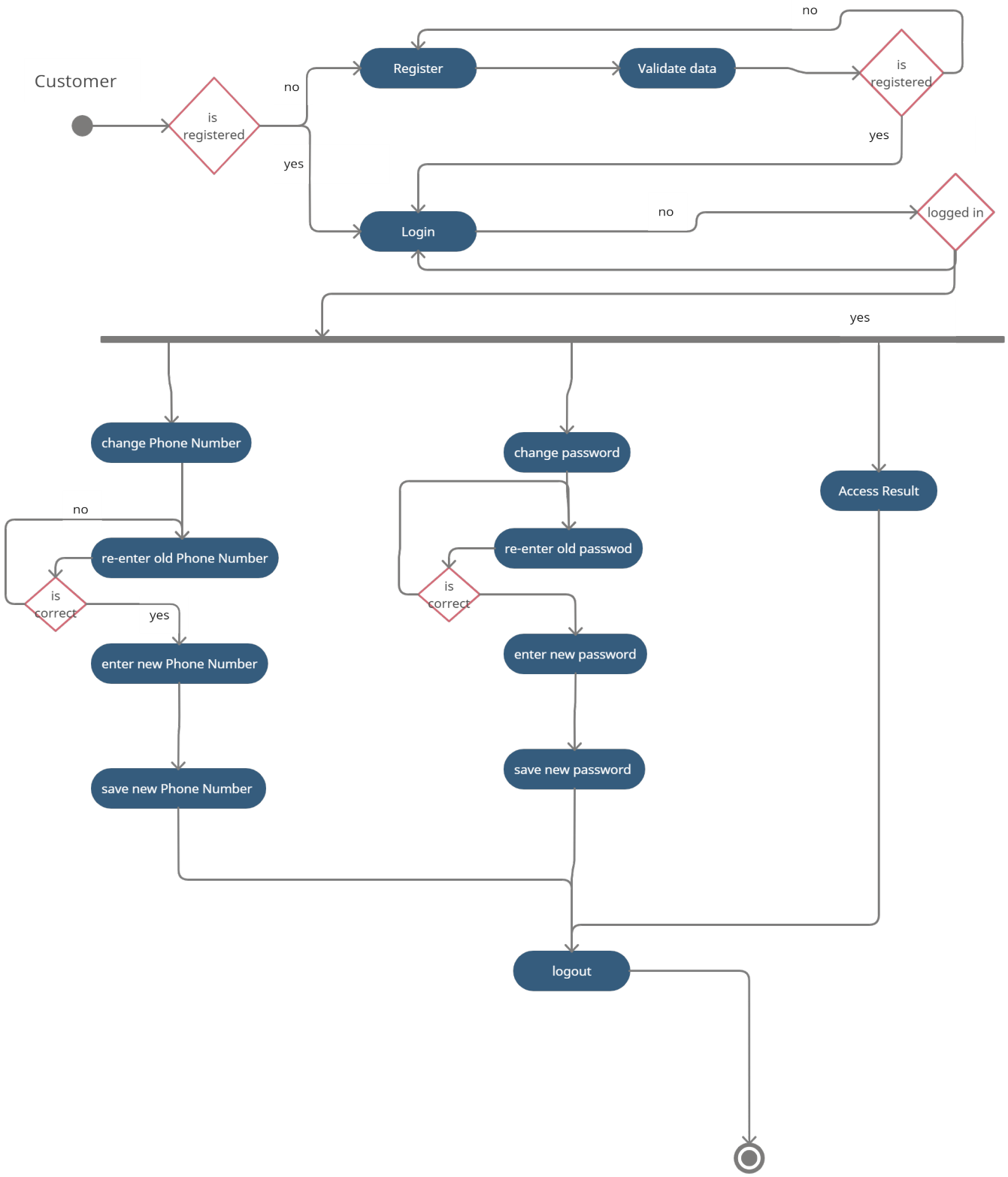


Figure 12: Customer Activity Diagram

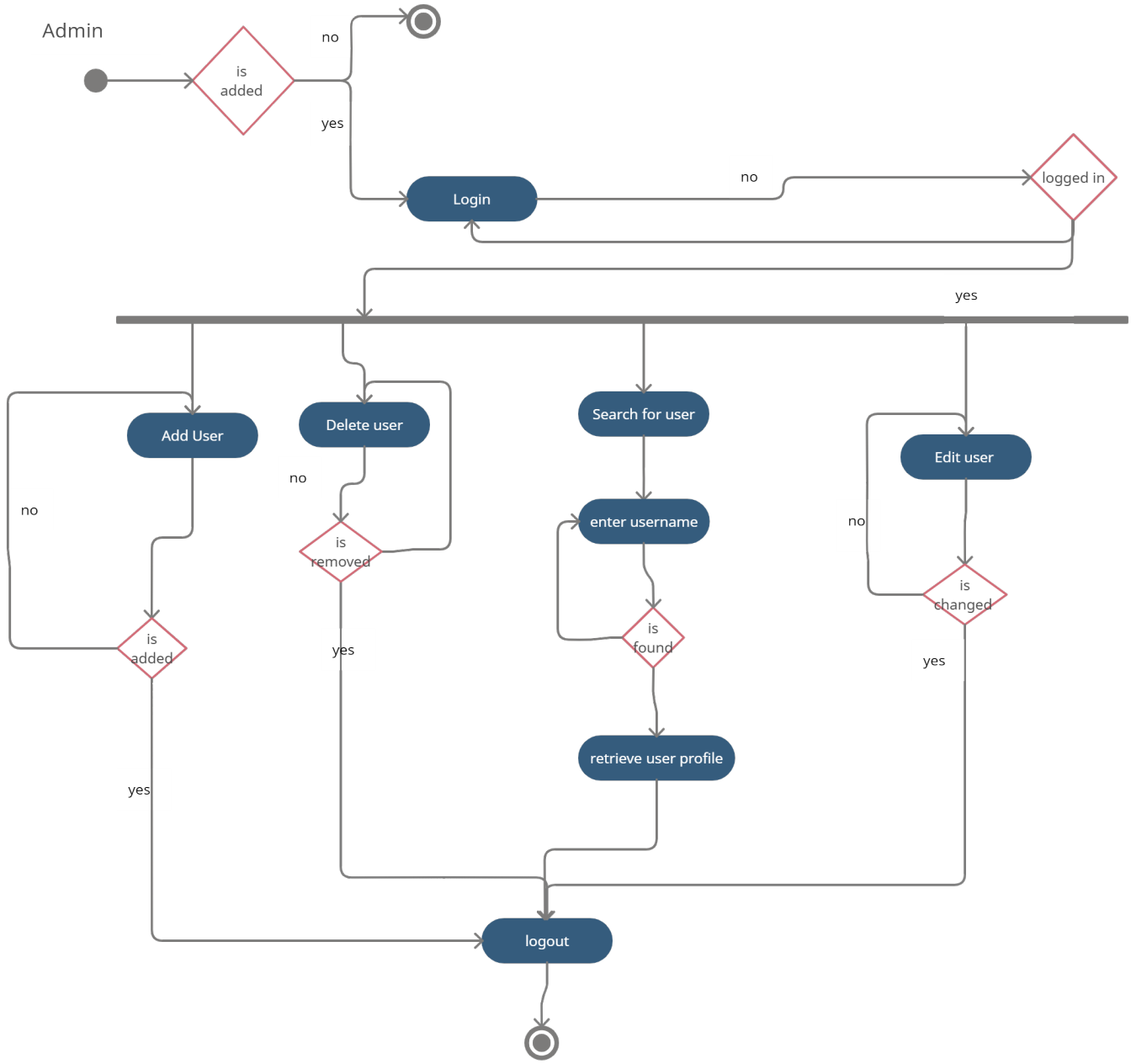


Figure 13: Admin Activity Diagram

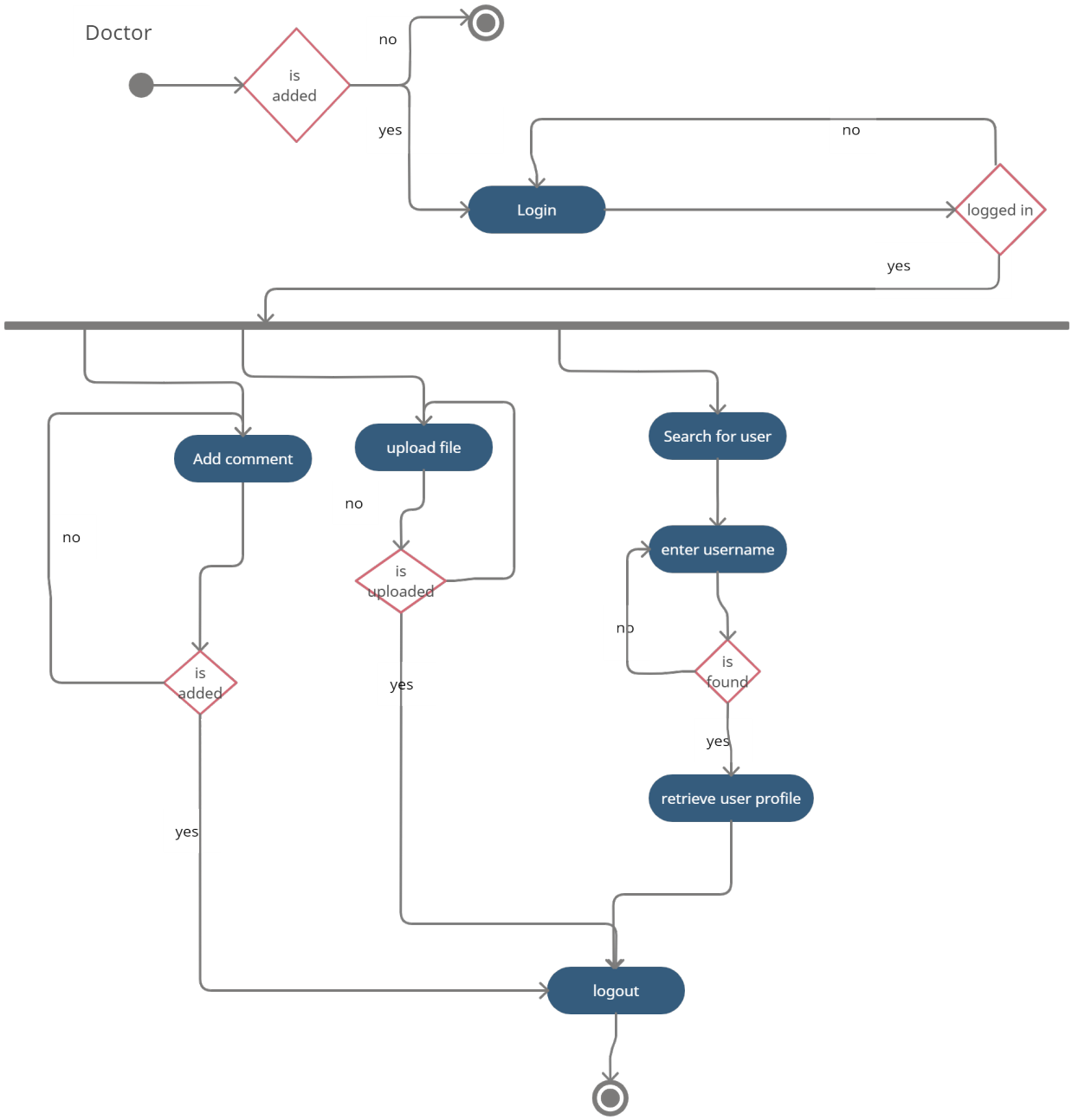


Figure 14: Doctor Activity Diagram

11 Project Plan

Task	From	To	PERCENT COMPLETE
Discussion about the ideas	25-Jul	14-Sep	100%
Researches about the idea	05-Aug	01-Oct	100%
collecting of dataset	03-Sep	11-Oct	80%
Writing proposal	28-Sep	22-Oct	100%
Implementation	30-Sep	22-Oct	10%
Proposal Phase 1	24-Oct		100%
