



**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. Mohammed Alhakim

Benha. June 2023

ACKNOWLEDGMENT

First of all, we would like to illustrate that our project idea came from our desire of using modern technologies in the field of computers and artificial intelligence in saving the lives of many people who die every day because of fatal cancer diseases. We finally succeeded in our first phase of the project but we could not have done this without our supervisors so we would like to express our sincere thanks to **Dr. Elsayed Badr** for giving us this great opportunity to make this project and for his valuable guidance and support over the first and second semesters. We would also like to express our gratitude towards **Eng. Mohamed AlHakim** for his help in various stages of the completion of the project. Without their support and suggestions, this project would not have been completed.

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: 22, JUN 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 with K_Fold is 1.0000 and regression is 0.9996.

Table Of Contents

List of Figures.....	ii
List of figures	1
Chapter 1: Introduction And Background	1
1.1 Introduction	1
1.2 Problem Definition	2
1.3 Problem Solution	3
Chapter 2: Planning AND REQUIREMENTS	4
2.1 Planning	4
2.1.1 Project scope	4
2.1.2 Project activities & duration.....	4
	5
2.2 Requirements	5
2.2.1 Business Requirements identification	5
2.2.2.1 User Functional Requirements	5
2.2.2.2 System Functional Requirements	6
2.2.2.3 System Non-Functional Requirements.....	7
Chapter 3: Analysis and design	9
3.1 Introduction.....	9
3.2 System Design	9
3.2.1 Use Case Diagram.....	9
3.2.2 DFD Diagrams	10
3.2.2.1 DFD level 0 (Context Diagram).....	10
3.2.2.2 DFD Level 1.....	11
3.2.2.3 DFD Level 2.....	12
3.2.3 ACTIVITY DIAGRAM.....	15
3.2.3.1 Doctor	15
3.2.3.2 Patient	16
3.2.4 Sequence Diagram	17
3.2.4.1 Patient	17
3.2.4.2 Doctor	18
3.2.5 ER Diagram.....	19
3.2.6 mapping.....	20
3.2.7 Class Diagram	21
3.2.8 Database:	22
3.2.8.1: Authentaction	22
3.2.8.2 Realtime database:.....	23
Chapter 4: Model.....	31
4.1 Data set.....	31
4.2 Model.....	34
4.3 GCN.....	36
Chapter 5: Results.....	43
Chapter 6: SYSTEM IMPLEMENTATION	46
6.1 System Overview:	46
6.2 Introduction Pages:	47
6.3 Admin pages	48
6.3.1 Login Page	48
6.3.2 Dashboard	49
6.3.3 Navigation view	50
6.3.4 Profile	51

Table of Contents

6.3.5 Appointment requests.....	52
6.3.6 Appointment Confirmation	53
6.3.7 Approved requests.....	54
6.3.8 Doctors	55
6.3.9 Add Doctors	56
6.4 Patient pages	57
6.4.1 Register Pages:	57
6.4.2 Login Page	58
6.4.3 Navigation	60
6.4.4 System Services (Home)	61
6.5 docyor pages	66
6.5.1 Doctor Login and Dashboard	66
6.5.2 Manage patients.....	67
6.5.3 Appointments	68
6.5.4 Model	69
Conclusion.....	72
References.....	72
Appendix.....	74
Parts Of Code	74
Run Screens	76

LIST OF FIGURES

Figure 3_1: use case diagram.....	9	
Figure 3_2: context diagram	10	
Figure 3_3: DFD level 1.....	11	
Figure 3_4: DFD level 2.....	12	
Figure 3_5: DFD level 2.....	13	
Figure 3_6: DFD level 2.....	14	
Figure 3_7: Doctor's Activity Diagram	15	
Figure 3_8: patient's Activity Diagram.....	16	
Figure 3_9: Patient's Sequence Diagram	17	
Figure 3_10: Doctor's Sequence Diagram	18	
Figure 3_11: ER Diagram	19	
Figure 3_12: Mapping.....	20	
Figure 3_13: Class Diagram.....	21	
Figure 3_14: Authentication.....	23	
Figure 3_15: real time database.	24	
Figure 3_16: real time database for doctors.	24	
Figure 3_17: real time database for admins.	25	
Figure 3_18: real time database for patients.....	26	
Figure 3_19: real time database for patients.....	26	
Figure 3_20: real time database for patients.....	27	
Figure 3_21: Machine learning models.....	27	
Figure 3_22: stored data.....	28	
Figure 3_23: stored data.....	28	
Figure 3_24: drug files.....	29	
Figure 3_25: HKL files.....	29	
Figure 3_26: Patient's files.	30	
Figure 4_1: Genomic expression.....	31	
Figure 4_2: Genomic expression dataset.....	32	
Figure 4_3: Genomic Mutation	32	
Figure 4_4: Genomic Mutation dataset	32	
Figure 4_5: Genomic Methylation Number	33	
Figure 4_6: Genomic methylation dataset.....	33	
Figure 4_7: Drugs dataset	34	
Figure 4_8: Model Framework	34	
Figure 4_9: Model framework (first part)	35	
Figure 4_10: Drug Representation	35	
Figure 4_11: GCN.....	36	
Figure 4_12: Convolution	37	
Figure 4_13: Dropout.....	38	
Figure 4_14: Pooling Layer.....	39	
Figure 4_15: Model framework (second part).....	39	
Figure 4_16: Omics Representation	40	
Figure 4_17: Data Representation	40	
Figure 4_18: CNN Layers	41	
Figure 4_19: Convolution layer	41	
Figure 4_20: Pooling layer	42	
Figure 4_21: FCI layer	42	
Figure 6_1: Logo	Figure 6_2: Welcome Page.....	47
Figure 6_3: Medical Rays	Figure 6_4: Medical Tests.....	47
Figure 6_5: Admin Login.....		48
Figure 6_5: Dashboard Page.		49
Figure 6_6: Navigation View Page.		50
Figure 6_7: Profile Page.....		51
Figure 6_8: Requests Page.		52
Figure 6_8: Confirmation Page.	Figure6_9: Confirmation message.....	53

List of figures

Figure 6_10: Approved Requests Page.....	54
Figure 6_11: Doctors page.....	55
Figure 6_13: Add Doctors page.....	56
Figure 6_14: Register Page 1.....	57
Figure 6_16: Login Page.....	58
Figure 6_17: Forget Password Page.....	59
Figure 6_19: Verification Page.....	59
Figure 6_21: Navigation Bar.....	60
Figure 6_22: Services Page1.....	61
Figure 6_24: Medical Rays.....	62
Figure 6_25: Booking appointment.....	63
Figure 6_26: Medical Tests.....	64
Figure 6_27: Medical Clinics.....	65
Figure 6_28: Doctor Login.....	66
Figure 6_30: Search Patient Page.....	67
Figure 6_32: Appointments Page.....	68
Figure 6_33: Model Page.....	69
Figure 6_34: Upload files Page.....	70
Figure 6_36: Predicting Page.....	71
Figure 7_1: GCN Layers.....	74
Figure 7_2: Train_Test_Split	75
Figure 7_3: K_Fold	75
Figure 7_4: Colab Pro	76
Figure: 7_5 Run of classification (Train_Test_Split).....	76
Figure: 7_6 Run of Regression (Train_Test_Split).....	77
Figure: 7_7 Run of classification (K_Fold).....	77
Figure: 7_8 Run of Regression (K_Fold).....	78

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 PLANNING AND REQUIREMENTS

2.1 PLANNING

2.1.1 Project scope

The system aims to assist doctors in organizing their appointments and choosing the right drug for each patient. Also, the application helps Patients to book appointment easily and go to the center on their time. Also, it helps admins to manage doctors' data and to accept or cancel requests. As cancer is one of the most dangerous diseases that affect humans, highlighting tumor centers and paying attention to solving existing problems is important and necessary, and this is what we try to do in our project.

2.1.1 Project activities & duration

Project Activities & Duration			
Task	Start Date	End Date	Duration in days
Understand paper	10-OCT-2022	20-OCT-2022	10
Collect Data	22-OCT-2022	28-OCT-2022	6
System requirements	1-NOV-2022	10-NOV-2022	9
System analysis	9-NOV-2022	20-NOV-2022	11
Building Model	22-NOV-2022	20-DEC-2022	28
Testing Model	22-DEC-2022	1-JAN-2023	9
Design GUI	20-JAN-2023	5-FEB-2023	15
Accuracy Improvement	1-FEB-2023	20-APR-2023	70
Building App	15-FEB-2023	1-MAY-2023	90
Connect models to app	2-MAY-2023	7-MAY-2023	5
Testing App	8-MAY-2023	15-MAY-2023	7

GANTT CHART



2.2 REQUIREMENTS

2.2.1 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.

2.2.2.1 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.

2.2.2.2 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).

2.2.2.3 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.

Chapter Three

3 ANALYSIS AND DESIGN

3.1 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.

3.2 SYSTEM DESIGN

3.2.1 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.

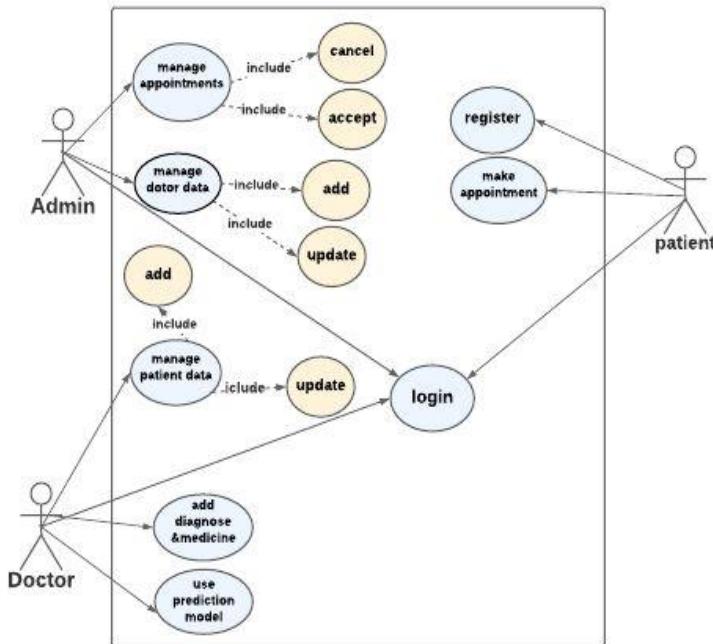


Figure 3_1: use case diagram.

