Faculty of Computers
& Artificial Intelligence

Benha University

Oncology Center Services and Prediction Techniques to Assist in Making Right Decision.

A senior project submitted in partial fulfillment of the requirements for the degree of Bachelor of Computers and Artificial Intelligence.

Medical Informatics Departement,

Project Team

1- Alaa Walid Mohamed Alshitany
2- Eman Tarek Taha Ali
3- Basma Talaat Mahmoud Mohamed
4- Abdelrahman Ahmed Abdelhady
5- Yara Ahmed Mohamed Swisa

Under Supervision of

Prof. Dr.Elsayed Badr

Eng.Mohamed Alhakem

Benha. January 2023

DECLARATION

We hereby certify that this material, which we now submit for assessment on the program of study leading to the award of Bachelor of Computers and Artificial Intelligence. we have exercised reasonable care to ensure that the work is original and does not to the best of our knowledge breach any law of copyright and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of our work.

Signed: _____

Date: 15, Feb 2023.

ABSTRACT

Cancer is any disease among many diseases characterized by the development of abnormal cells that divide uncontrollably and is the leading cause of death for 10 million people, or approximately one death in every 6 deaths. Accurate prediction of response to cancer drugs is difficult due to uncertainty of drug efficacy and heterogeneity of cancer patients. Due to It is one of the most dangerous diseases that affect humans, so highlighting tumor's centers and paying attention to solving existing problems is important and necessary, and this is what we try to do in our project. We try to organize the process of making appointments and help doctors to choose the right drug efficiently using the deep learning model.

We have integrated 2 public databases in our study: GDSC and CCLE In our study, we operate on 256 genes and 238 drugs. GDSC database provides us IC50 values for large-scale drug screening data, when IC50 value is small, it is evidence that the effectiveness of the drug is high. CCLE database provide us Genomic, transcriptomic and epigenomic data of more than a thousand cancer cell lines. For the three data omics, we focused on gene mutation data, gene expression data, and DNA methylation data. we can get omics data by using microarray.

The goal of the model is to determine if the drug is effective for the cancer patient or not depending on cancer cell profiles (genomic mutation, gene expression and DNA methylation data). In the model we use CNN and GCN.

The result is the process become more accurate for doctors to determine if the drug will be efficient to a particular patient or not and the effectiveness of the drug. Through the android application we are developing, doctors can find it easily for showing their appointments in an organized manner and use the newest methods for diagnosis patient and recognize the disease. Also, patient can make an appointment online and choose which services he needs such as for test, rays or for diagnosis and medicine. Regarding the model, first, we used train_test_split to split the data and K-fold. Second, we used KNN with GCN, decision tree with GCN, random forest with GCN. The accuracy of Classification is 87.66 and regression is 95.17.

Table Of Contents

List of Figures……..……...…ii

Chapter One: Introduction And Background 1

Introduction 1

Problem Definition 2

Problem Solution 3

Chapter Two: Analysis And Design 4

Introduction 4

Business Requirements Identification 4

User Functional Requirements 4

System Functional Requirements 5

System Non-Functional Requirements 6

System Design 8

Use Case Diagram 8

Dfd Diagrams 9

Dfd Level 0 (Context Diagram) 9

Dfd Level 1 10

Dfd Level 2 11

Activity Diagram 14

Doctor 14

Patient 15

Sequence Diagram 16

Patient 16

Doctor 17

Er Diagram 18

Mapping 18

Class Diagram 19

GUI 20

Chapter Three: Model 26

Data Set 26

Model 27

CNN 30

GCN 32

Chapter Four: Results 34

Conclusion 35

Refernces 36

LIST OF FIGURES

Figure 2_1: Use Case Diagram…….…..8
Figure 2_2: Context Diagram…….....9
Figure 2_3:DFD Level1…….....10
Figure 2_4:DFD Level2…….....11
Figure 2_5:DFD Level2…….....12
Figure 2_6:DFD Level2…….....13
Figure 2_7: Doctor's Activity Diagram …….....14
Figure 2_8: Patient's Activity Diagram …….....15
Figure 2_9: Patient's Sequence Diagram ……...…….....16
Figure 2_10: Doctor's Sequence Diagram ……...…….....17
Figure 2_11:ER Diagram…….....18
Figure 2_12:Mapping…….....18
Figure 2_13: Class Diagram…….....19
Figure 3_1: Genomic Expression……..……..26
Figure 3_2: Genomic Mutation……...……..…..27
Figure 3_3: Genomic Methylation…….……...……..27
Figure 3_4: Model Farmwork…..……..……..28
Figure 3_5: Drug Representation…..……...……..……..28
Figure 3_6: Omics Representation...…...……..……..29
Figure 3_7: Data Representation.……..……..29
Figure 3_8: CNN layers……..……..……..30
Figure 3_9: Convolution layer.……..……...…..30
Figure 3_10: pooling layer.……..……..……..……..…..31
Figure 3_11: FCI layer.……..……..……..…..……..…..31
Figure 3_12: GCN.……..……..……..……..…..32
Figure 3_13: convolution...……..……..……..……..…..32
Figure 3_14: convolution.……..……..……....……..…..33
Figure 3_15: pooling layer.……..……..……..……..…..33

Chapter One

1 INTRODUCTION AND BACKGROUND

1.1 INTRODUCTION

Oncology refers to a branch of medicine that deals with the study, treatment, diagnosis, and prevention of cancer. [1] Cancer refers to any disease among many diseases characterized by the development of abnormal cells that divide uncontrollably and can infiltrate and destroy normal body tissues [2]. Cancer often can spread throughout the body, and cancer is the second leading cause of death in the world. But survival rates are improving for many types of cancer thanks to improvements in cancer detection, treatment, and prevention.[3]

Cancer is caused by changes or mutations in the DNA within cells. The DNA inside a cell is grouped into many individual genes, each containing a set of instructions that tell the cell what functions to perform, as well as how to grow and divide. Errors in these instructions can stop the cell from performing its normal function and may allow the cell to become cancerous.[2]

Accurate prediction of response to cancer drugs is difficult due to uncertainty of drug efficacy and heterogeneity of cancer patients.[5] Therefore, designing new drugs with desirable efficacy for cancer patients is a major challenge. Therefore, shedding light on tumor centers and paying attention to solving existing problems is important and necessary. Here is the role of machine learning and software engineering that help in diagnosing and treatment.[4]

Machine learning techniques and software are used these days in diagnose and treat diseases such as cancer. Deep learning techniques also used in pharmacogenomics as to predict drug effectiveness or develop new drug.[6] We expect this software will help patients and doctors. Patients need this application to know all services the center offers from tests to rays and clinics also. they can also book appointments. Doctors can use the prediction model that uses multi omics data and drugs to assist doctors in diagnose and treatment.

1.2 PROBLEM DEFINITION

Cancer is one of the most dangerous diseases that affect humans. An estimated 20 million people have been diagnosed with cancer, and 10 million people have died from it. Therefore, shedding light on tumor centers and paying attention to solving existing problems is important and necessary.

- **problems specific to doctors:**

  -**Time:** is one of the most important problems that doctors face, as the doctor goes to the oncology center and there, he knows how many appointments he has.

  -**Diagnose & Drugs:** there are many doctors who are beginners and need a tool to help them through diagnose and writing medicine.

  -**Patient Record:** records of patient always saved in hospitals, so doctors cannot view them in any time.

- **problems specific to patients:**

  -**Booking:** patients always face problems with booking appointments as they need to visit the center for booking then another visit for diagnose and this consumes health and time.

  -**Services:** patients don't know all services that center provide such as medical tests, medical rays, or medical clinics.

- **problems specific to organization and management of the center:**

  -**Appointments:** In most healthcare facilities in Egypt, there are problems organizing and booking appointments. The existing scenario is that first, patients try to make appointment through telephone. if that is not available, then patients go to the center to book appointment and this appointment may be in the same day after many hours or may be after some days. In both, the patient wasted a lot of his time.

  - **Paper patient records:** most health care hospitals in Egypt follow paper system and this consumes resources such as place to save all the records, also more paper to cover all patients in all departments...etc.

1.3 PROBLEM SOLUTION

To solve the problems mentioned, we have provided some solutions to most of the problems, such as:

1) For each doctor, we provide them with a page with his appointments, so he goes to the center, knowing the number of patients and the time for each patient, so we have overcome the problem of time.

2) We introduced a deep learning model that predicts drug effectiveness for each patient instead of trying more than one type of drug that the patient may not respond to, and this will benefit junior doctors who are not experienced enough.

3) For patients, we try to organize and facilitate the process of booking as we provide online booking in our system and the patient can choose which service to book, so we have overcome the problem of booking.

4) Patients and Doctors can view the services of the oncology center, so can make tests, rays and take diagnose and medicine in one place and it saves much time, so we have overcome the problem of services and also choosing services type in the booking helps in organize the center.

5) We have a person called the admin, he is responsible for organizing the appointments, whether accepting, rejecting, or modifying the appointments. Also, he is responsible for digitalizing the paper records, so the records become accessible by doctors at any time.

6) As the patient records become digital meaning that they no longer been saved in hospital, they are saved on server and doctors and admins can access them at any time

Chapter Two

2 ANALYSIS AND DESIGN

2.1 INTRODUCTION

This chapter focuses on the analysis phase of the system development life cycle. The main objective of this chapter is to provide a concise idea of the system and the activities taken to develop the system. A detailed description of the functional and non-functional requirements of the system in addition to the user requirements is included. Also, an overall description of the system is included.

2.2 BUSINESS REQUIREMENTS IDENTIFICATION

Obtaining the business requirements of the system is an important stage of the system development life cycle. Using iterative model as a software development method made it possible for us to fix and enhance the requirements of the system as the project advanced from one phase to another. Interviews and meetings with the client were held to gain the complete understanding of the system specifications. The following illustrates the software requirements, the functional requirements, and nonfunctional requirements of the system.

2.2.1 User Functional Requirements

This section illustrates the user requirements of the system. The following requirements are suggested by the system developers themselves to be verified later by the client. For example, if there are three types of users for the Project System: Patients, Doctors, Admin.

- Patients can register to be able to have an email that helps in seeing our services.
- Patients can log in if they have an account.
- Patients can book an appointment.
- Patients can log out from his account.
- Patients can receive SMS with any change to their appointment.

- Doctors can also log in.
- Doctors can use prediction model.
- Doctors can view appointments.
- Admin can log in.
- Admin can add, delete, modify the data of doctors.
- Admin can add, delete, modify the data of appointments.
- Admin can organize doctor's time.

2.2.2 System Functional Requirements

1. The system should display a welcome screen containing the system name and logo.
2. The system should display an intro slider which displays information about different features within the application.
3. The system should provide a screen where each patient can sign up using his information such as (First Name, Last Name, Birthdate, Gender, Email, Address, Phone, Password, Password confirmation)
4. The system should allow each patient to log in using his email and password.
5. The system will enable patient to book an appointment using his name, phone, and email.
6. The system should send message to patient with his name, phone, date & time of the appointment.
7. The system will enable patient to view the center services and doctors available.
8. The system should provide an about us button for patients to know more about our center.
9. The system will log out from the patient profile if the patient clicks log out button. (Log out from account).
10. The system should allow each doctor to log in using his email and password. (Provided by the admin).
11. The system should allow doctor to view patients who have reservations with him.
12. The system will log out from the doctor profile if the doctor clicks log out button. (Log out from account).
13. The system should allow admin to log in using his code and password. (Provided by the developer).
14. The system allows admin to add, modify, delete doctor data.

15. The system allows admin to manage appointments of patients.
16. The system allows admin to organize doctor's time.
17. The system allows admin to make backup.
18. The system will log out from the admin profile if the admin clicks log out button. (Log out from account).

2.2.3 System Non-Functional Requirements

This section defines system attributes such as security, reliability, performance, maintainability, scalability, and usability. They serve as constraints or restrictions on the design of the system across the different backlogs.

- **Performance:**

  The responsiveness of the application and the time it takes it to finish a job are indicators of its performance. For instance, the initial patient, doctor, or administrator screen should load in no more than three seconds when an application first starts up. Additionally, it should be ensured that the software won't obstruct user input.

- **Scalability:**

  App should be able to adjust itself to higher usage or be able to manage more data as time progress. For instance, the app should be able to handle patients (personal, medical, and payment) without delay by optimizing how storage is done and retrieved.

- **Responsiveness:**

  The application must respond to input from the doctor and the patient or to any external interruption with the greatest priority and restore to the previous state. For an instance, when an app is interrupted by anything, it should be able to store its state and return to the same page that it was in before the interruption.

- **Usability:**

  Doctors and patients should be able to utilize the app without assistance from professionals or manuals, meaning that both parties should be able to quickly.

comprehend how it works. User experience is not good if it needs to be explained.

- **Reliability:**

The application must be trustworthy in order to handle patient data and payments; for example, when a user takes an essential action, it must be confirmed.

- **Security:**

All app data should be protected and encrypted with the absolute minimal security requirements in order to prevent it from both external and internal attacks.

As an illustration, all authentication tokens ought to be retained on local devices for comparison and access requires user authorization.

- **Availability:**

The user should be able to access your application on a common platform to install it and check for updates on a regular basis. give criticism.

2.3 SYSTEM DESIGN

In this section, we will discuss the design of our system such as:

2.3.1 USE CASE DIAGRAM

Use case diagram give us an overview on the system including the **actors**, in our system are (doctor, patient, and admin) and the **main process** that each actor can do.

Figure 2_1: use case diagram.

Doctor can login to the system, use the model to help him in his diagnose, view the appointments that with him, add medicine and diagnose to the patient and manage the data of patients. **Patient**: can login or register to the system and make appointment with doctor. **Admin:** can login and manage the appointments and doctors' data.