Proposal presentations	27-Oct		
collecting of dataset	30-Oct	10-Nov	20%
Writing SRS	10-Nov	14-Dec	100%
Implementation	01-Nov	14-Dec	35%
Pesentation of SRS	14-Dec	21-Dec	
Writing SDD	25-Dec	14-Feb	100%
Implementation	01-Jan	14-Feb	65%
Presentation of SDD	15-Feb	21-Feb	
Implementation	10-Mar	20-Apr	80%
Writing paper	29-Feb	02-May	100%
Deliver Paper			
Testing			
Implementation	21-Apr	21-May	90%
Implementation			100%
Thesis	01-May	15-Jun	100%
Final Presentation			

Figure 15: This is Our plan

12 Appendices

12.1 Definitions, Acronyms, Abbreviations

12.2 Supportive Documents

13 References

References

[1] Noriyuki Fuku, Rafael Alis, Thomas Yvert, et al. “Muscle-related polymorphisms (mstn rs1805086 and actn3 rs1815739) are not associated with exceptional longevity in Japanese centenarians”. In: *PLoS One* 11.11 (2016), e0166605.

[2] Rakesh John, Mandeep Singh Dhillon, and Sidak Dhillon. “Genetics and the Elite Athlete: Our Understanding in 2020”. In: *Indian Journal of Orthopaedics* (2020), pp. 1–8.