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.

3.2.2 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.

3.2.2.1 DFD LEVEL 0 (CONTEXT DIAGRAM)

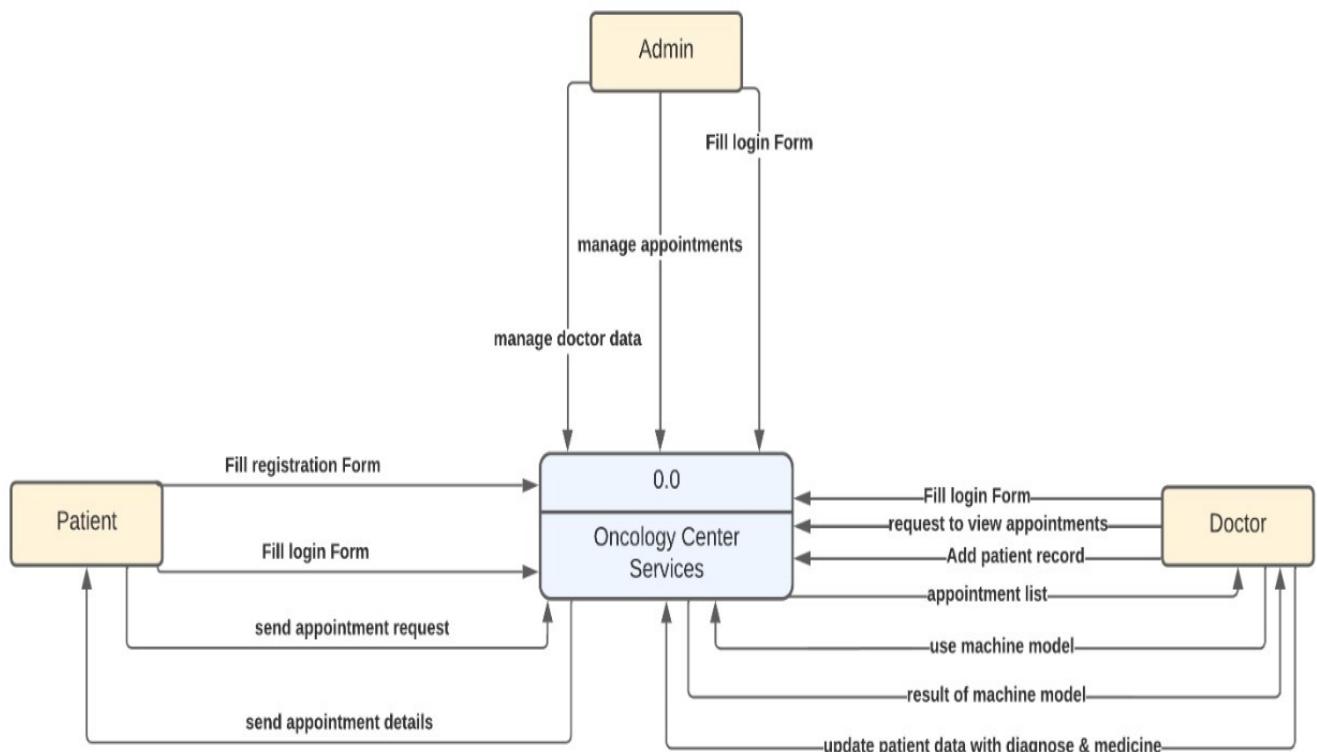


Figure 3_2: context diagram

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.

3.2.2.2 DFD Level 1

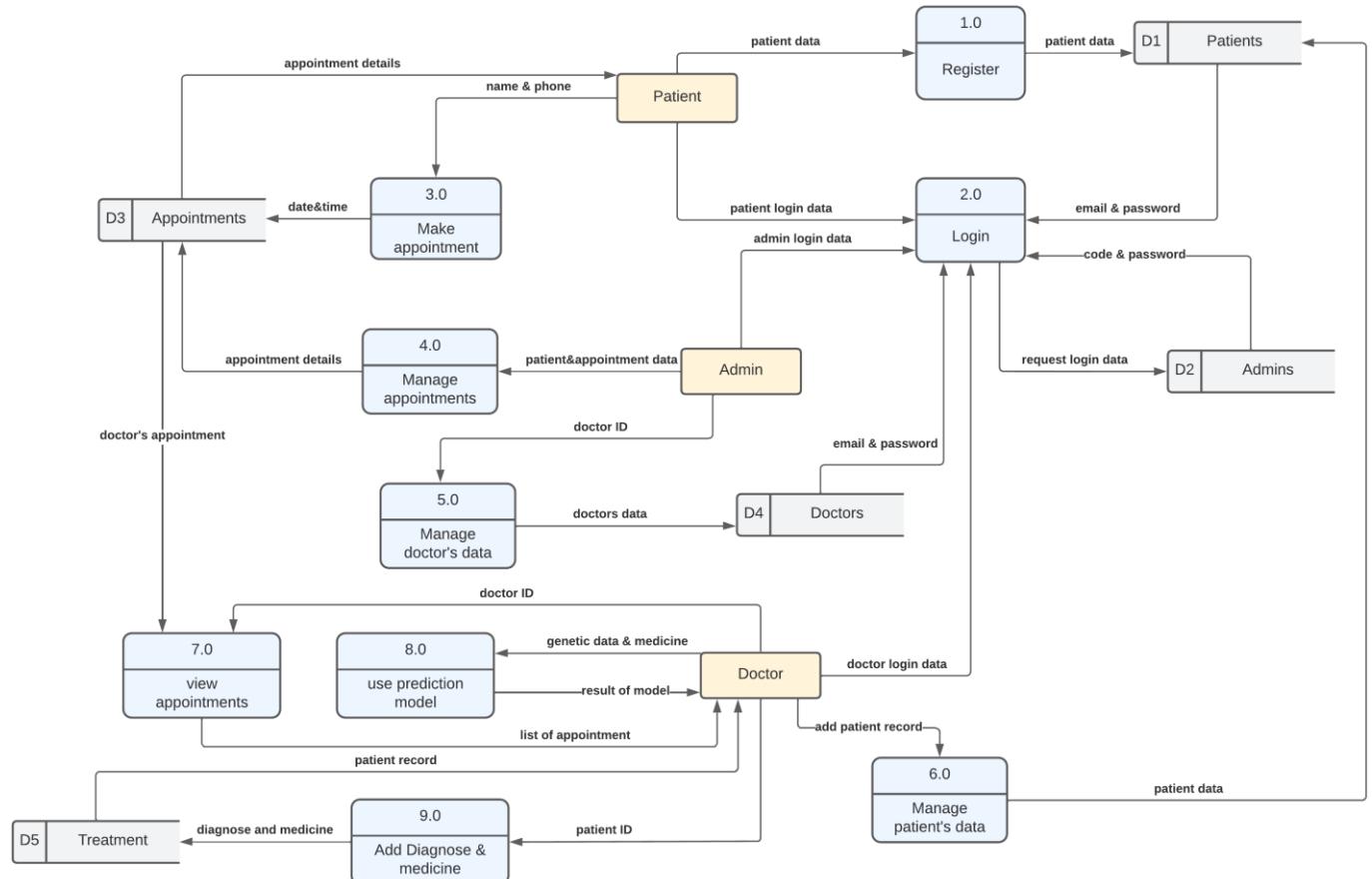


Figure 3_3: DFD level 1

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.