2.3.2 DFD DIAGRAMS

Data Flow Diagram: showing the entities that interact with it (patient, doctor, admin, and treatment) and represent the flow of the data between the system and each external entity. It has many levels such as level 0, level 1 and level 2.

2.3.2.1 DFD LEVEL 0 (CONTEXT DIAGRAM)

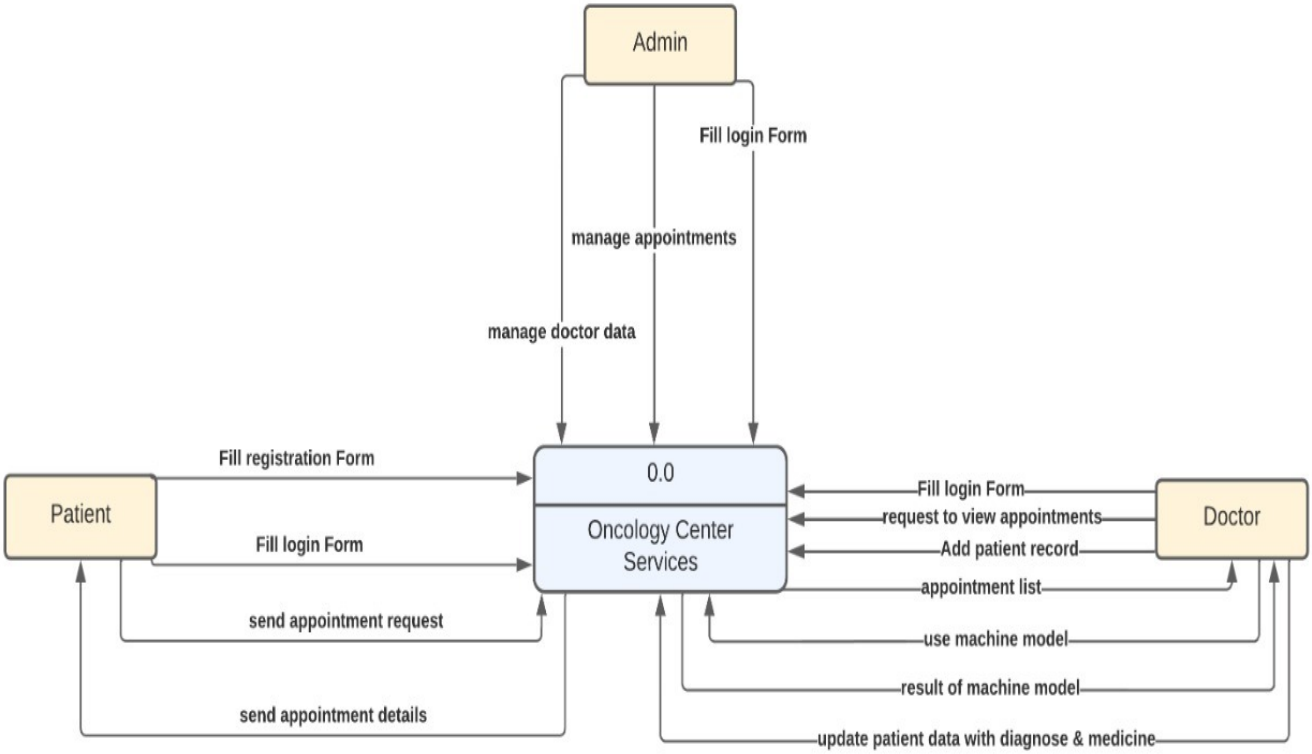


Figure 2_2: context diagram

Context diagram: patient can fill registration form, fill login form, and send appointment request and the system send it an appointment detail. Admin can fill login form, manage appointment, and manage the data of the doctor. Doctor can fill login form, request to view appointment, add patient record, use machine model, and update patient data with diagnose and medicine and the system send it an appointment list and result of machine model.

2.3.2.2 DFD Level 1

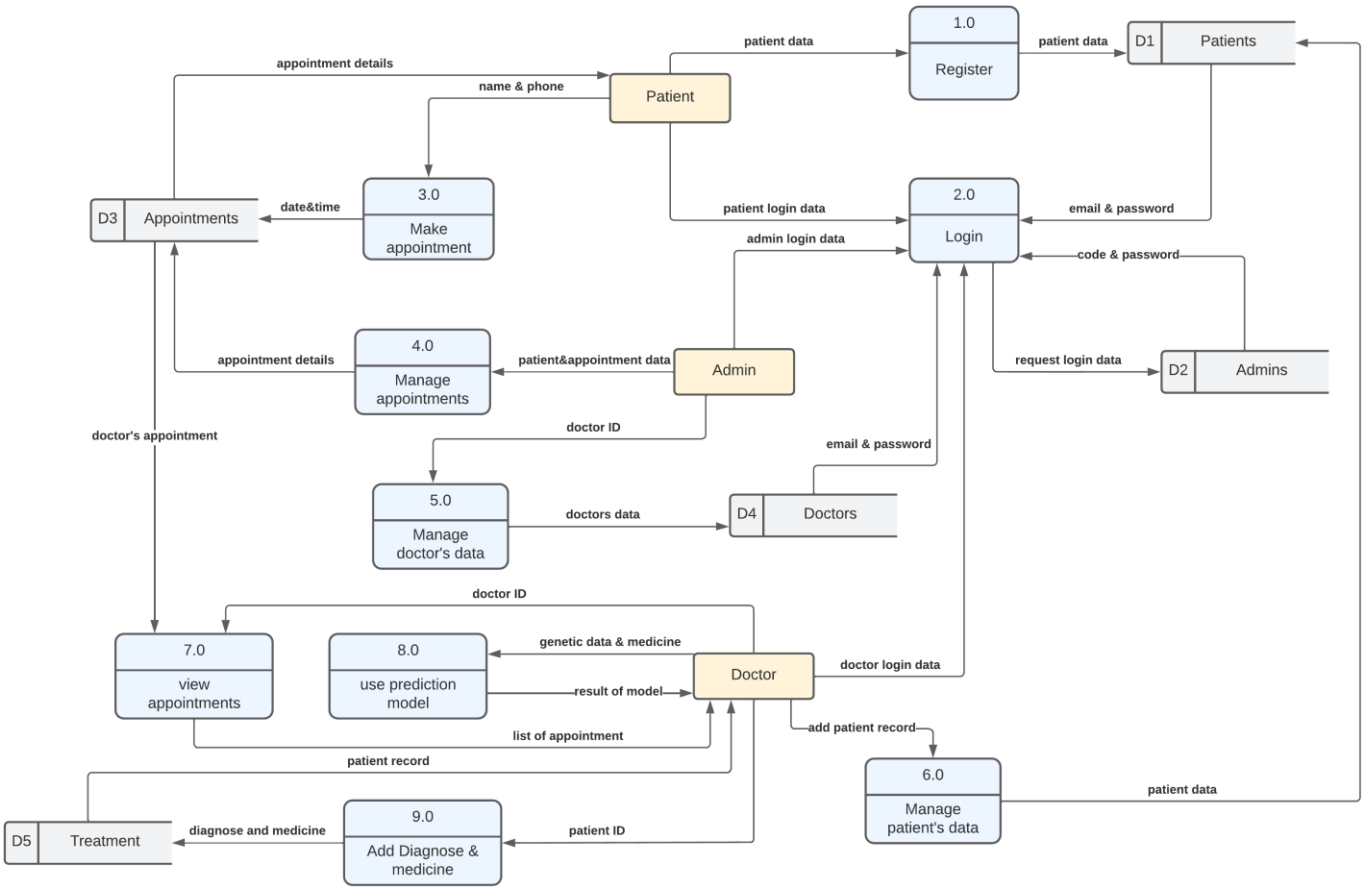


Figure 2_3: DFD level 1

DFD Level 1, they go into more detail than a context diagram. The single process node from the context diagram is broken down into sub-process, which include data store such as(patient, admin, doctor, and appointments) that store the data, process such as (register, login, manage patient's data, make appointment, manage appointment, manage doctor's data, view appointments, use prediction model, and add diagnoses and medicine), external entity such as(patient, admin, doctor), who does the process and also include arrows which represent the flow of the data among external entity, process, and data store.

2.3.2.3 DFD Level 2

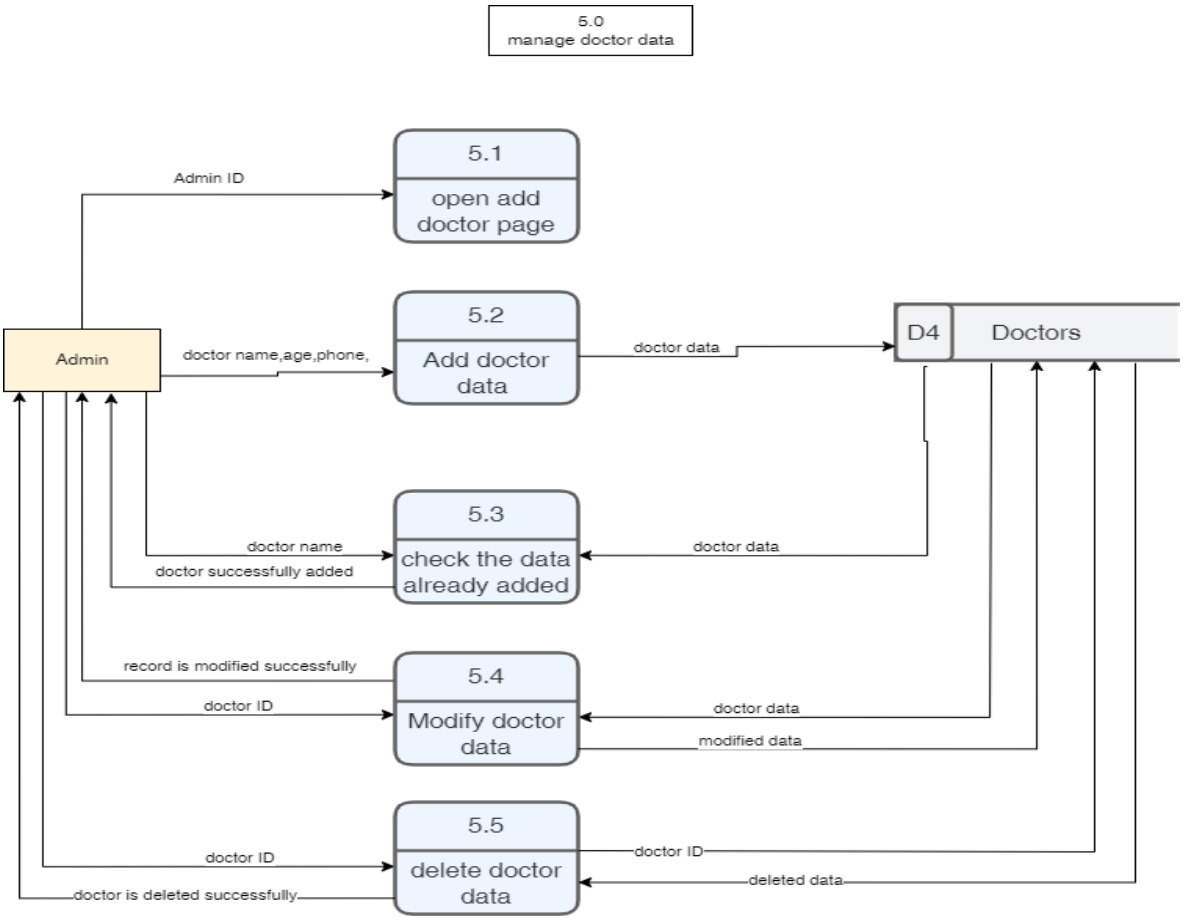


Figure 2_4: DFD level 2

DFD Level 2, goes one step deeper into parts of 1-level DFD. Here, we choose the process number (5.0) named (manage doctor data) which show how the data of the doctor being managing by admin (external entity) and store data into data store (doctor) and use arrows to explain the flow of the data among external entity, process, and data store.