3.2.2.3 DFD Level 2

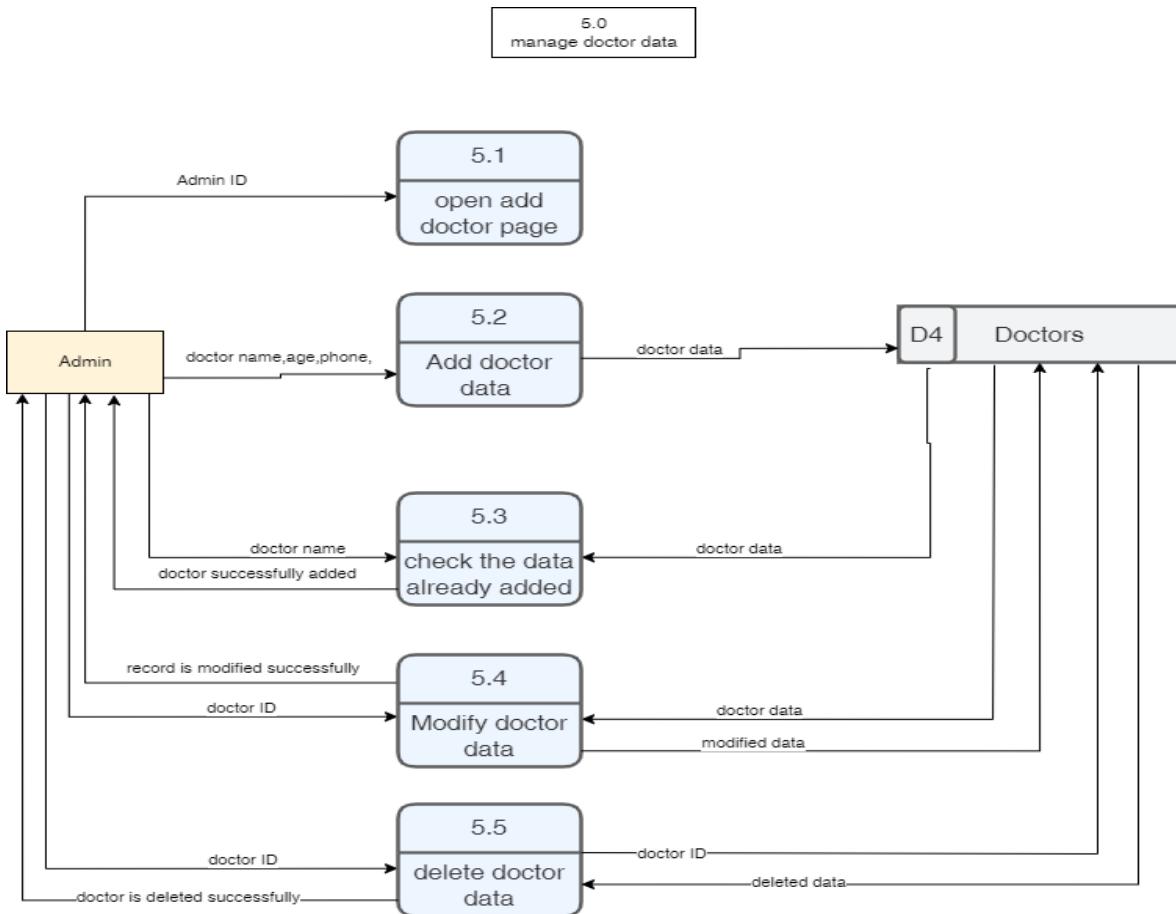


Figure 3_4: DFD level 2

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.

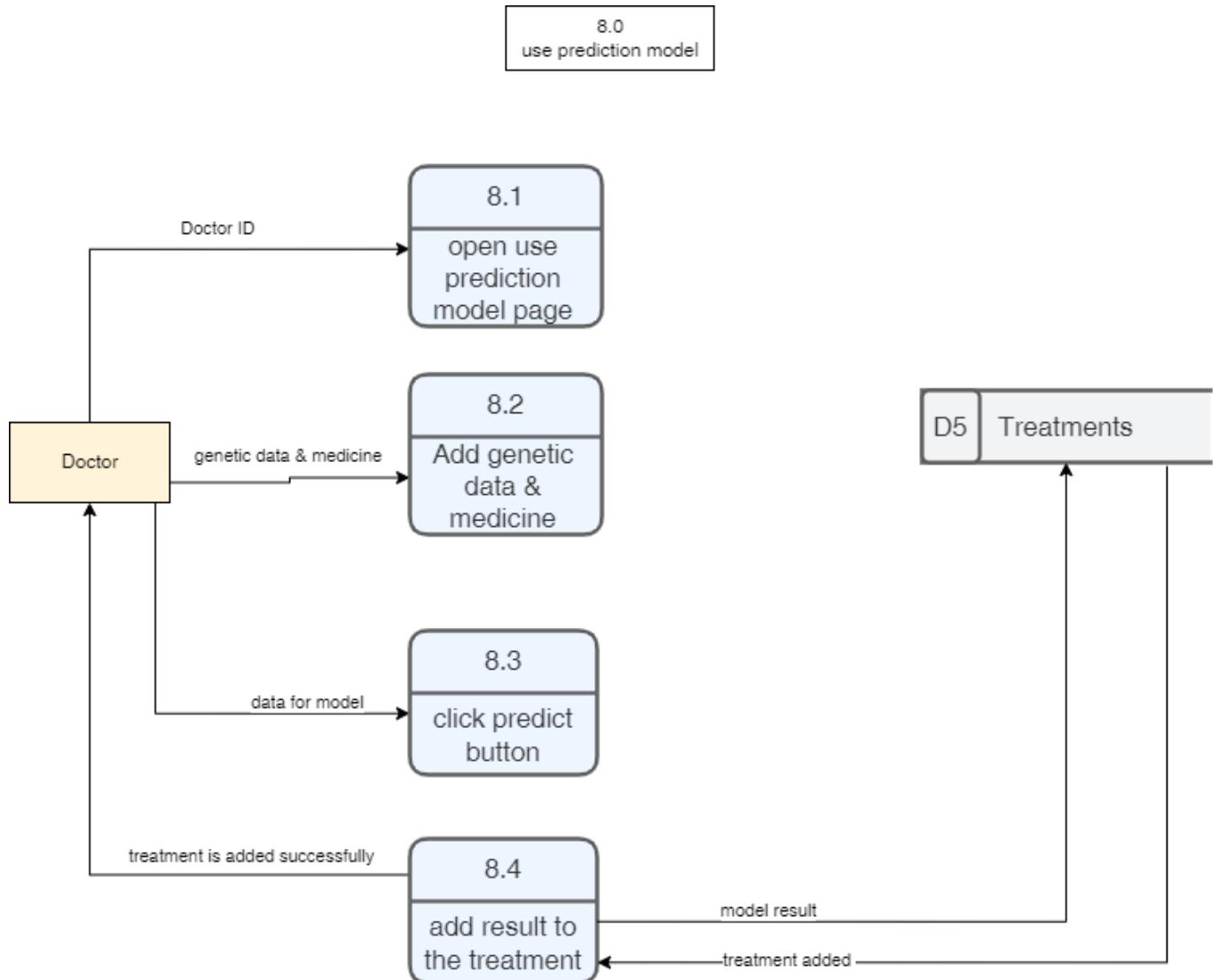


Figure 3_5: DFD level 2

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.

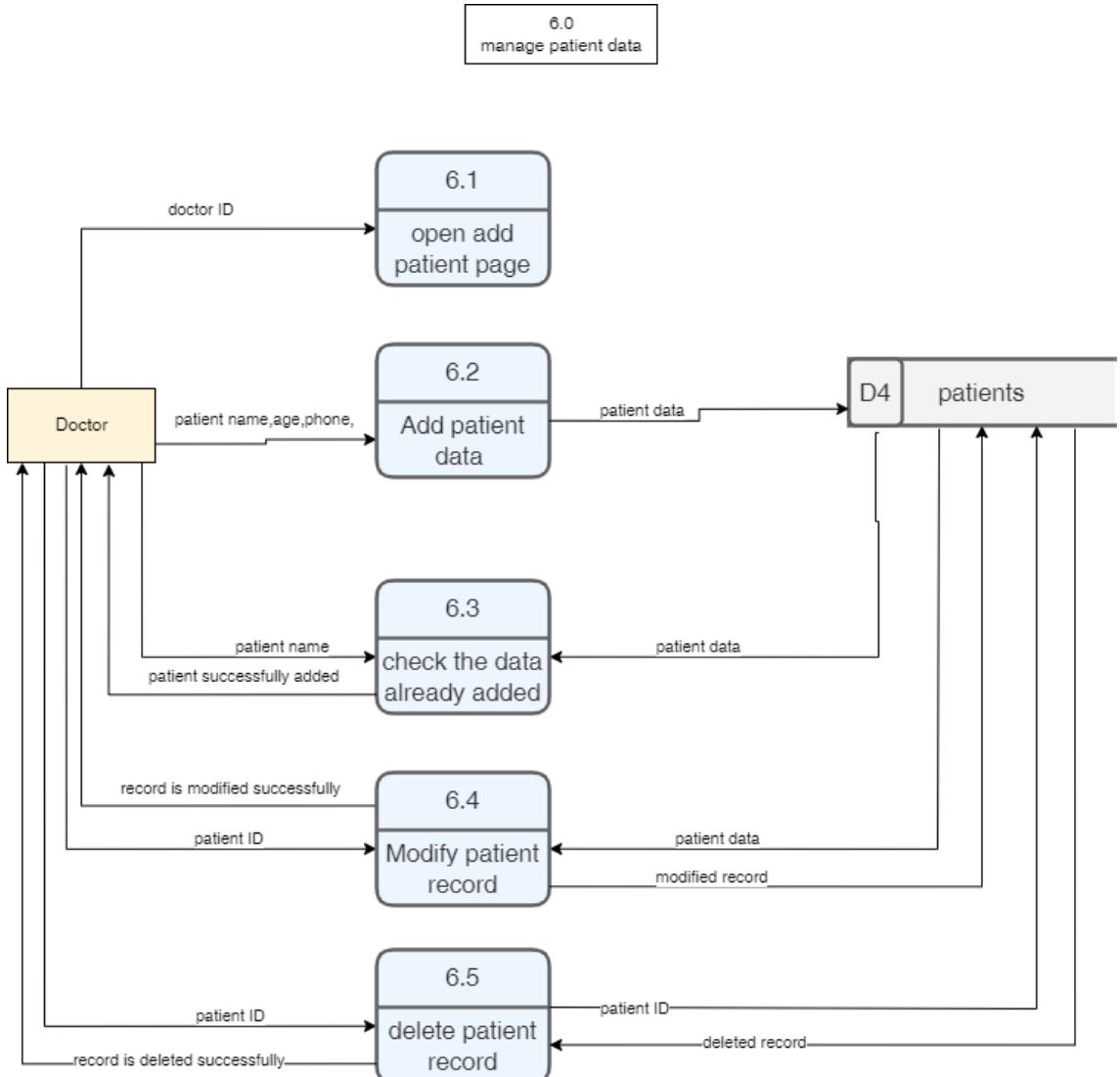


Figure 3_6: DFD level 2

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.