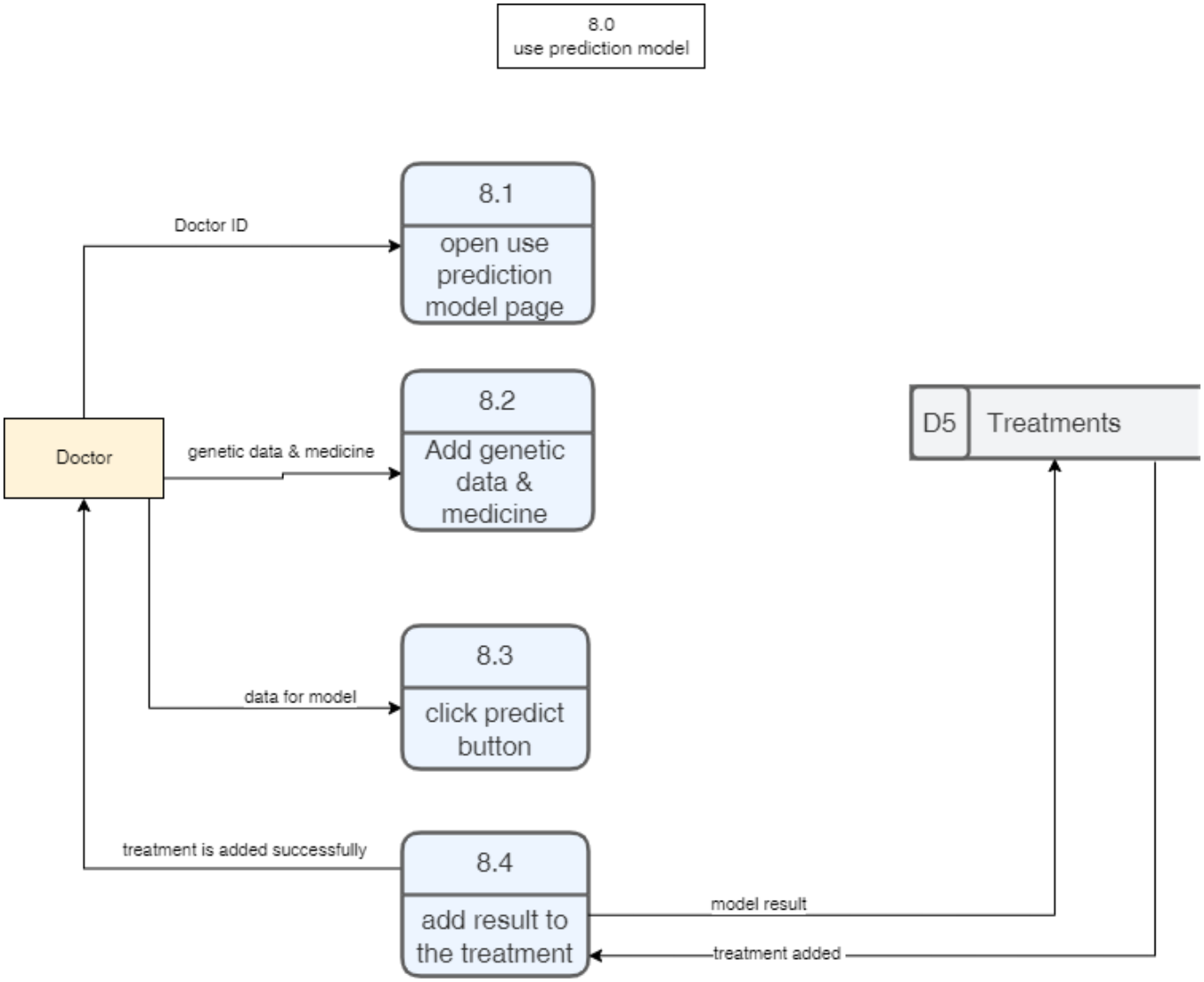


Figure 2_5: DFD level 2

DFD Level 2, goes one step deeper into parts of 1-level DFD. Here, we choose the process number (8.0) named (use prediction model) which show the process to use the prediction model by doctor (external entity) and store data into data store (treatments) and use arrows to explain the flow of the data among external entity, process, and data store.

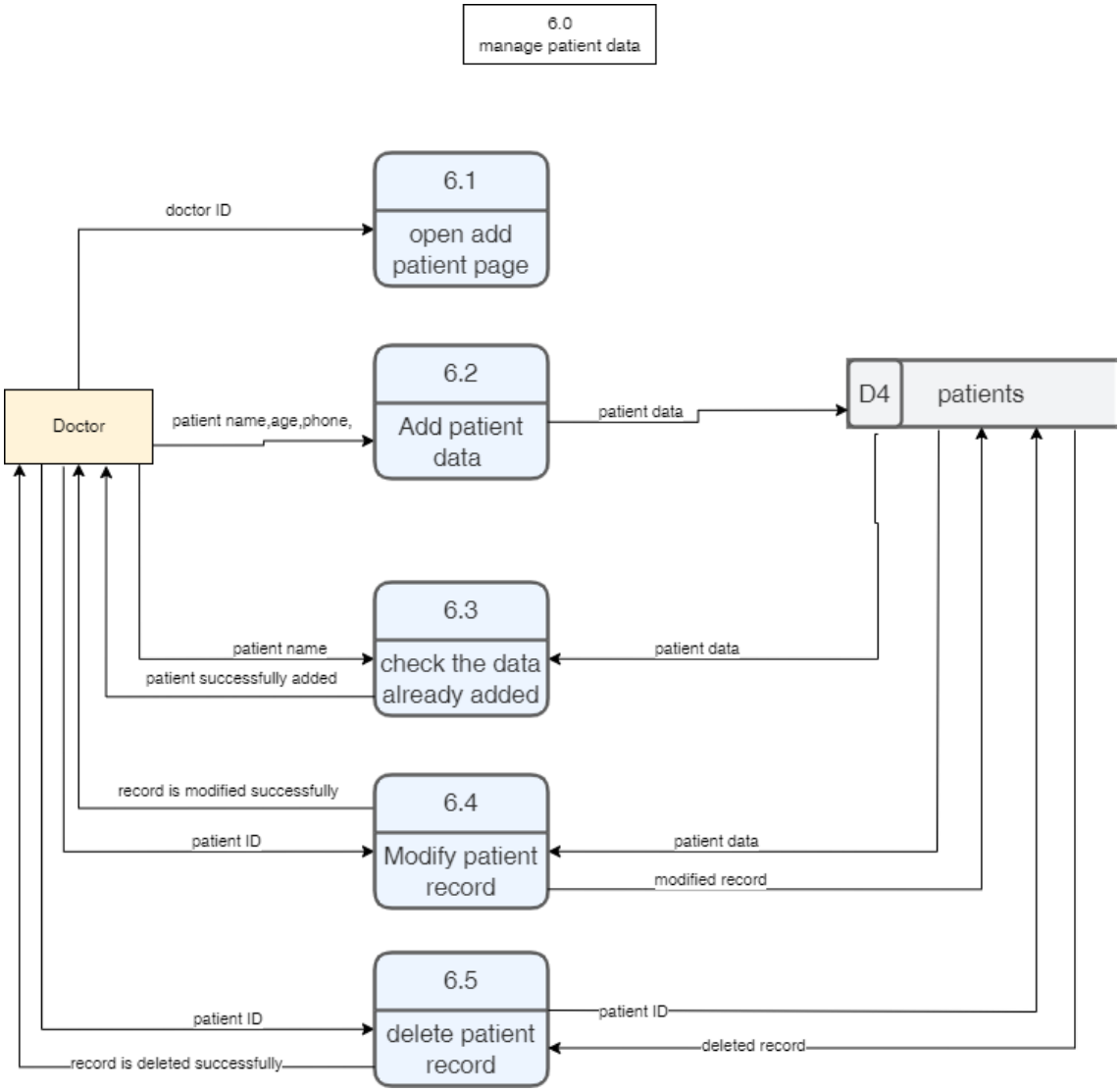


Figure 2_6: DFD level 2

DFD Level 2, goes one step deeper into parts of 1-level DFD. Here, we choose the process number (6.0) named (manage patient data) which show how the data of the patient being managed by doctor (external entity) and store data into data store (patients) .and use arrows to explain the flow of the data among external entity, process, and data store.

2.3.3 ACTIVITY DIAGRAM

2.3.3.1 DOCTOR

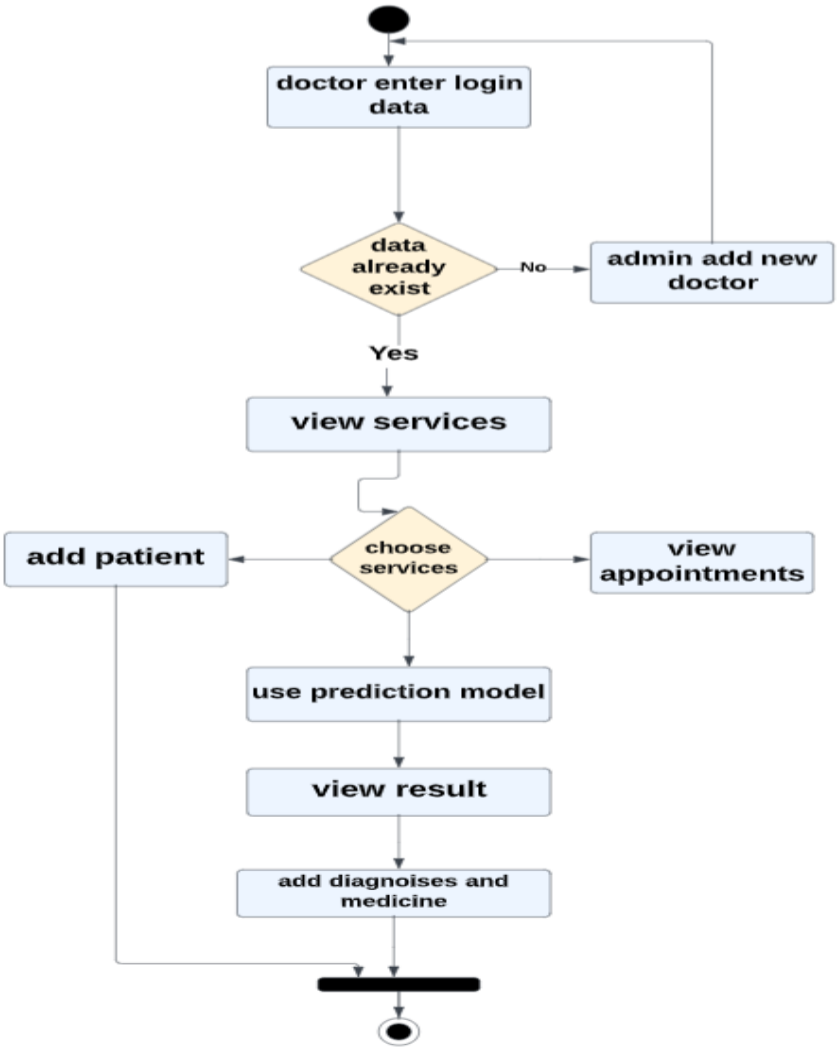


Figure 2_7: Doctor's Activity Diagram

The activity diagram of the doctor: explain to us what the doctor can do with the system. To login into the system: The system checks the login data if it is right or not. If the data is right then the doctor has access to his services as use the model to help him in his diagnose, view the appointments that with him, add medicine and diagnose to the patient and manage the data of patients. If the data is wrong, this means that this doctor isn't added to the system so the admin will add him and then login.

2.3.3.2 PATIENT

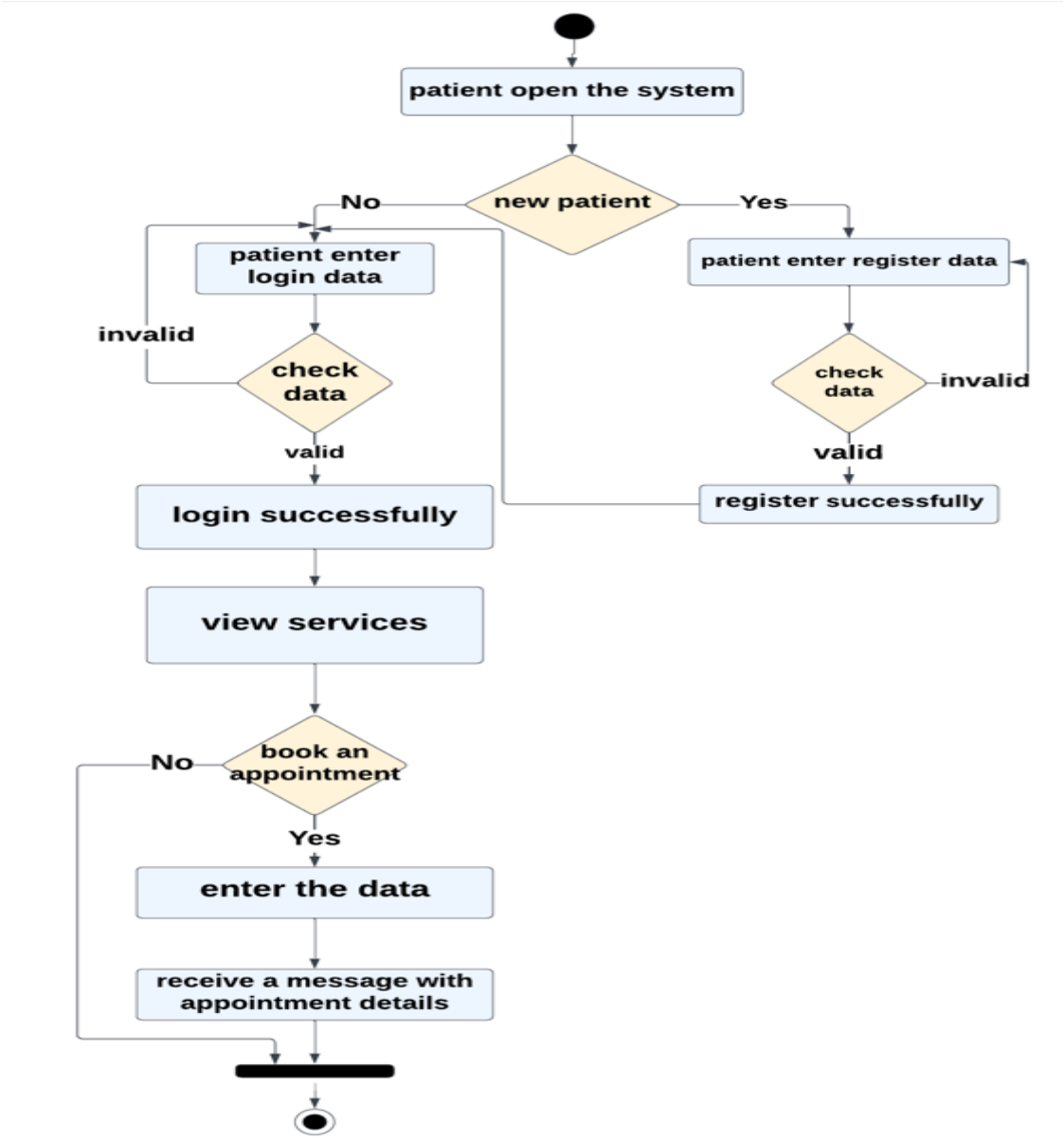


Figure 2_8: patient's Activity Diagram

The activity diagram of the patient: explain to us what the patient can do with the system. First patient must register to the app if it's his first time to use the app. Then patient can login if he entered login data (email and password) correctly. After successful login, patient can view services (including Rays, Tests, and Clinics) and also patient can request to book an appointment.