3.2.3 ACTIVITY DIAGRAM

3.2.3.1 DOCTOR

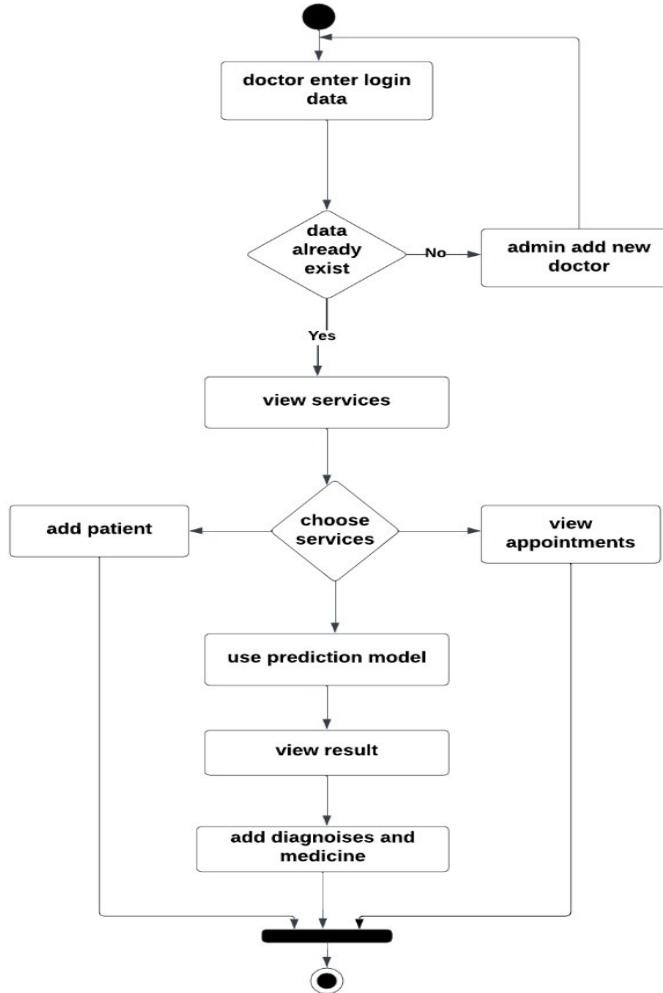


Figure 3_7: Doctor's Activity Diagram

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.

3.2.3.2 PATIENT

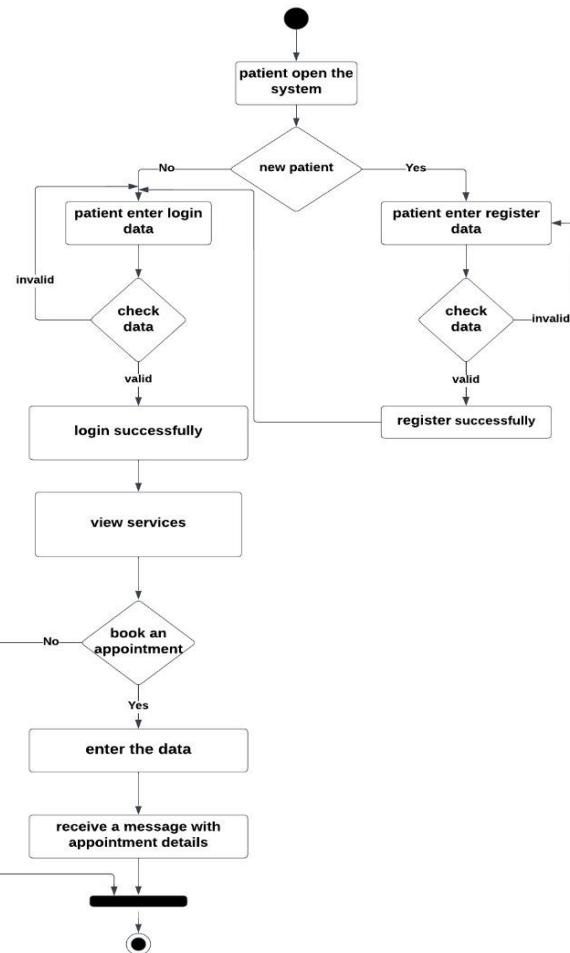


Figure 3_8: patient's Activity Diagram

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.

3.2.4 SEQUENCE DIAGRAM

3.2.4.1 PATIENT

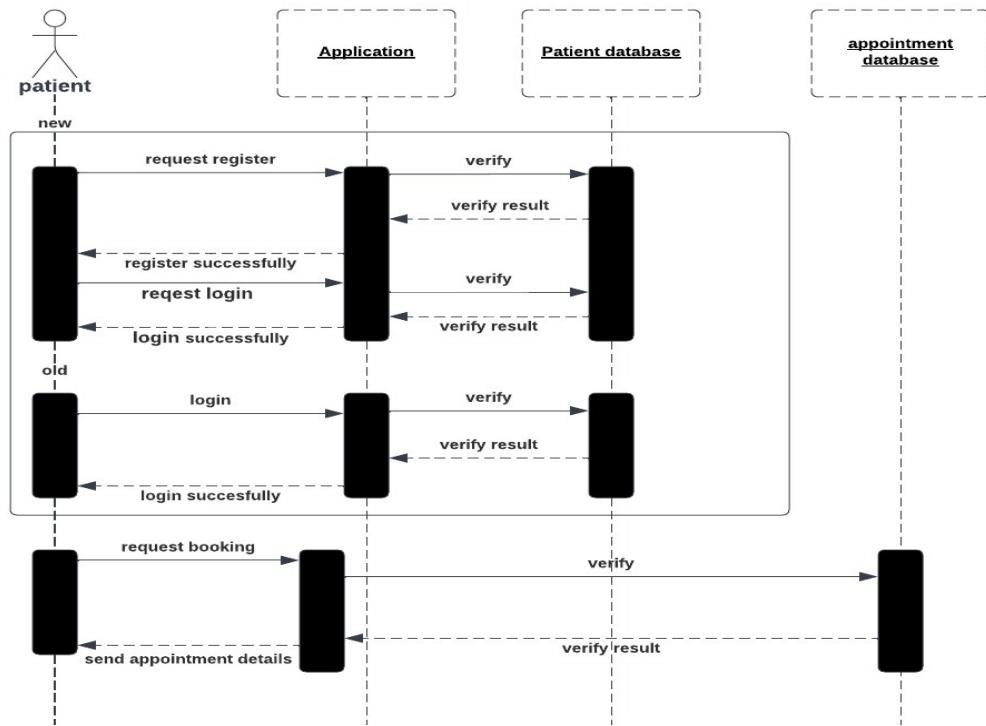


Figure 3_9: Patient's Sequence Diagram

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.

3.2.4.2 DOCTOR

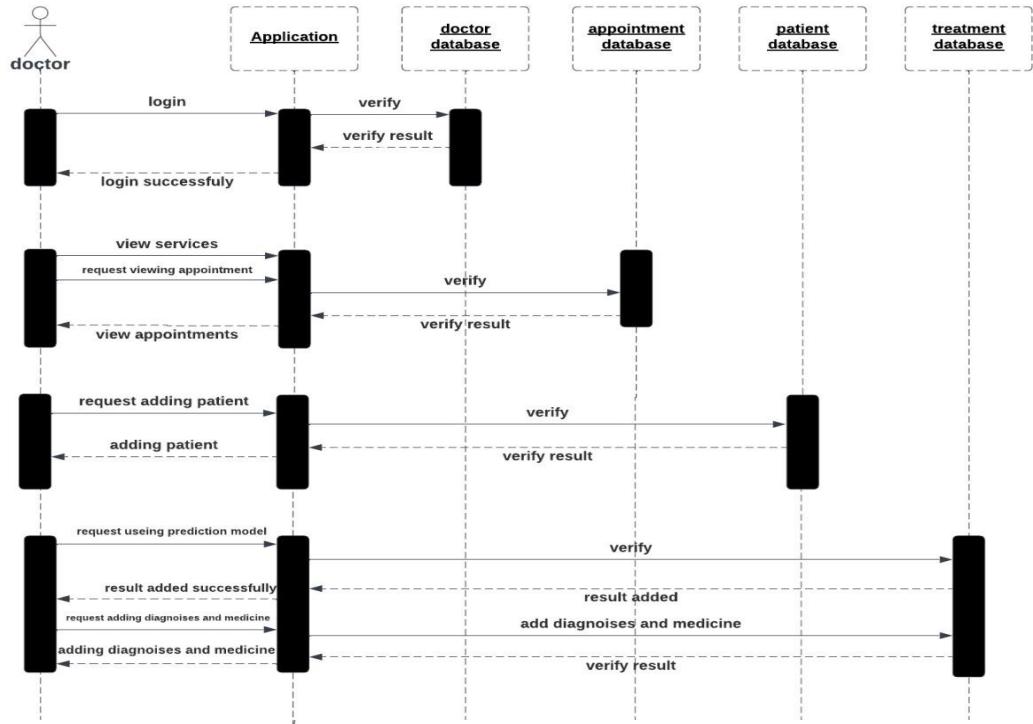


Figure 3_10: Doctor's Sequence Diagram

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.