2.3.4 SEQUENCE DIAGRAM

2.3.4.1 PATIENT

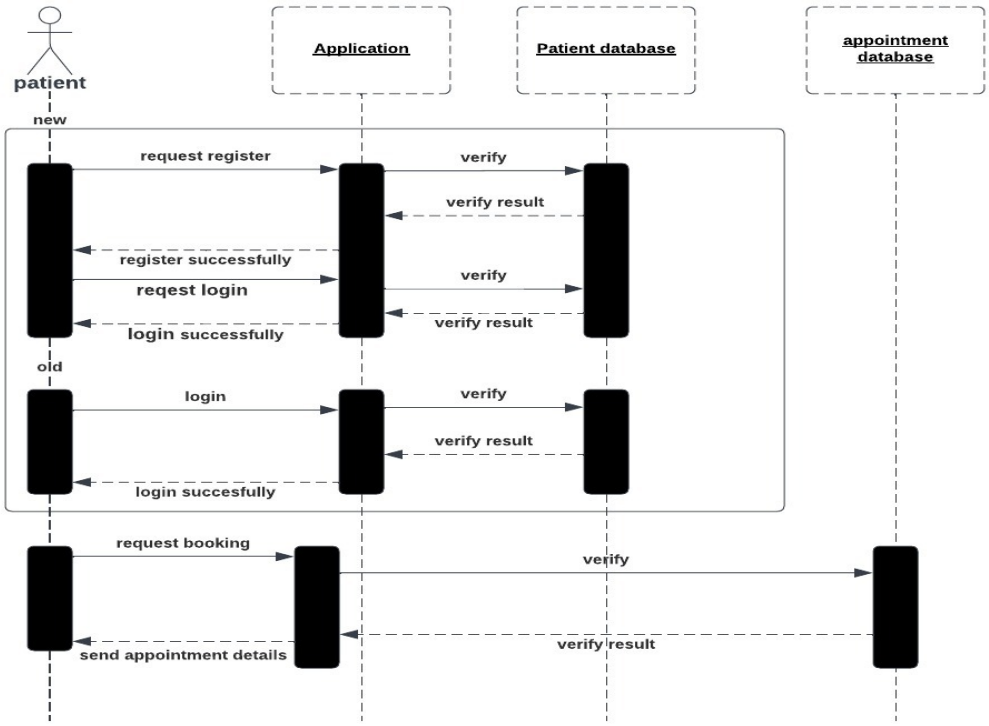


Figure 2_9: Patient's Sequence Diagram

Sequence diagram of the patient: explain to us the steps that the system makes after the patient act with the system and what the messages that will appear to the patient in each case. As the patient enter the data then the system checks this data if it is correct then print to him login successfully and then let him book appointment then check the data in appointment if right send to him the details of appointment as date and name of doctor.

2.3.4.2 DOCTOR

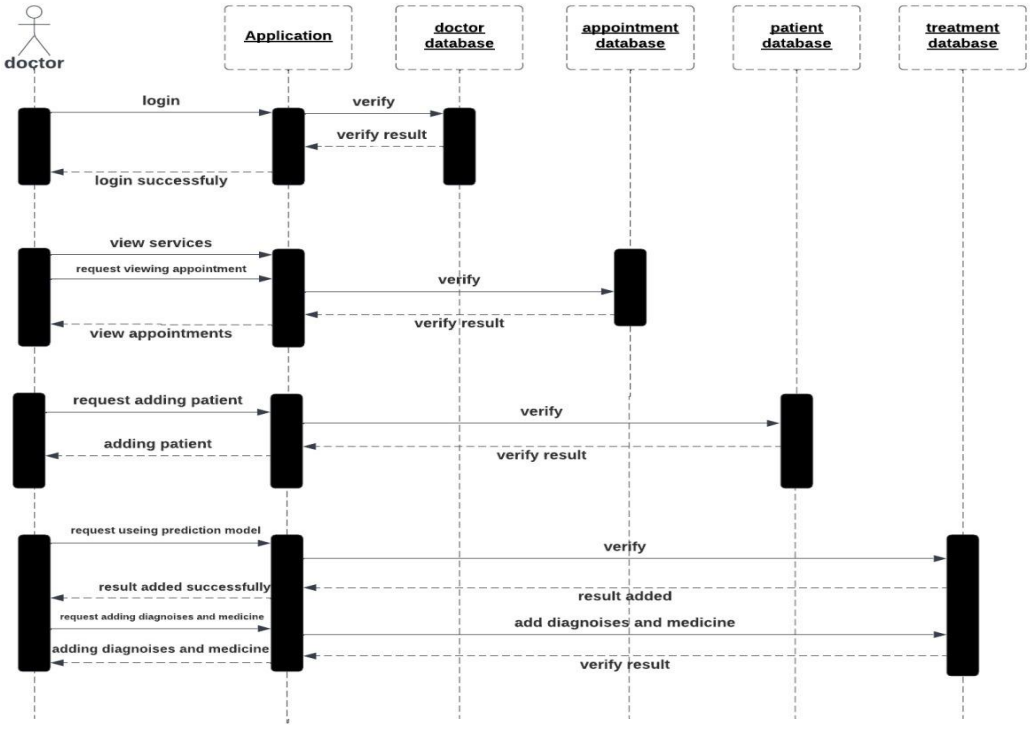


Figure 2_10: Doctor's Sequence Diagram

Sequence diagram of the doctor: explain to us the steps that the system makes after the patient act with the system and what the messages that will appear to the doctor in each case. As the doctor enter the data then the system checks this data if it is correct then print to him login successfully and then let him make his services as adding patient then check the data if it is right print to him the patient was added, predicting with model that the doctor enters the data then the data added to model then the result printed to the doctor.

2.3.5 ER DIAGRAM

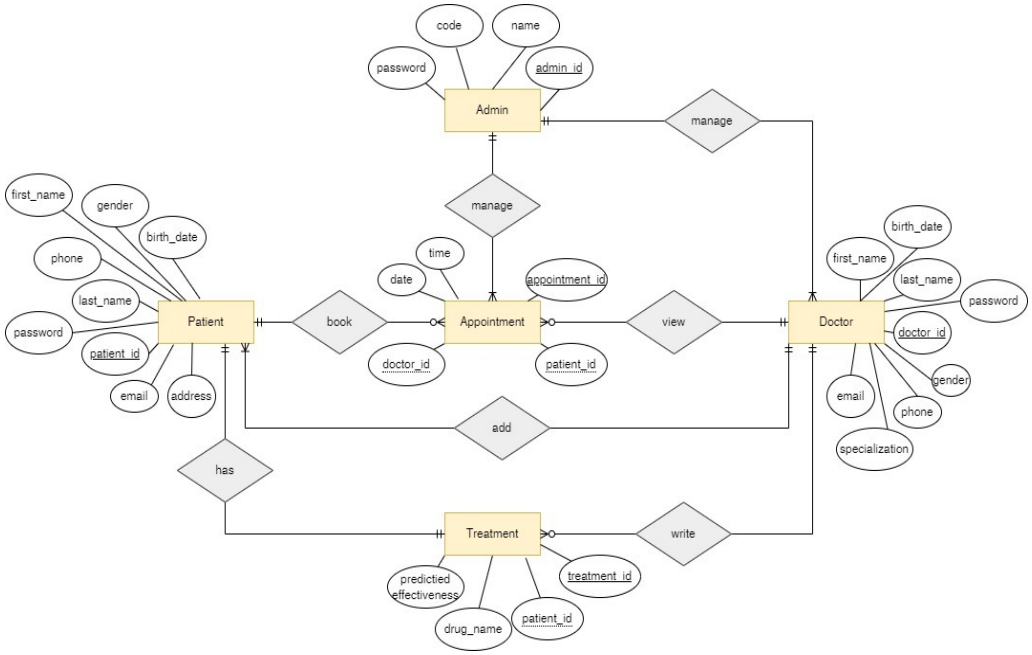


Figure 2_11: ER Diagram

Entity Relationship Diagram: this diagram describes the relationships between entities in a database, such as persons, things, or concepts. Also, ERD describes the properties of these entities. In our system, main entities are (admin, patient, doctor, appointment, and treatment). Each of these entities has attributes such as (id, name, phone, age…. etc.) as shown in diagram.

2.3.6 MAPPING

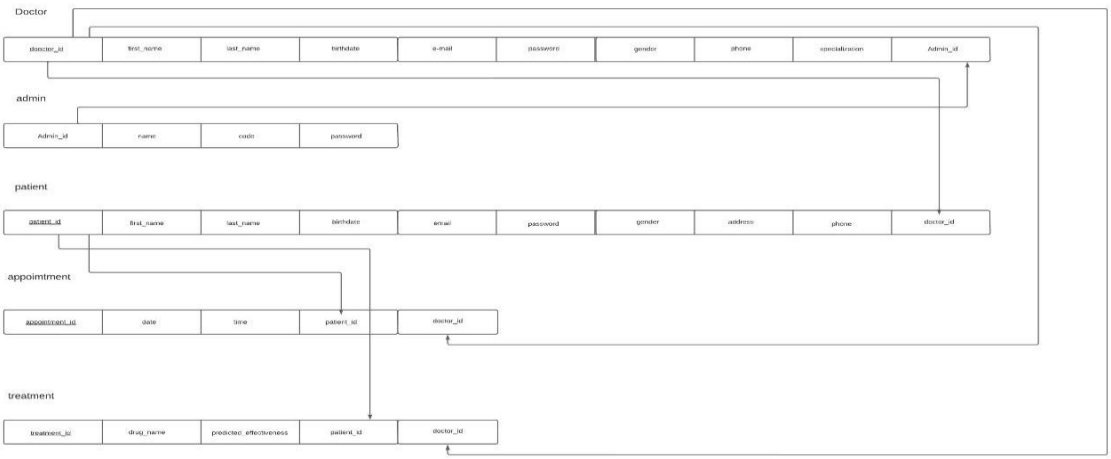


Figure 2_12: Mapping

Mapping: this figure explains the relationship between tables in database, what are the primary key and from which primary key are foreign keys derived.

2.3.7 CLASS DIAGRAM

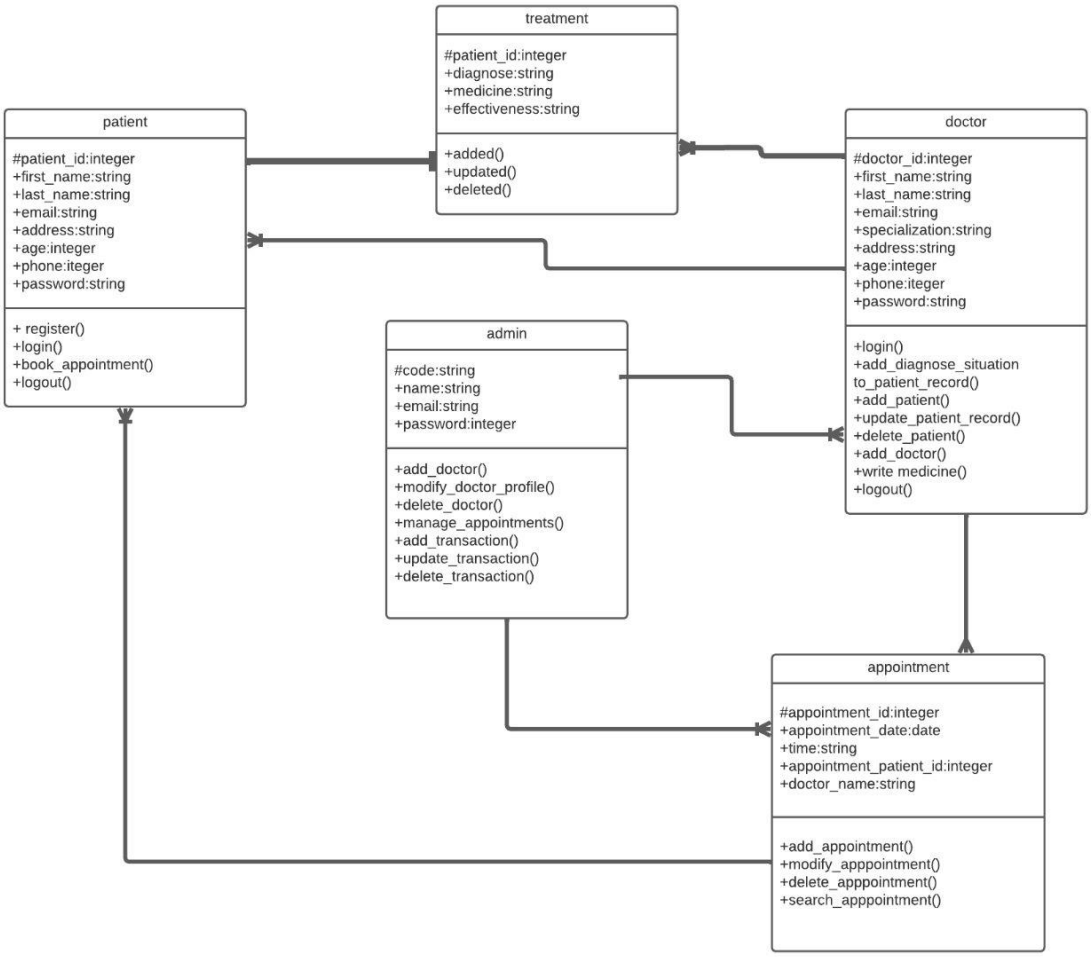


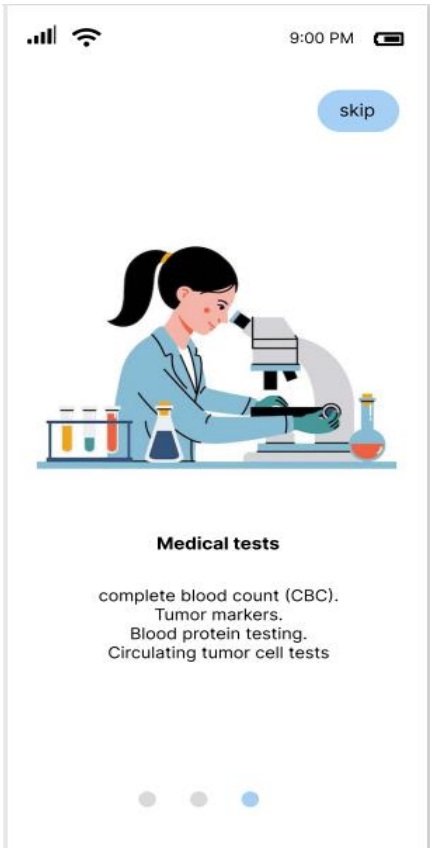
Figure 2_13: Class Diagram

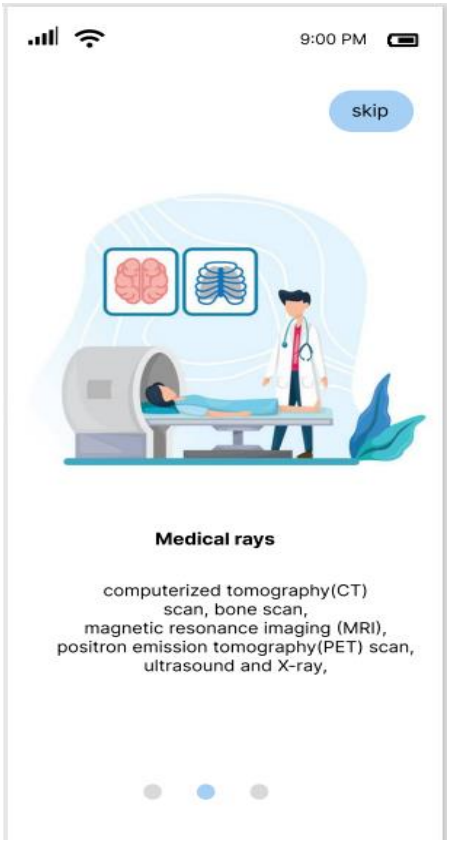
Class Diagram: this diagram shows the relationships between the objects and define what those objects perform and the services they offer. Class diagrams are important at various phases of system design because they display each user's properties and functions. The relations between entities may be one-one, many-many, one-many. Also, it describes the cardinality such as mandatory, optional.

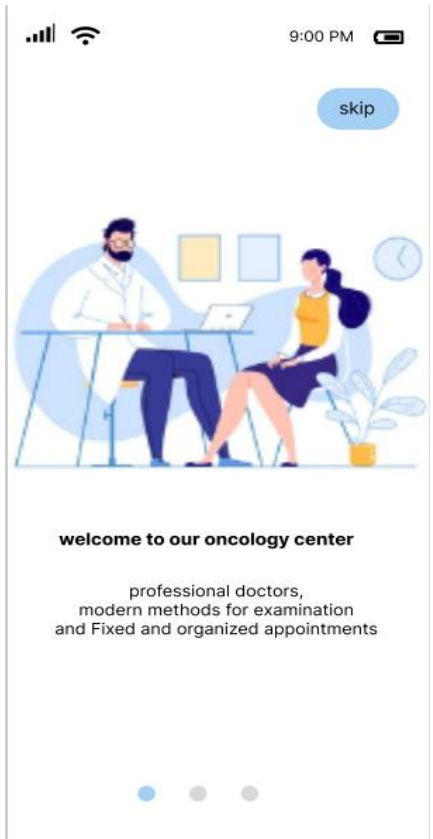
-Relation between appointment and patient is one-many, meaning that each patient may have one or more appointment.

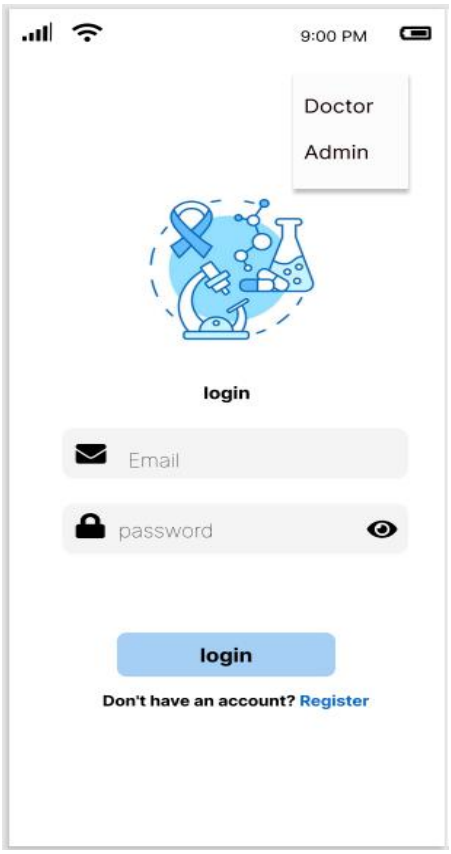
2.3.8 GUI

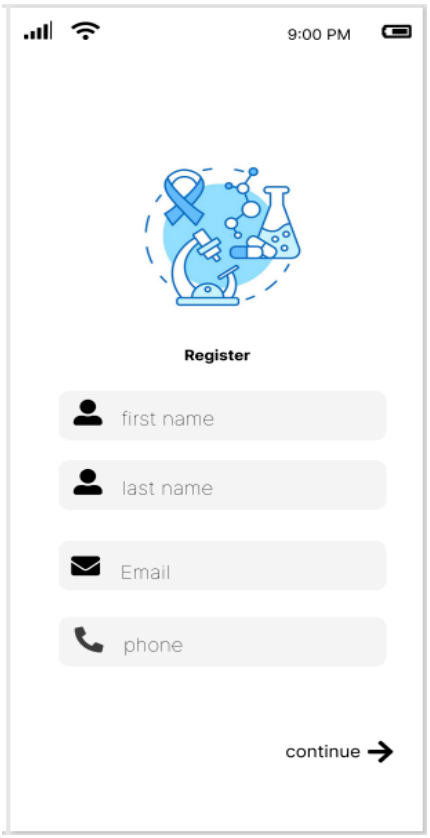
9:00 PM
OCS
Oncology Center Services

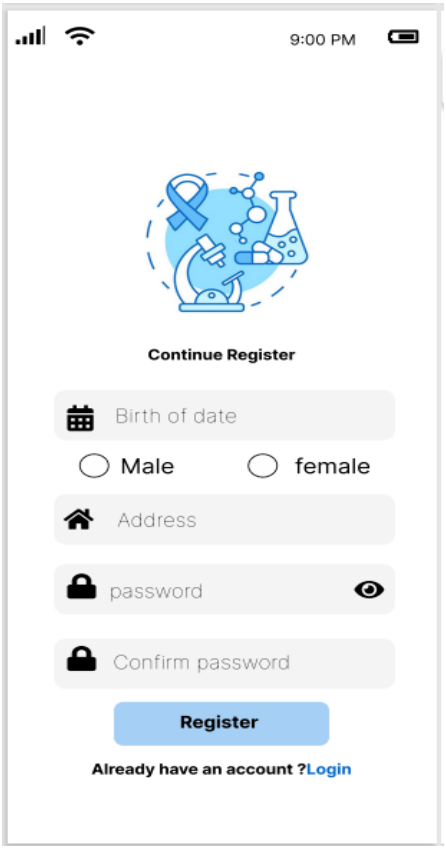

9:00 PM
skip
Medical tests
complete blood count (CBC).
Tumor markers.
Blood protein testing.
Circulating tumor cell tests

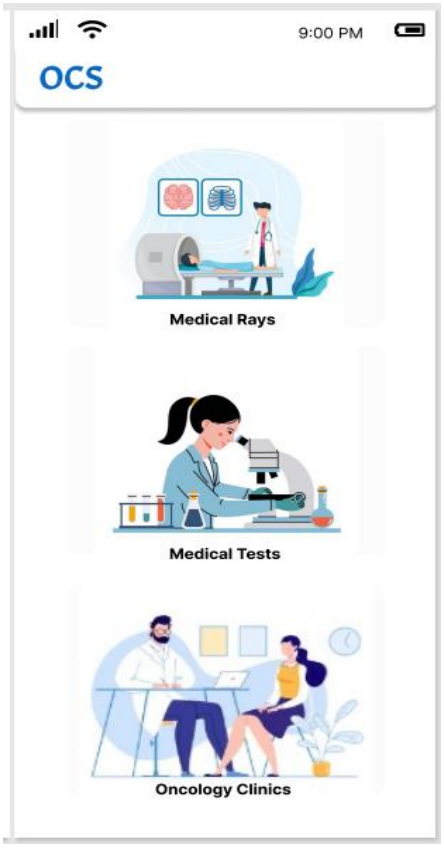

9:00 PM
skip
Medical rays
computerized tomography(CT)
scan, bone scan,
magnetic resonance imaging (MRI),
positron emission tomography(PET) scan,
ultrasound and X-ray,

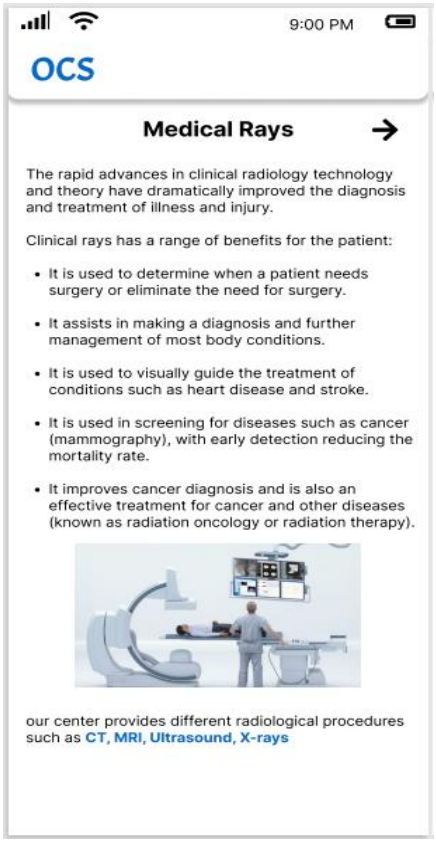

9:00 PM
skip
welcome to our oncology center
professional doctors,
modern methods for examination
and Fixed and organized appointments

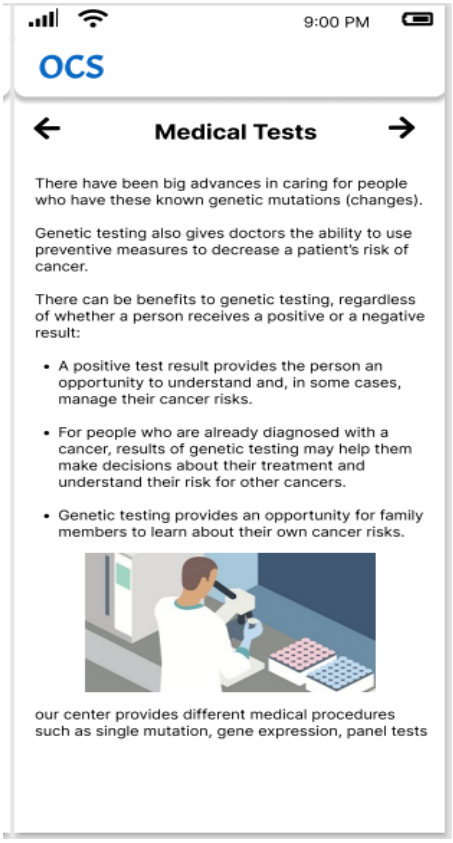

9:00 PM
Doctor
Admin
login
Email
password
login
Don't have an account? Register


9:00 PM
Register
first name
last name
Email
phone
continue

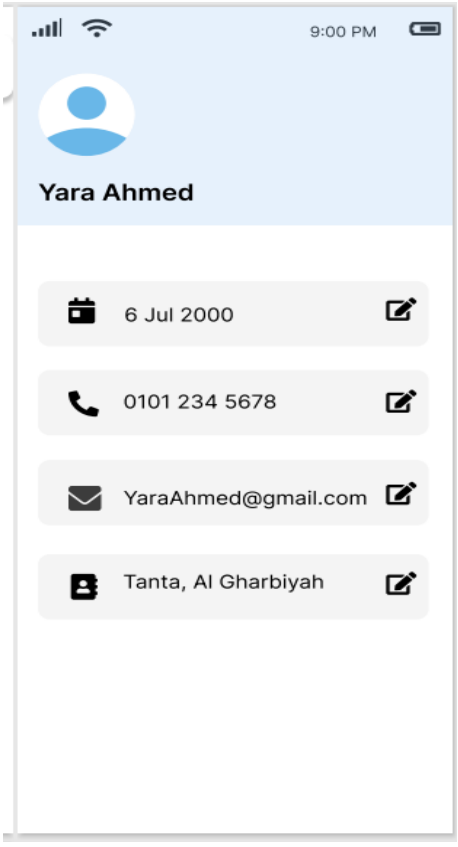

9:00 PM
Continue Register
Birth of date
Male
female
Address
password
Confirm password
Register
Already have an account ?Login