3.2.5 ER DIAGRAM

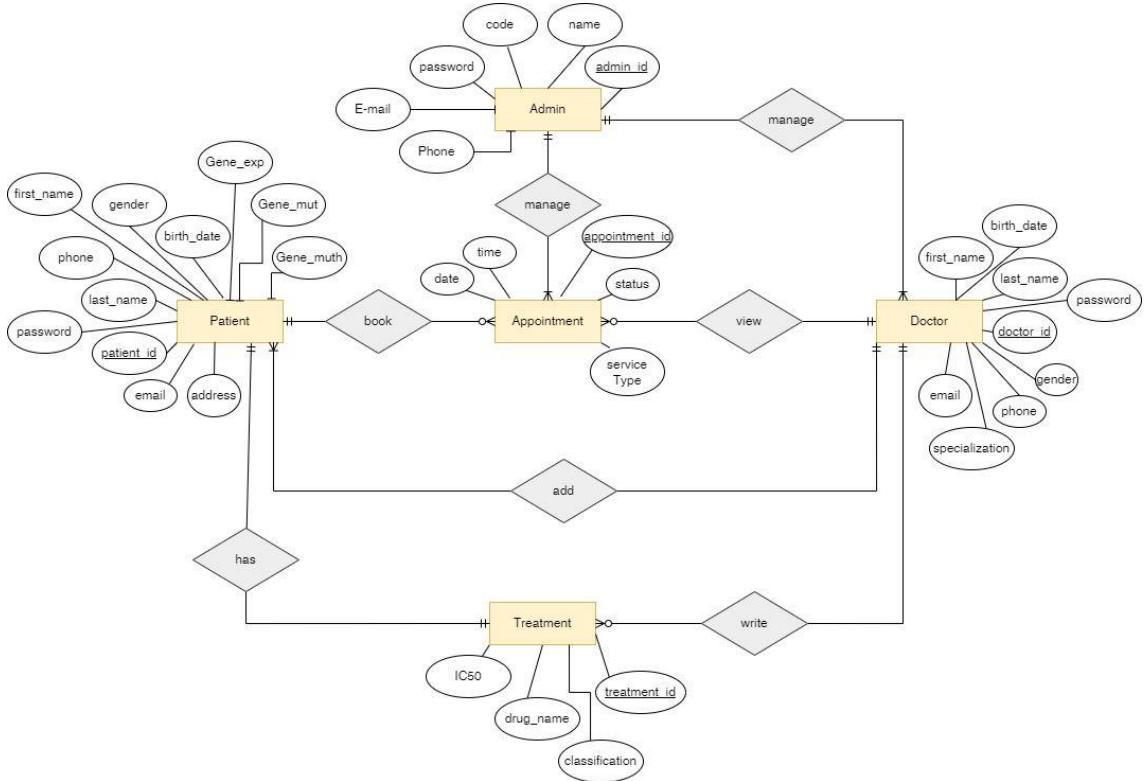


Figure 3_11: ER Diagram

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.

3.2.6 MAPPING

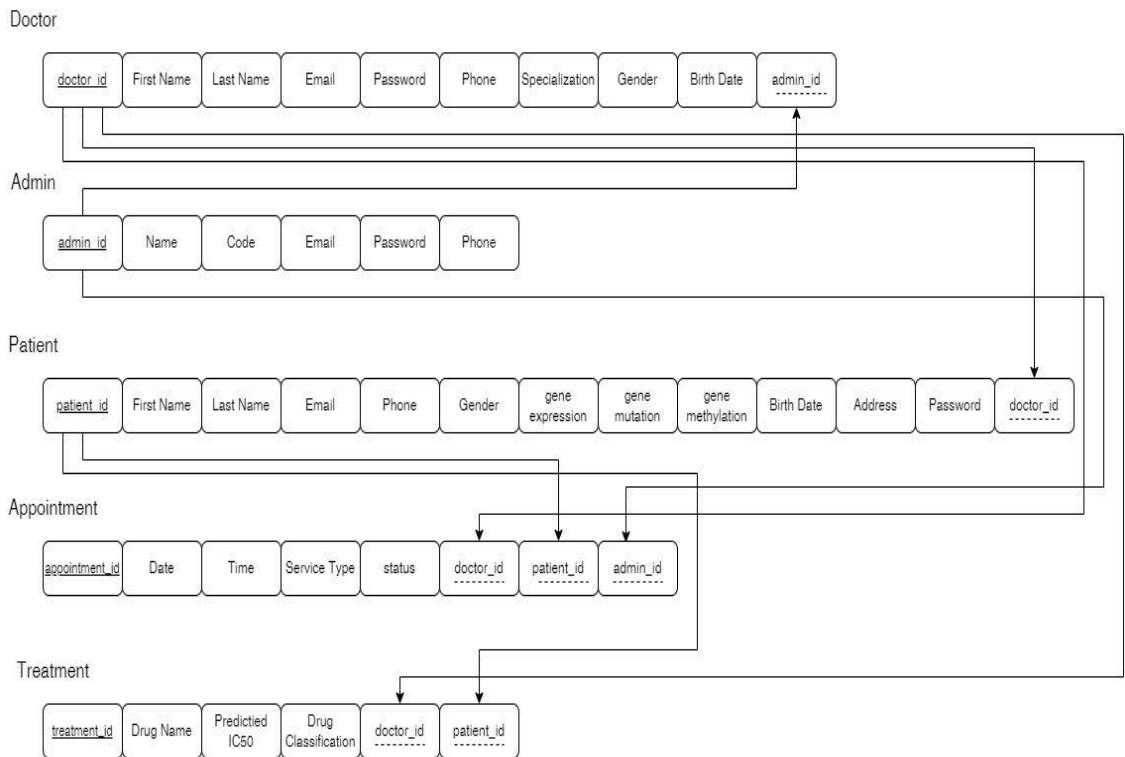


Figure 3_12: Mapping

Mapping: this figure explains the relationship between tables in database, what are the primary key and from which primary key are foreign keys derived, in our system main entities are (admin, patient, doctor, appointment, and treatment), mapping explain the primary and foreign key of each entity as shown in figure.

3.2.7 CLASS DIAGRAM

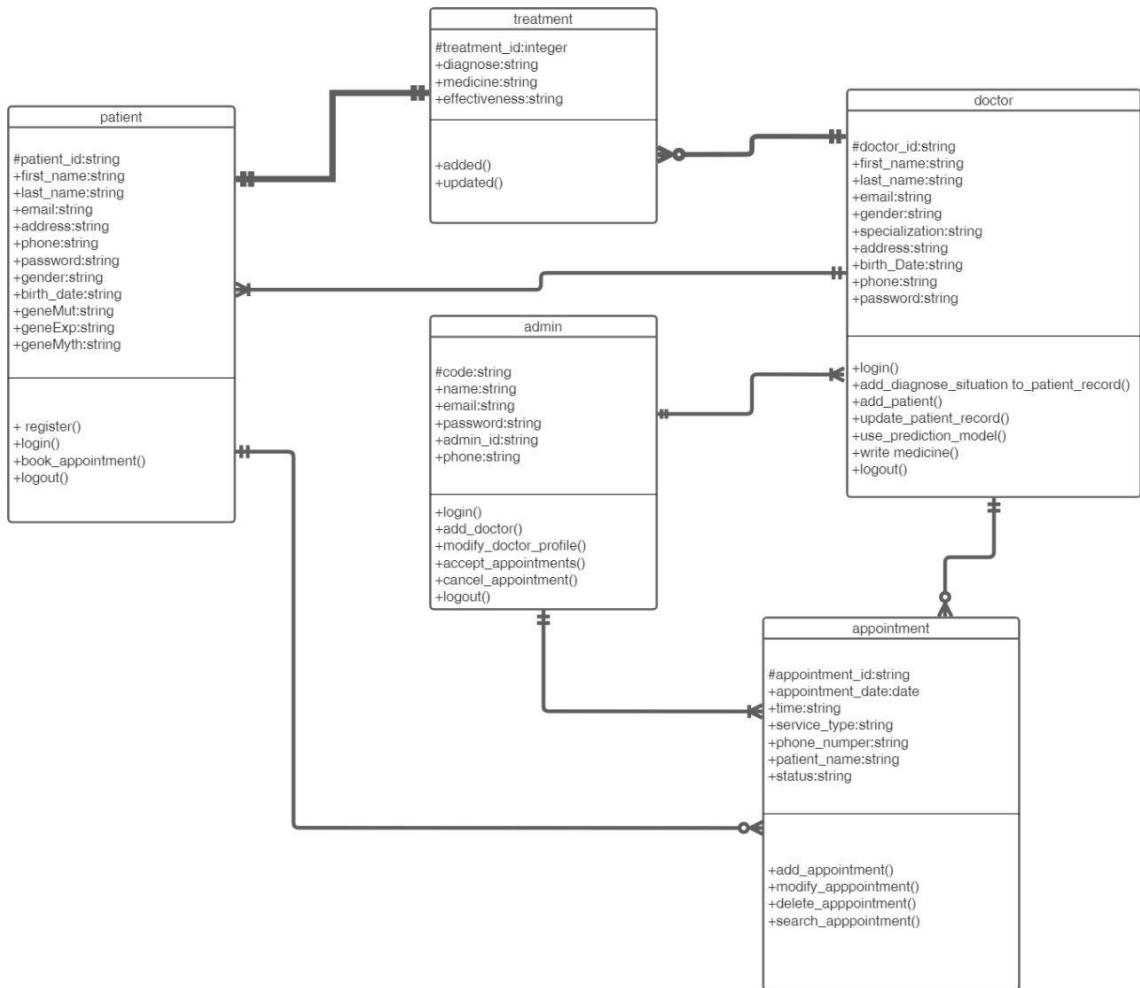


Figure 3_13: Class Diagram

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 cardinality such as mandatory, optional.

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

3.2.8 DATABASE:

We use Firebase, it is a mobile and web application development platform, owned by Google. It provides a suite of services that developers can use to build and manage their applications, including:

Real time Database: A cloud-hosted NoSQL database that allows developers to store and sync data in real time.

Authentication: A service that allows developers to add user authentication to their applications, including support for social media logins and two-factor authentication.

Storage: A service that allows developers to store and serve user-generated content, such as images and videos.

Machine Learning: A service that allows developer to upload machine learning model and save it as .tf lite.

3.2.8.1: AUTHENTACTION

Authentication is the process of verifying the identity of a user or device. Firebase Authentication is a service provided by Firebase that allows authenticating users to app using various authentication methods, such as email and password authentication, phone number authentication.

Firebase Authentication provides several benefits, such as:

Security: Firebase Authentication provides secure and reliable authentication methods, and handles the complexity of securely storing and transmitting user credentials.

User management: Firebase Authentication provides tools for managing user accounts, such as creating new users, deleting users, and resetting passwords.

Integration: Firebase Authentication can be easily integrated with other Firebase services, such as Cloud Fire store, Cloud Functions, and Cloud Storage.

The screenshot shows the Firebase Project Overview page with the 'Authentication' service selected. The main area displays a table of user accounts under the 'Users' tab. The columns include Identifier, Providers, Created, Signed In, and User UID. The table lists ten users, each with their email address, creation date, sign-in date, and a unique User UID.

Identifier	Providers	Created	Signed In	User UID
ahmed2@gmail.com	Email	May 29, 2023	May 29, 2023	aymen1UmYEHoX4aaNb31ocStAg...
ahmedhasin@gmail.com	Email	May 27, 2023	May 28, 2023	gWEAJoTrnGNSWt69xfdz9HyJ32...
nismar Ahmed@ocs.com	Email	May 22, 2023	May 29, 2023	LUbrGUYihhgBQ2G9A4Rv6xkzcJ2...
walidalsalihary@ocs.com	Email	May 21, 2023	May 21, 2023	zoiaGzvyDfMWkbaJLbcDKnwvCBRB2...
lbtsem23@ocs.com	Email	May 9, 2023	May 9, 2023	wFxBM6MfTS9GVDrpU0dxAoL...
eman91@ocs.com	Email	May 9, 2023	May 9, 2023	dd4OpUJmFTAgNvNx6fMpVaqV2...
nordan2023@ocs.com	Email	May 5, 2023	May 5, 2023	ICSeSsFHNBdHrWlqe0uJzea3G2...
selama2ged@ocs.com	Email	May 4, 2023	May 4, 2023	IDOMj4FqgBMeDCo2dBO7Gzf5fY1...
ahmed23@gmail.com	Email	May 4, 2023	May 4, 2023	rwe5Uaxc2YwOAVoqvhGeogy...
admin5234@ocs.com	Email	May 3, 2023	May 3, 2023	k7n0Jhmz7PauYZKTTA30s0tD7Z...
basmatelata756@gmail.co...	Email	May 2, 2023	May 2, 2023	XisVLy6csgefXctWwS1OfuPAWh1...

Figure 3_14: Authentication.

When a user signs in to our app using Firebase Authentication, Firebase returns a unique user ID token that we can use to identify and authenticate the user in our app.

3.2.8.2 REALTIME DATABASE:

It refers to the Real time Database service, which is a cloud-hosted NoSQL database that allows developers to store and sync data in real time. This means that any changes made to the database are immediately propagated to all connected clients, providing a seamless and responsive experience for users.

Firebase Real time Database uses a technique called data synchronization, which allows multiple clients to listen for changes to the same data and receive updates in real time. When a client makes a change to the database, Firebase synchronizes that change with other clients listening to that data. This allows all clients to have a consistent view of the data, regardless of where the changes originated.

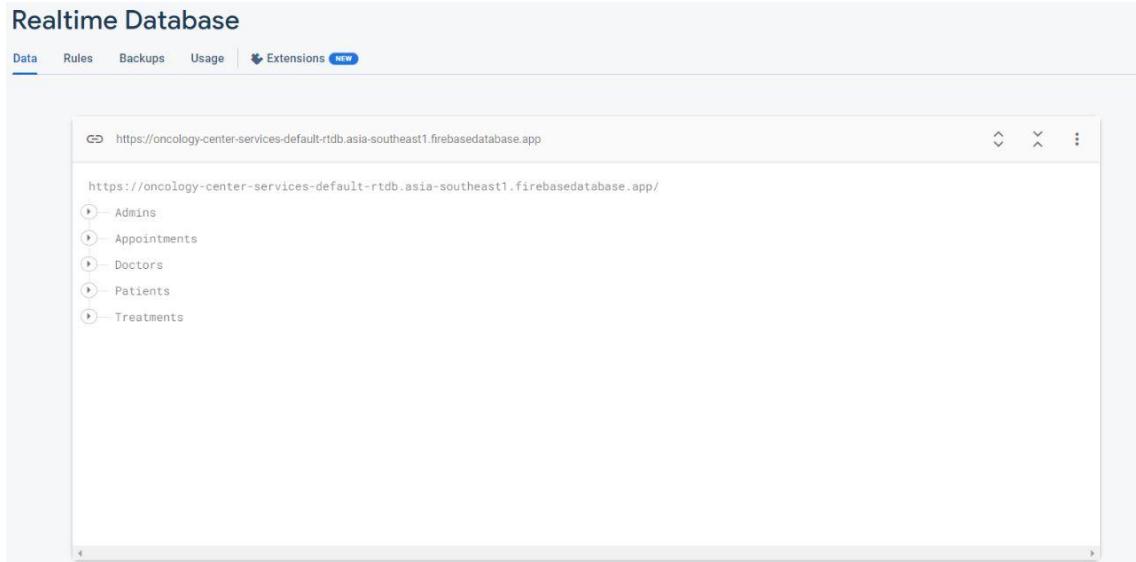


Figure 3_15: real time database.

Our real time database contains data about Admins, Appointments, Doctors, Patients and Treatments. When click on each of them its data will appear.

3.2.8.2.1 for doctors:

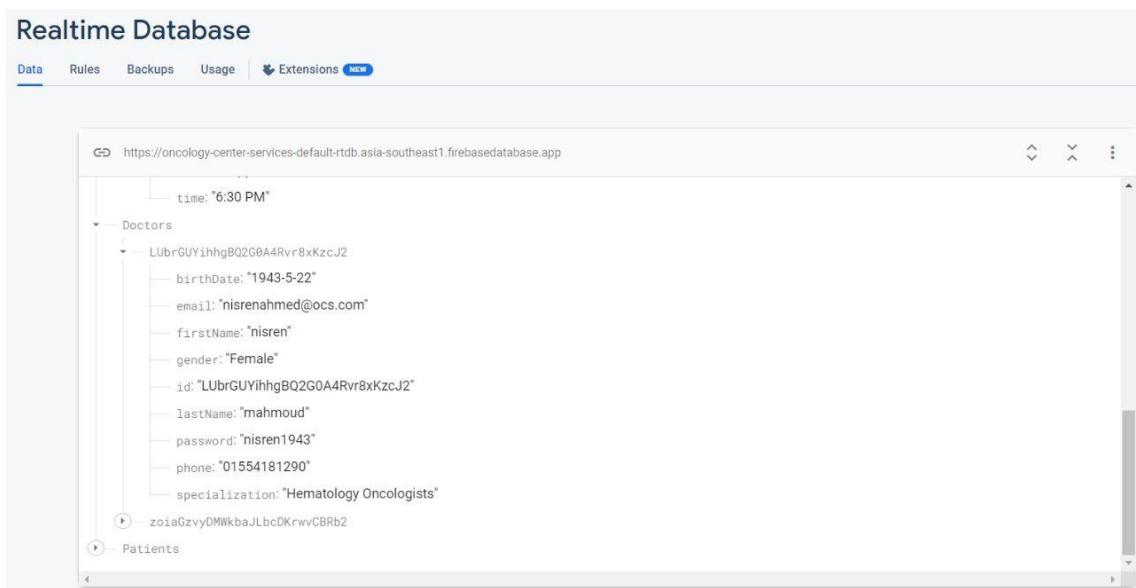


Figure 3_16: real time database for doctors.

It contains the doctor's ID, when we click on it, the doctor's information will appear such as (birthdate, email, first Name, gender, id, last Name, password, phone, specialization).

3.2.8.2.2: For admins

The screenshot shows the Firebase Realtime Database interface. At the top, there are tabs for Data, Rules, Backups, Usage, and Extensions. The URL in the address bar is https://oncology-center-services-default-rtbd.firebaseio.com/. The main view displays a hierarchical database structure under the path https://oncology-center-services-default-rtbd.firebaseio.com/. The structure includes a 'Admins' node with a single child node 'k7n0QhmZfPauYZKTTA303dDi7zs2'. This child node contains several fields: code ('23o4c9s87'), email ('admin5234@ocs.com'), id ('k7n0QhmZfPauYZKTTA303dDi7zs2'), name ('Ahmed Ibrahim'), password ('1galp02ws'), and phone ('01001006024'). Below the 'Admins' node, there are three other nodes: 'Appointments', 'Doctors', and 'Patients', each represented by a circular icon.

Figure 3_17: real time database for admins.

It contains the admin's ID, when we click on it, the admin's information will appear such as (code, email, id, name, password, phone).

3.2.8.2.3 For patients

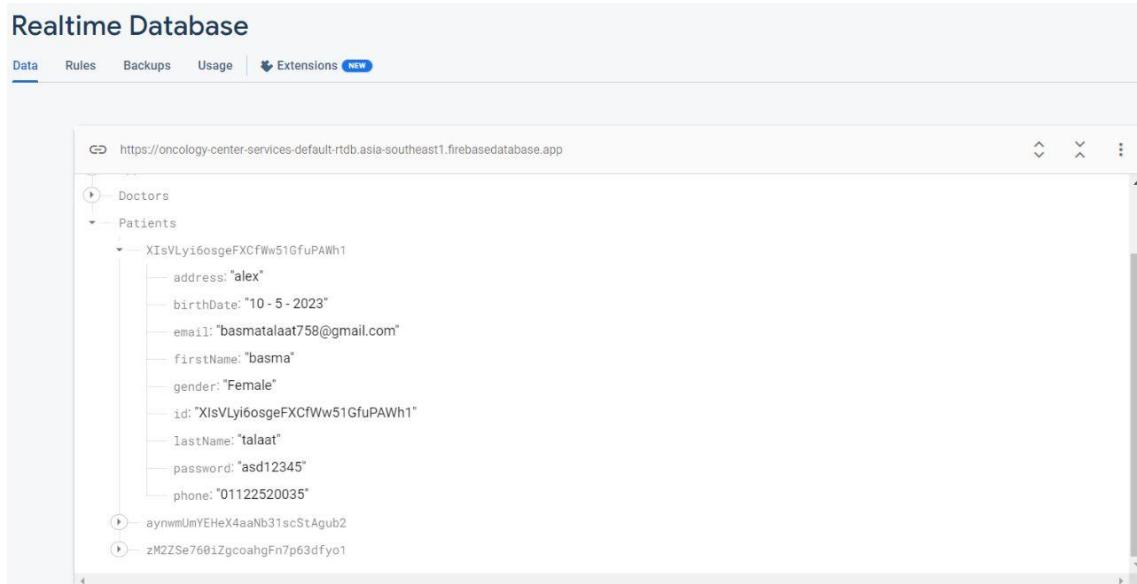


Figure 3_18: real time database for patients.