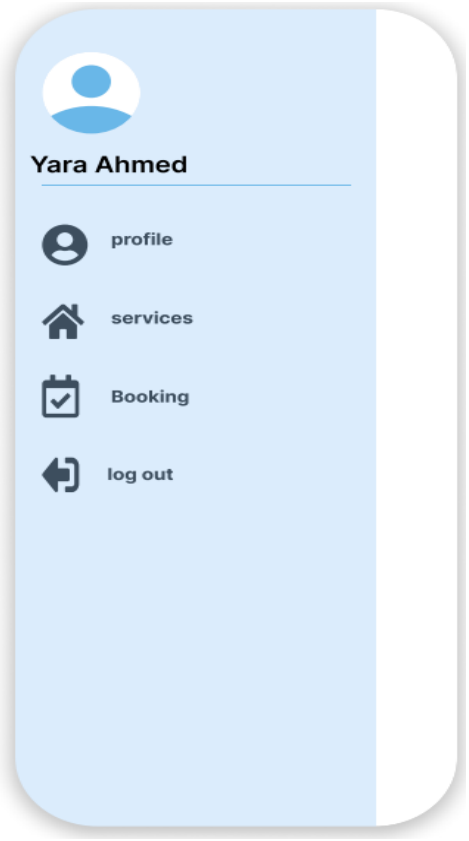

9:00 PM
OCS
Medical Rays
Medical Tests
Oncology Clinics

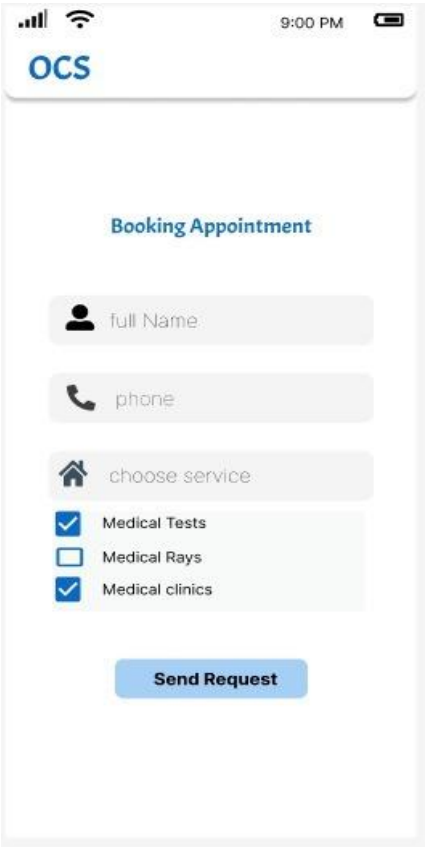

9:00 PM
OCS
Medical Rays
The rapid advances in clinical radiology technology and theory have dramatically improved the diagnosis and treatment of illness and injury.
Clinical rays has a range of benefits for the patient:
• It is used to determine when a patient needs surgery or eliminate the need for surgery.
• It assists in making a diagnosis and further management of most body conditions.
• It is used to visually guide the treatment of conditions such as heart disease and stroke.
• It is used in screening for diseases such as cancer (mammography), with early detection reducing the mortality rate.
• It improves cancer diagnosis and is also an effective treatment for cancer and other diseases (known as radiation oncology or radiation therapy).
our center provides different radiological procedures such as CT, MRI, Ultrasound, X-rays


9:00 PM
OCS
Medical Tests
There have been big advances in caring for people who have these known genetic mutations (changes).
Genetic testing also gives doctors the ability to use preventive measures to decrease a patient's risk of cancer.
There can be benefits to genetic testing, regardless of whether a person receives a positive or a negative result:
• A positive test result provides the person an opportunity to understand and, in some cases, manage their cancer risks.
• For people who are already diagnosed with a cancer, results of genetic testing may help them make decisions about their treatment and understand their risk for other cancers.
• Genetic testing provides an opportunity for family members to learn about their own cancer risks.
our center provides different medical procedures such as single mutation, gene expression, panel tests

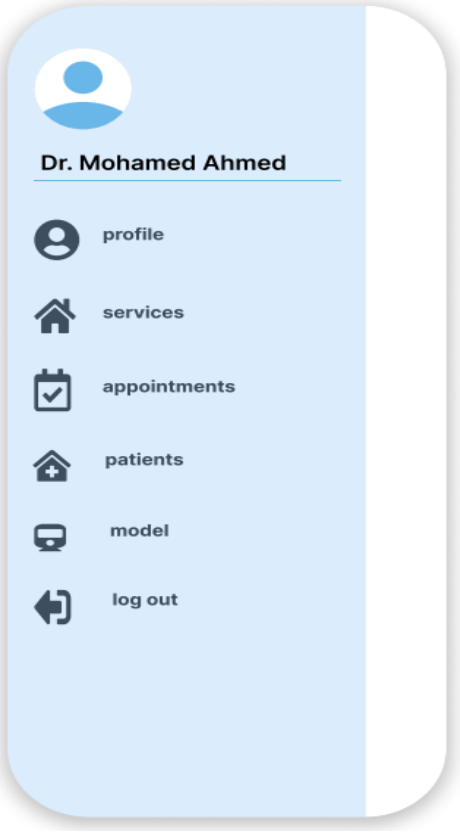
9:00 PM
OCS
Oncology Clinics
Providing health care for cancer is different than for any other disease.
So in OCS, we have widely varying levels of experience in diagnosing and treating the various types of cancer.
Complex cancers are best treated by specialists who have experience in treating patients with that type of cancer — not a general oncologist who treats multiple cancer types.
In our center we follow pharmacogenom, which means each patient has it's own diagnose and medicine even if the disease is one.
We wish you good health and speedy recovery

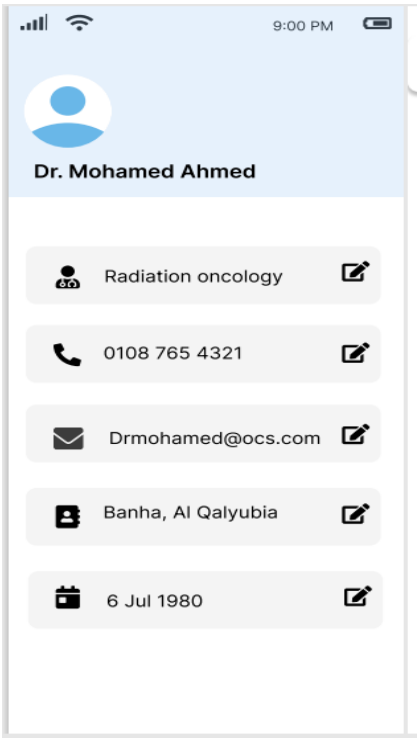

9:00 PM
Yara Ahmed
6 Jul 2000
0101 234 5678
YaraAhmed@gmail.com
Tanta, Al Gharbiyah

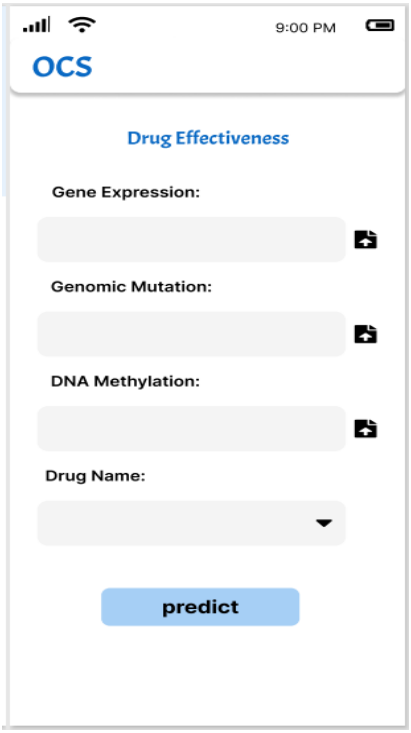

Yara Ahmed
profile
services
Booking
log out

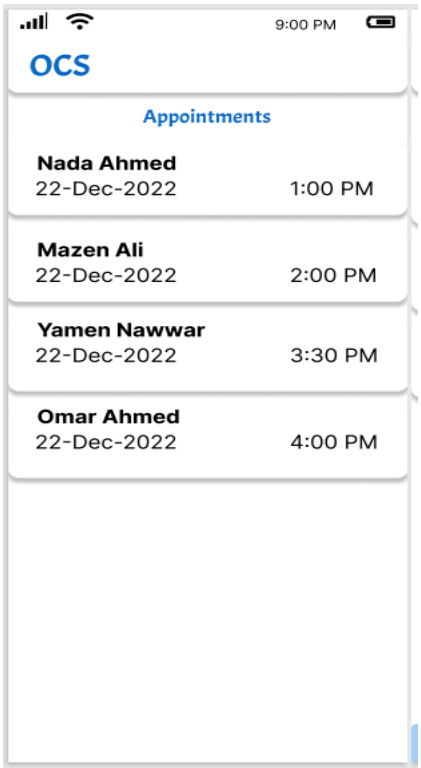

9:00 PM
OCS
Booking Appointment
full Name
phone
choose service
Medical Tests
Medical Rays
Medical clinics
Send Request

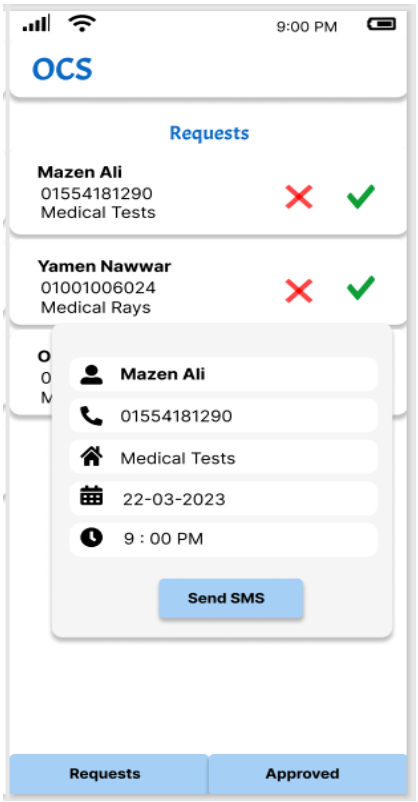
9:00 PM
OCS
Patient Record
Name: Mohammed Alaa
Email: Mohammed555@yahoo.com
Phone: 01209632578
Diagnosis: Blood Cancer
Medicine: Alectinib 150 mg

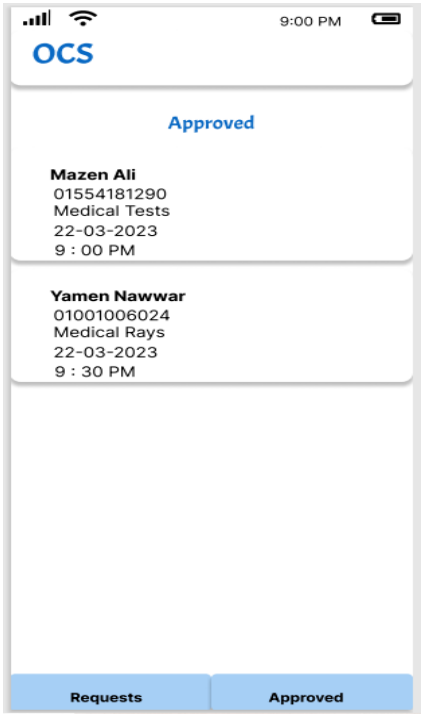

Dr. Mohamed Ahmed
profile
services
appointments
patients
model
log out

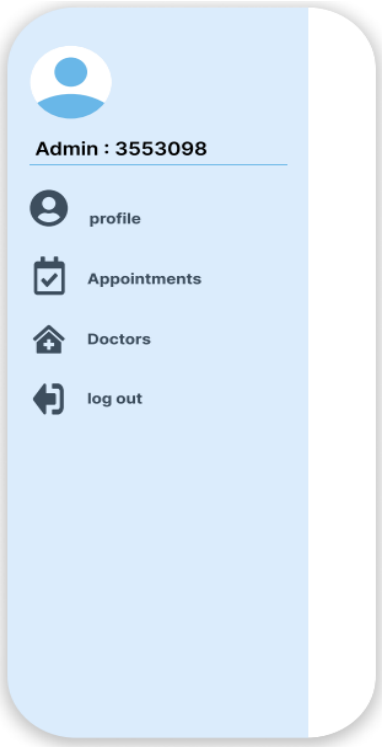

9:00 PM
Dr. Mohamed Ahmed
Radiation oncology
0108 765 4321
Drmohamed@ocs.com
Banha, Al Qalyubia
6 Jul 1980

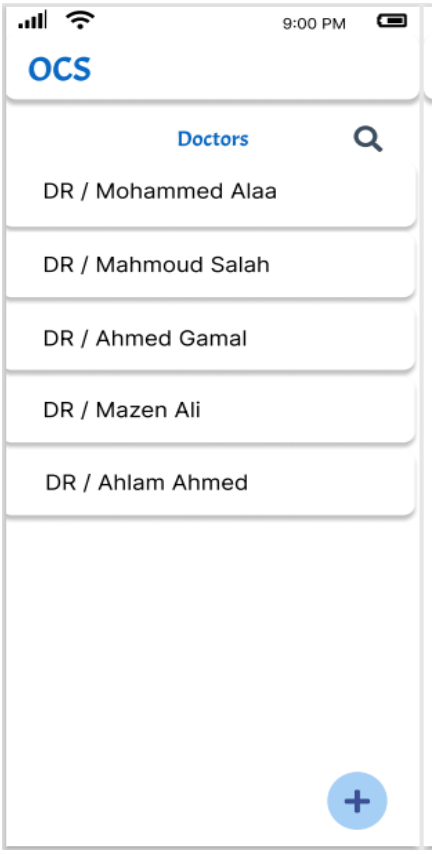

9:00 PM
OCS
Drug Effectiveness
Gene Expression:
Genomic Mutation:
DNA Methylation:
Drug Name:
predict