It contains the patient's ID, when we click on it, the patient's information will appear such as (address, birthdate, email, first Name, gender, id, last Name, password, phone).

3.2.8.2.4 for appointments

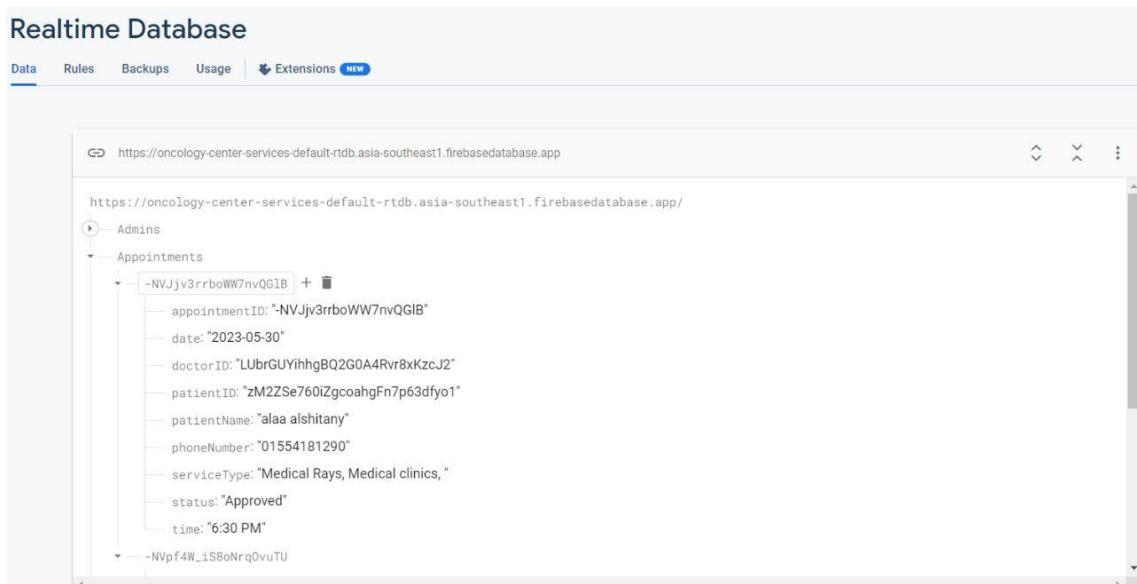


Figure 3_19: real time database for patients.

It contains the appointment's ID, when we click on it, the appointment's information will appear such as (appointment ID, date, doctor ID, patient ID, patient Name, phone Number, service Type, status, time).

3.2.8.2.5 for treatments

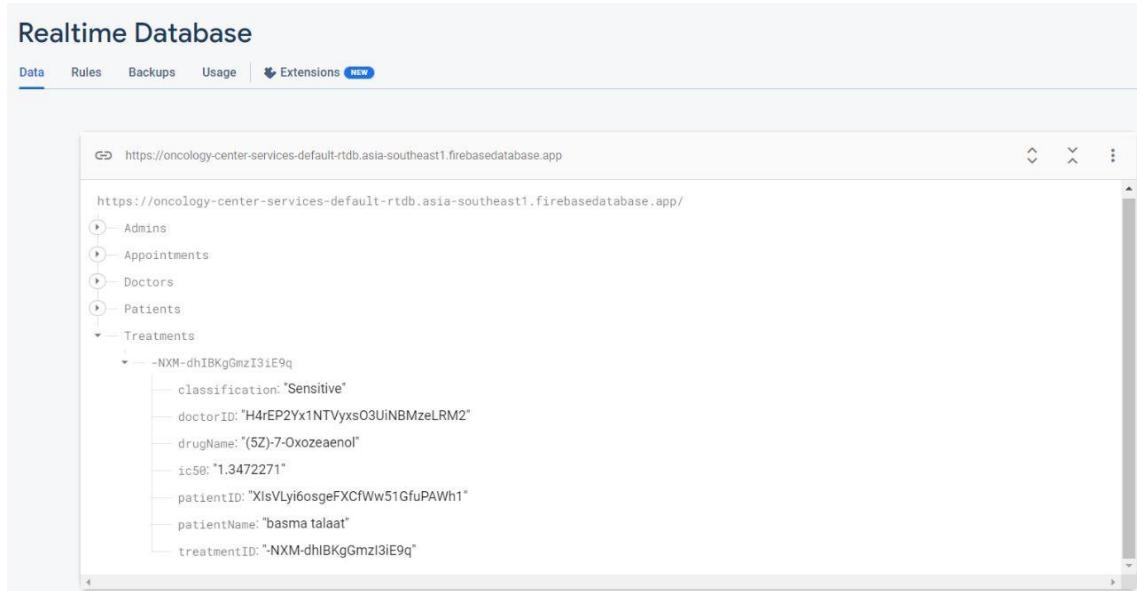


Figure 3_20: real time database for patients.

It contains the treatment's ID, when we click on it, the treatment's information will appear such as (classification, Drug Name, doctor ID, patient ID, patient Name, ic50, treatment ID).

3.2.8.3: Machine learning

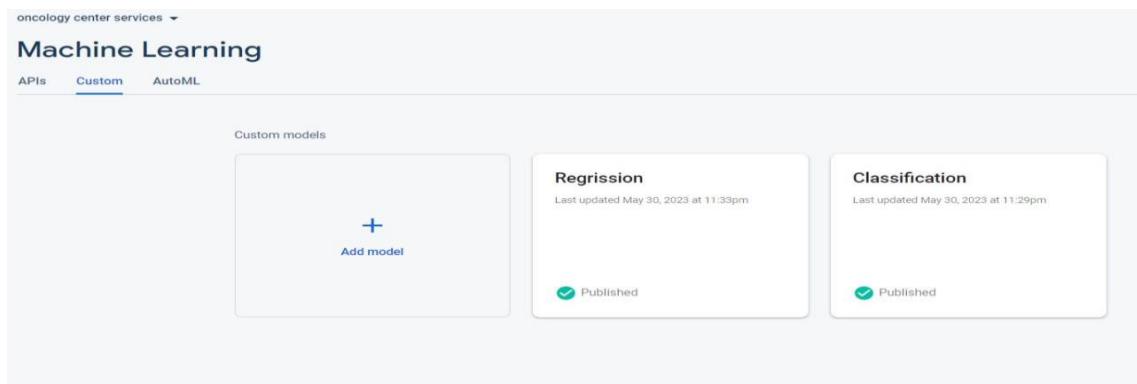


Figure 3_21: Machine learning models.

Firebase enables upload machine learning models and save it as .tf lite , there are our two models(regression and classification).

3.2.8.4 Storage

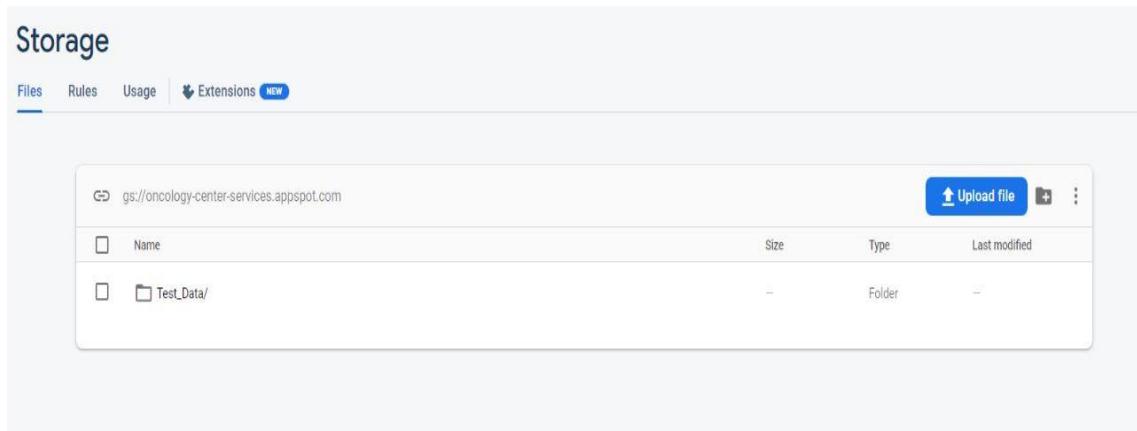


Figure 3_22: stored data.

When open storage this path will appear, and when open this path the following page will appear.