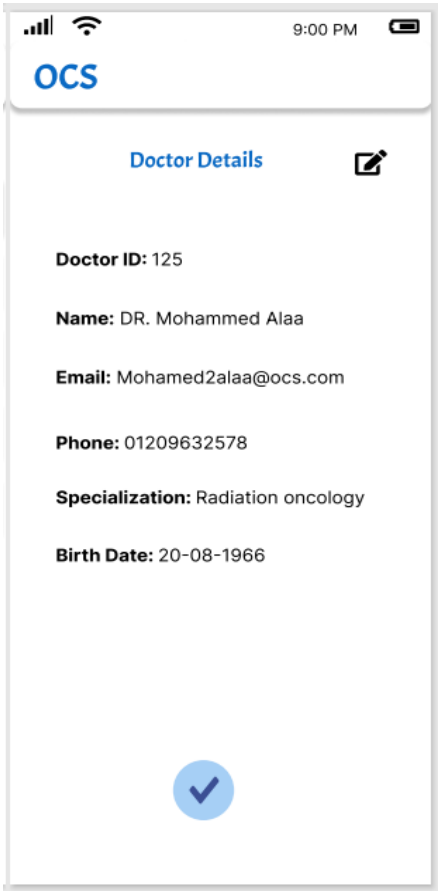

9:00 PM
OCS
Appointments
Nada Ahmed
22-Dec-2022
1:00 PM
Mazen Ali
22-Dec-2022
2:00 PM
Yamen Nawwar
22-Dec-2022
3:30 PM
Omar Ahmed
22-Dec-2022
4:00 PM


9:00 PM
OCS
Requests
Mazen Ali
01554181290
Medical Tests
Yamen Nawwar
01001006024
Medical Rays
Mazen Ali
01554181290
Medical Tests
22-03-2023
9 : 00 PM
Send SMS
Requests
Approved


9:00 PM
OCS
Approved
Mazen Ali
01554181290
Medical Tests
22-03-2023
9 : 00 PM
Yamen Nawwar
01001006024
Medical Rays
22-03-2023
9 : 30 PM
Requests
Approved


Admin : 3553098
profile
Appointments
Doctors
log out


9:00 PM
OCS
Doctors
DR / Mohammed Alaa
DR / Mahmoud Salah
DR / Ahmed Gamal
DR / Mazen Ali
DR / Ahlam Ahmed


9:00 PM
OCS
Doctor Details
Doctor ID: 125
Name: DR. Mohammed Alaa
Email: Mohamed2alaa@ocs.com
Phone: 01209632578
Specialization: Radiation oncology
Birth Date: 20-08-1966

Chapter Three

3 MODEL

3.1 DATA SET

We have integrated 2 public databases in our study: **GDSC and CCL**. In our study, we operate on 256 genes and 238 drugs. **GDSC** database provide us: IC50 values for large-scale drug screening data, of which each IC50 value corresponds to a pair of cancer cell line and drug interaction.

Half-maximal inhibitory concentration (IC50): is the most widely used and informative measure of a drug's efficacy. It indicates how much drug is needed to inhibit a biological process by half, thus providing a measure of potency of an antagonist drug in pharmacological research .and when its value is small, it is evidence that the effectiveness of the drug is high.

CCLE database provide us: Genomic, transcriptomic and epigenomic data of more than a thousand cancer cell lines. For the three data omics, we focused on gene mutation data, gene expression data, and DNA methylation data.

1. **Genomic expression**: Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product that enables it to produce end products, protein, or non-coding RNA. In other words: Gene expression is that process of turning on a specific gene to start making messenger RNA.

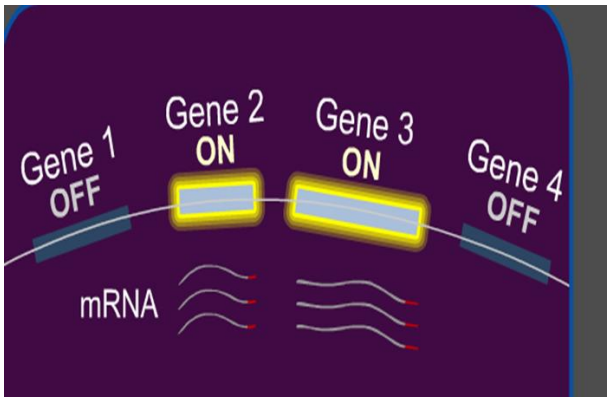


Figure 3_1: Genomic expression

2. **Genomic Mutation**: Genetic mutations are changes to your DNA sequence that happen during cell division when your cells make copies of themselves.

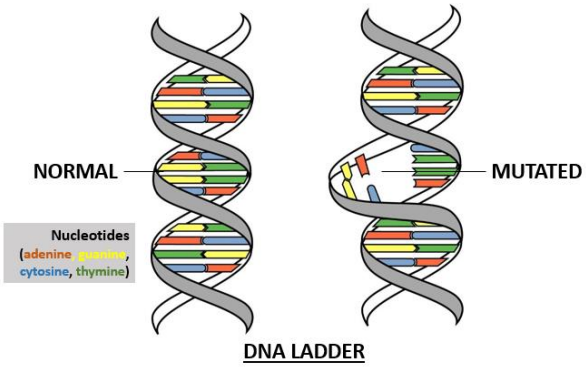


Figure 3_2: Genomic Mutation

3. **Genomic Methylation**: DNA methylation is a biological process by which methyl groups are added to the DNA molecule. Gene is inactivated by addition of methyl group to the DNA-cytosine forming methyl cytosine. Methyl cytosine combines with specific protein to form methyl cytosine protein complex which blocks transcription (stop copying).

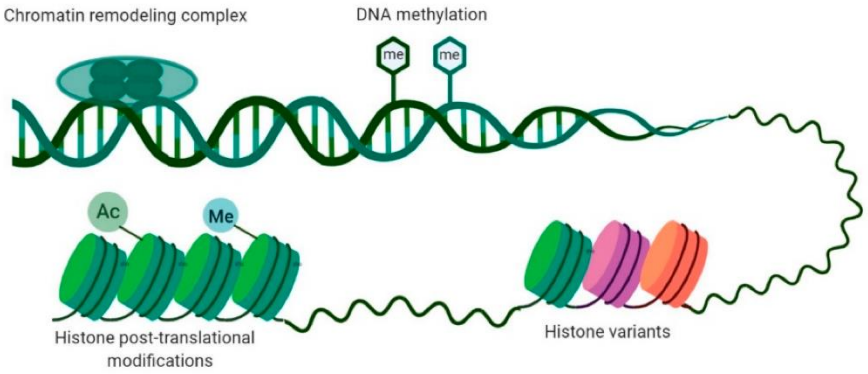


Figure 3_3: Genomic Methylation

3.2 MODEL

The goal of the model is to determine if the drug is effective for the cancer patient or not depends on cancer cell profiles (genomic mutation, gene expression and DNA methylation data).

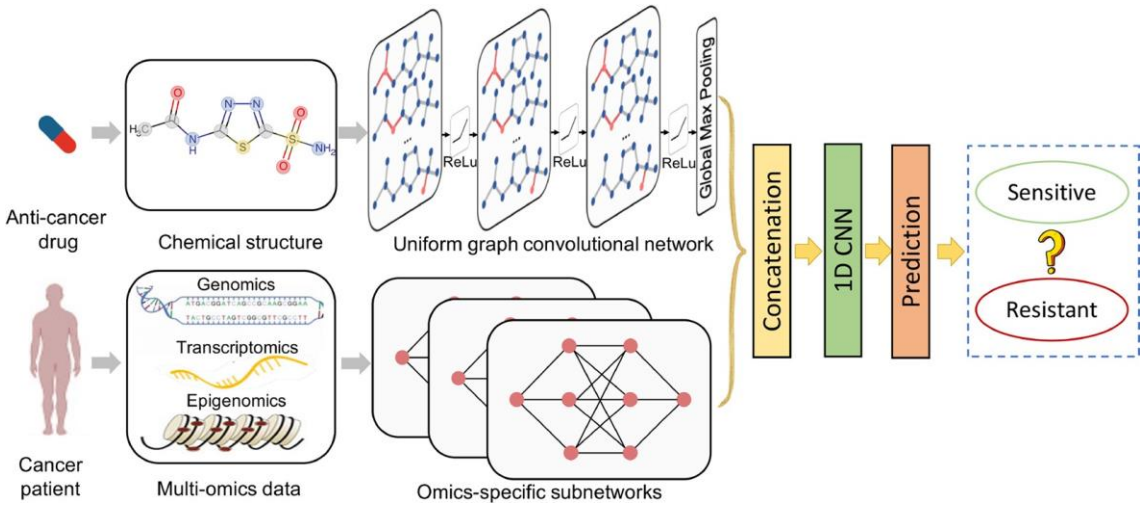


Figure 3_4: Model Framework

The overview framework of model contains a UGCN and three subnetworks for processing drug structures and cancer cell profiles (genomic mutation, gene expression and DNA methylation data) respectively. Model takes a pair of drug and cancer cell profiles as inputs and predicts the drug sensitivity (IC50, which denotes the effectiveness of a drug in inhibiting the growth of a specific cancer cell line.) (regression) or claims the drug to be sensitive or resistant (classification).

Drug is represented as a graph based on the chemical structure before by the **GCN.**

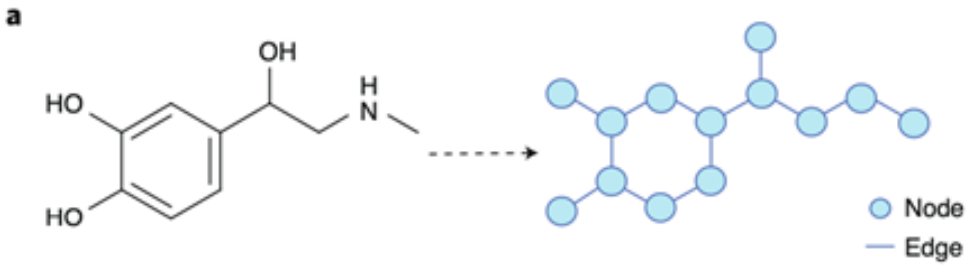


Figure 3_5: Drug Representation

Each Omics data is represented by subnetwork and then integrated to one network.

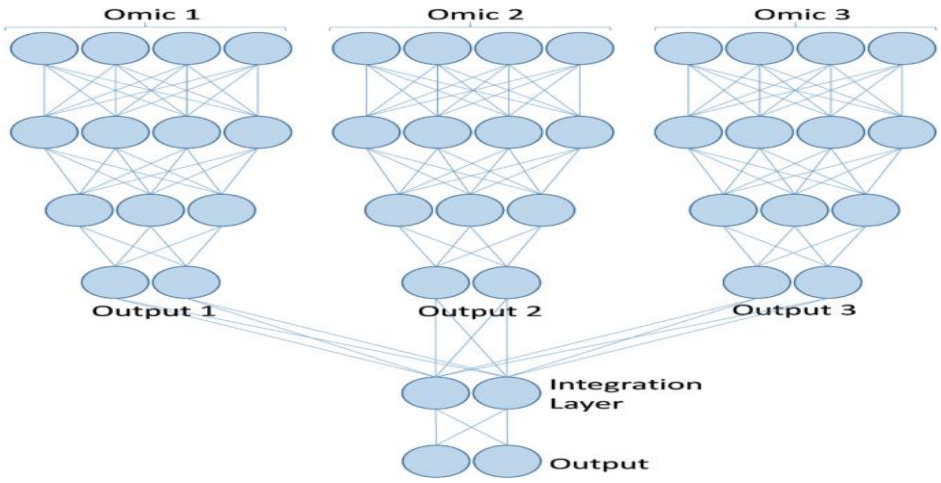


Figure 3_6: Omics Representation

Then concatenate the network (multi-omics-data) and graph (drug) and fed to CNN.

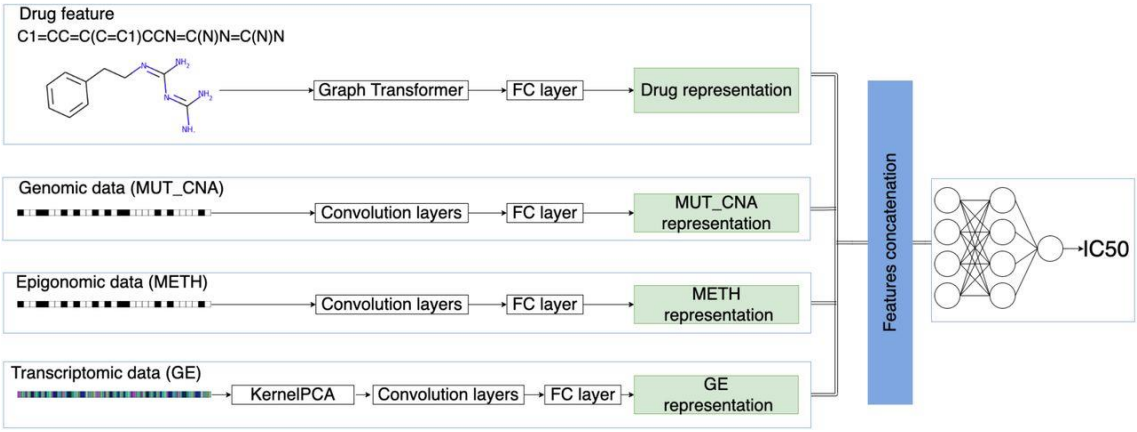


Figure 3_7: Data Representation

Finally predicts the drug sensitivity (IC50) (regression) or classifies the drug to be sensitive or resistant (classification).

3.3 CNN

A convolutional neural network (CNN) is a class of deep neural network most applied to analyze visual imagery.

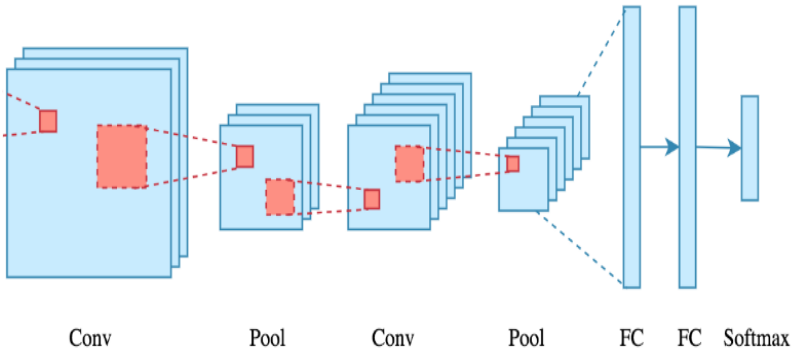


Figure 3_8: CNN Layers

Convolution layer: convolutional neural network applies filter (kernel) to an input to create a feature map that summarize the presence of detected features in the input.

Filter: is a window that scans the image and enable CNNs to learn features from neighboring cells and take their value by training.

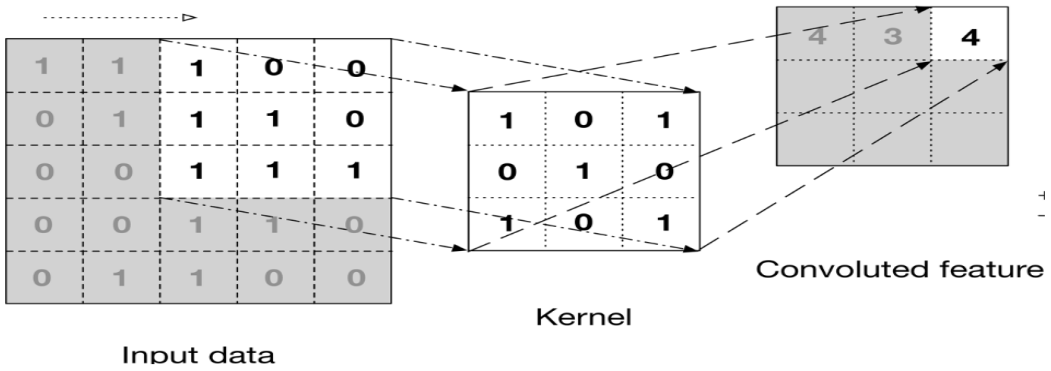


Figure 3_9: Convolution layer

Pooling Layer: is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data by reducing the dimensions. There are two types of pooling average pooling and max pooling.

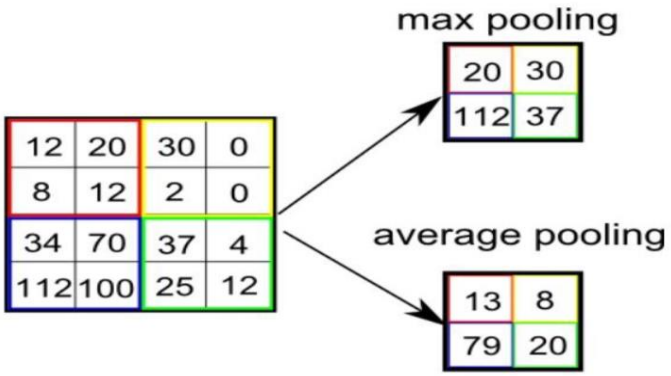


Figure 3_10: Pooling layer

FCI Layer: The feature outputs will be flattened to column vector and feed-forward it to FCL, then this vector will be inputs for neural network.

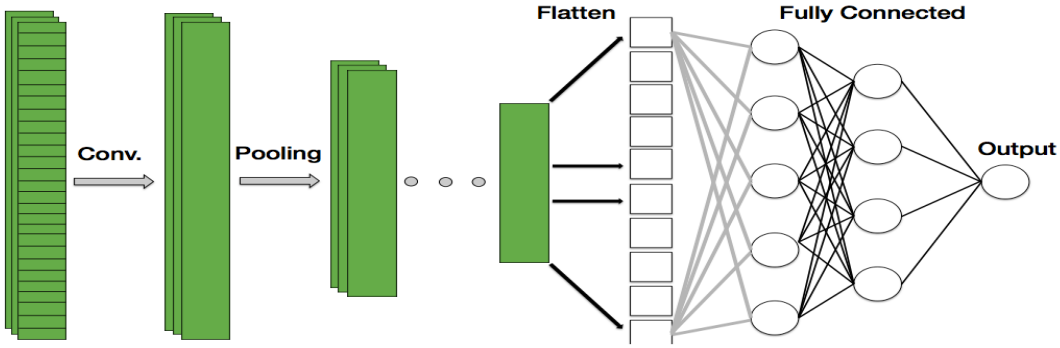


Figure 3_11: FCI layer

3.4 GCN

GCNs are a very powerful neural network architecture for machine learning on graphs. Graph convolutional network (GCN) is a neural network that operates on graphs. Given a graph G = (Nodes, Edges).

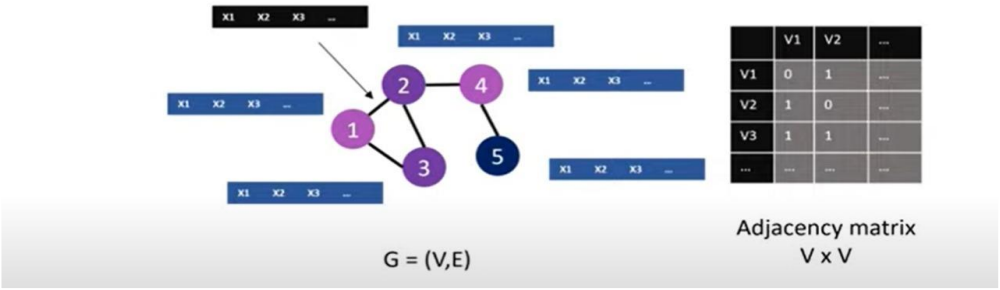


Figure 3_12: GCN

Convolution: For each node, we get the feature information from all its neighbors. Each node takes messages from other nodes in more than round. Each node will be represented by number that means its similarity (Embedding).

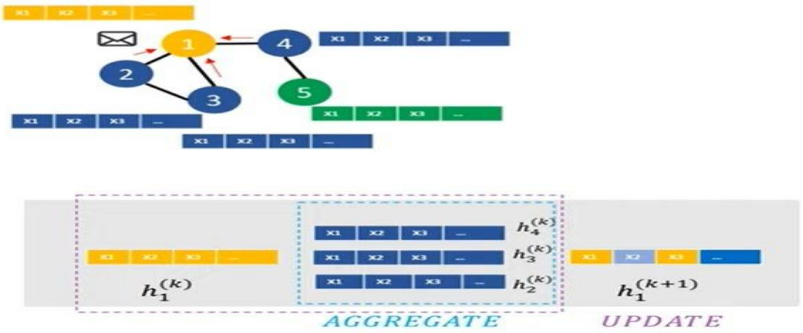


Figure 3_13: Convolution

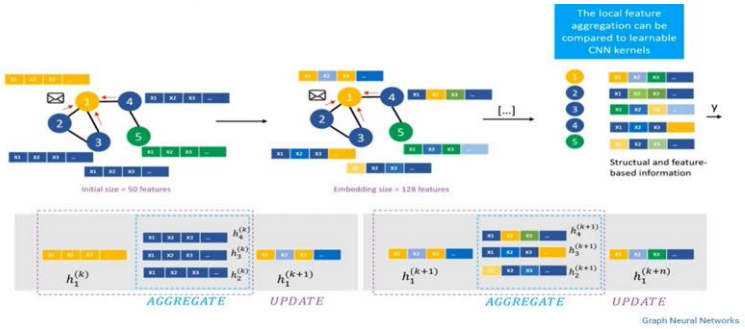


Figure 3_14: Convolution

Pooling Layer: is responsible for reducing the spatial size of the Convoluted Feature. This is to decrease the computational power required to process the data by reducing the dimensions.

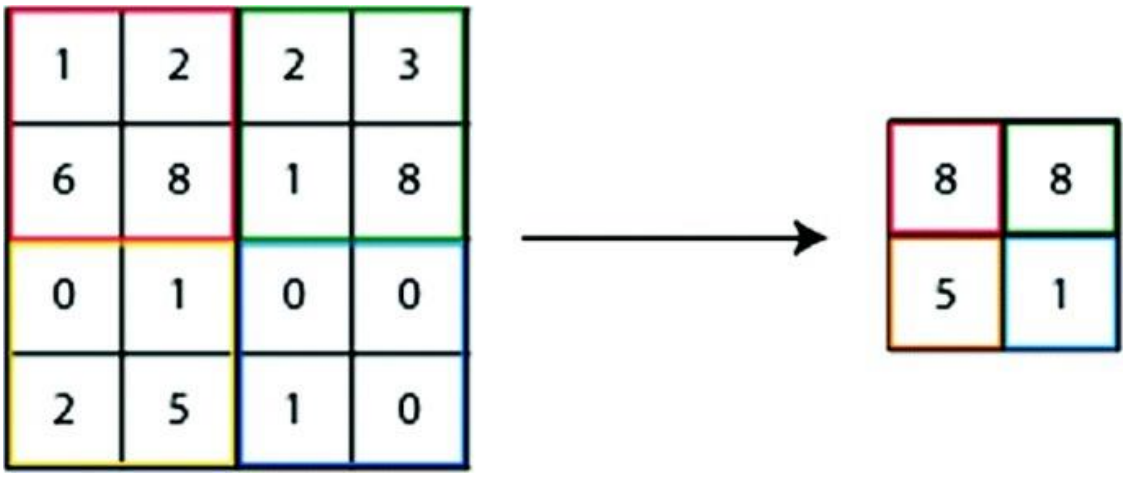


Figure 3_15: Pooling layer

In our model, we use Max pooling which takes maximum value for patches of a feature map and uses it to create a down sampled (pooled) feature map. After that we concatenate this matrix with omics data matrix and then fed to CNN. Which predicts IC50 in case of regression or claims sensitive or resistance in case of classification.

Chapter Four

4 RESULTS

After we built the model, we made many changes and applied many methods to improve the Accuracy.

First: We used train-test split in the data splitting step of preprocessing and we obtained a result in classification: 84.1, precision of regression: 92.7. Then we used k-fold and found an increase in Accuracy classification: 87.66, precision of regression: 95.17.

Second : We changed the building of the model (GCN), so that **after the flatten layer**, we put machine learning model, as it turns out to us through searches that it gives a better result. We use the **k-nearest neighbor** (KNN) using the try and error method. We found that at k =4 the best result was classification Accuracy: 86.04, precision of regression: 94.54. Then we used a **Decision Tree**, and the result was classification Accuracy: 85.24, precision of regression: 93.33. Then we used **Random Forest**, and the result was classification Accuracy: 84.29, precision of regression: 92.91.

Changes	**IDE**	**Run Time**	**Classification Accuracy**	**Pearson's correlation**
Without Changes	Colab Pro	regression: 25m, classification: 20m	84.1	92.7
Random forest with GCN	Colab Pro	regression: 15m, classification: 12m	84.29	92.91
Decision tree with GCN	Colab Pro	regression: 16m, classification: 14m	85.24	93.33
KNN (K=4) with GCN	Colab Pro	regression :14m, classification:9m	86.04	94.54
k-fold	Colab Pro	regression: 30m, classification: 22m	87.66	95.17

At the end, it became clear to us from the previous that when we used.

-k-fold

We got a good initial result (Accuracy classification: 87.66, Pearson's correlation: 95.17).

CONCLUSION

Diagnosing and treating cancer is not easy and need the most effective tools to help doctors and patients. OCS is the application we developed to assist in cancer field, and we hope it helps. We used deep learning techniques such as CNN, GCN. OCS can be served as an application for exploring drug sensitivity with large-scale cancer multi-omics profiles. OCS outperforms multiple baselines, and our analysis illustrates how our method can help prioritize therapeutic targets for anti-cancer drug discovery. And we targeted android devices to this app as most people in world use android so the app will be most used and spread.

REFERNCES

[1] Maureen McCutcheon. Where Have My Eyebrows Gone?. Cengage Learning, 2001. ISBN 0766839346. Page 5.

[2] Cancer. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed Feb. 16, 2021.

[3] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer Vision – ECCV 2014. Cham: Springer International Publishing; 2014. p. 818–33.

[4] ZhangN. et al. (2015) Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. *PLoS Comput. Biol*., 11, e1004498.

[5] Qiao Liu, Zhiqiang Hu, Rui Jiang, Mu Zhou, DeepCDR: a hybrid graph convolutional network for predicting cancer drug response, December 2020, Pages i911–i918.

[6] DalyA.K.(2017) Pharmacogenetics: a general review on progress to date. *Br. Med. Bull*., 124, 1–79.