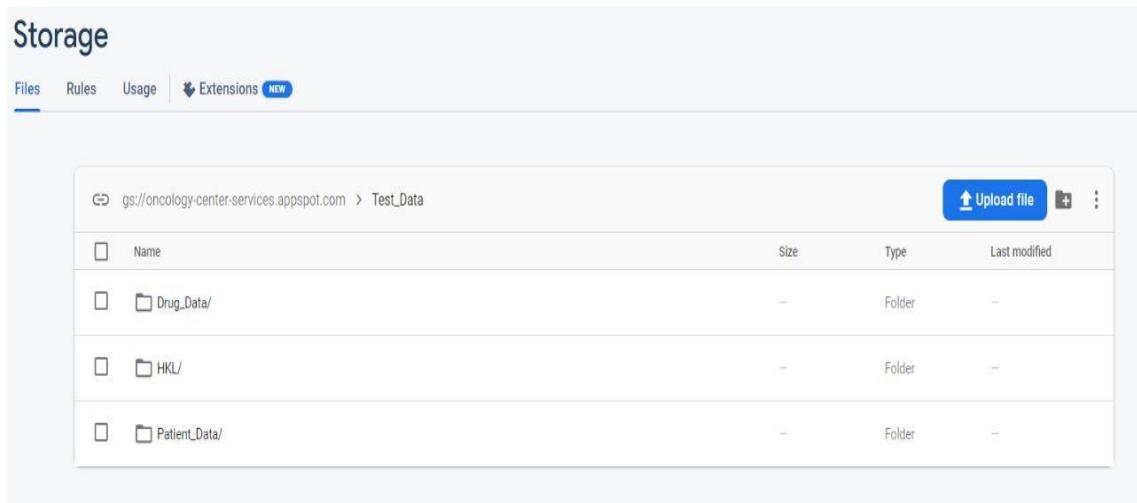
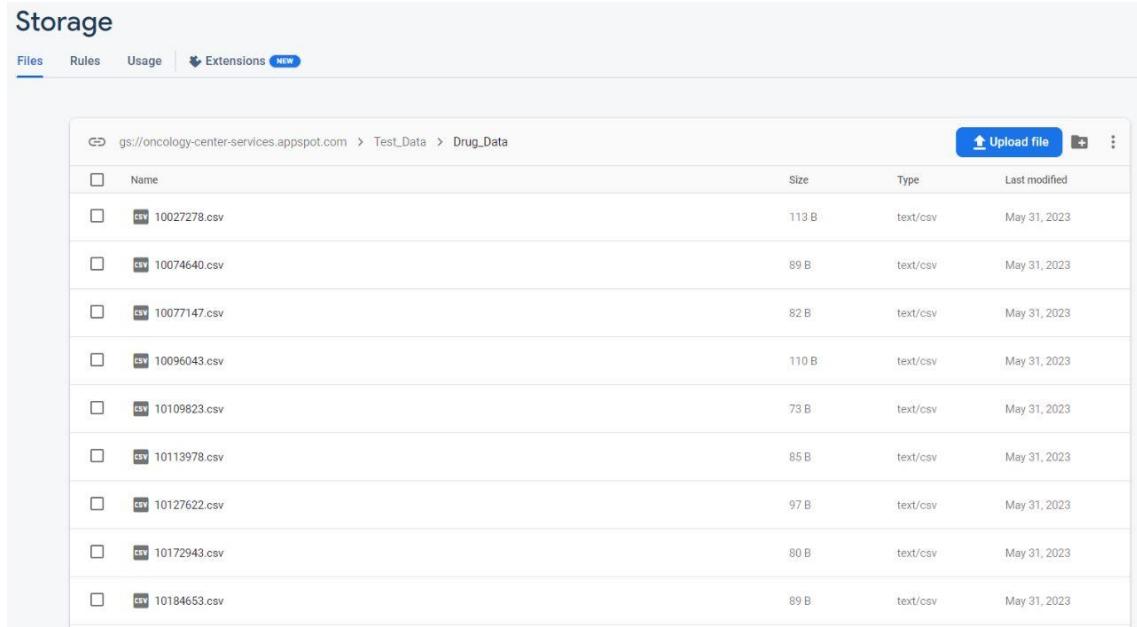


Figure 3_23: stored data.

It contains data files which consists of drug's files, HKL files and patient's files, we can upload or delete files.

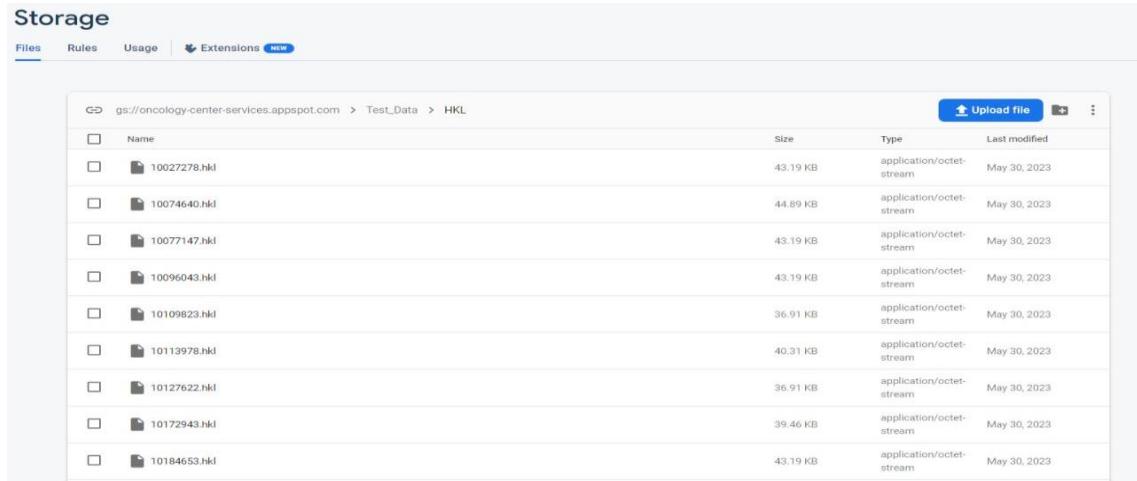


The screenshot shows a Google Cloud Storage interface. The path is gs://oncology-center-services.appspot.com / Test_Data / Drug_Data. There are ten CSV files listed:

Name	Size	Type	Last modified
10027278.csv	113 B	text/csv	May 31, 2023
10074640.csv	89 B	text/csv	May 31, 2023
10077147.csv	82 B	text/csv	May 31, 2023
10096043.csv	110 B	text/csv	May 31, 2023
10109823.csv	73 B	text/csv	May 31, 2023
10113978.csv	85 B	text/csv	May 31, 2023
10127622.csv	97 B	text/csv	May 31, 2023
10172943.csv	80 B	text/csv	May 31, 2023
10184653.csv	89 B	text/csv	May 31, 2023

Figure 3_24: drug files.

It contains drug files that each of them contain drugs information such as (ID, name, synonyms, targets, PubChem, sample size, count). We call these files during prediction.



The screenshot shows a Google Cloud Storage interface. The path is gs://oncology-center-services.appspot.com / Test_Data / HKL. There are ten HKL files listed:

Name	Size	Type	Last modified
10027278.hkl	43.19 KB	application/octet-stream	May 30, 2023
10074640.hkl	44.89 KB	application/octet-stream	May 30, 2023
10077147.hkl	43.19 KB	application/octet-stream	May 30, 2023
10096043.hkl	43.19 KB	application/octet-stream	May 30, 2023
10109823.hkl	36.91 KB	application/octet-stream	May 30, 2023
10113978.hkl	40.31 KB	application/octet-stream	May 30, 2023
10127622.hkl	36.91 KB	application/octet-stream	May 30, 2023
10172943.hkl	39.46 KB	application/octet-stream	May 30, 2023
10184653.hkl	43.19 KB	application/octet-stream	May 30, 2023

Figure 3_25: HKL files.

It contains HKL files, the HKL file is an important tool for understanding the structure and function of proteins and for designing new drugs that target specific proteins.

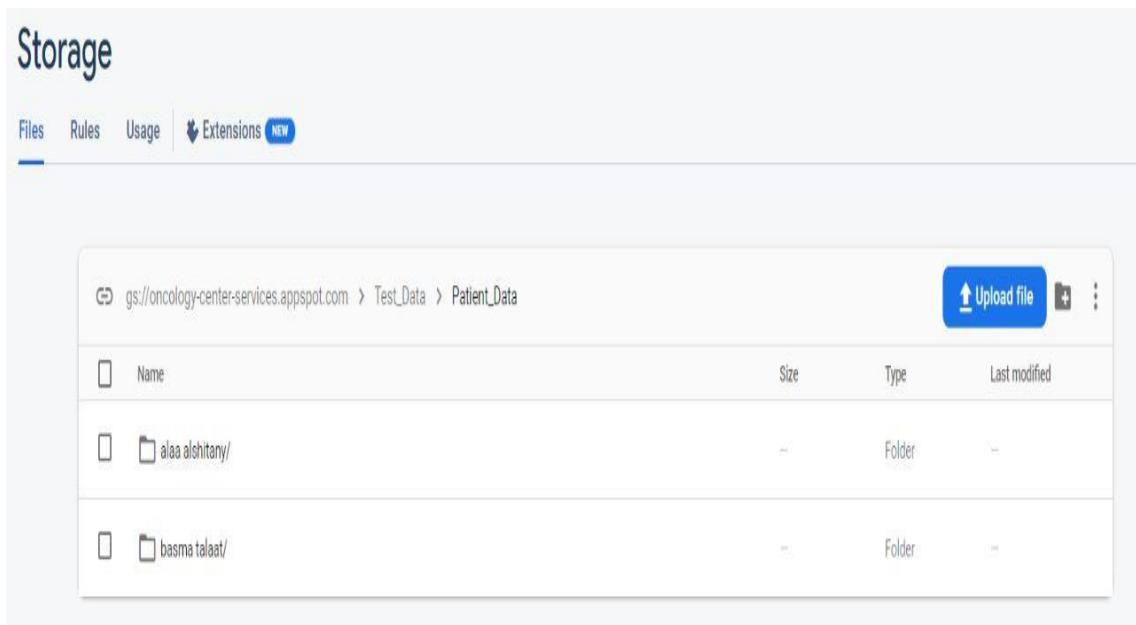


Figure 3_26: Patient's files.

It contains patient's files (gene expression, gene mutation, gene methylation, we call these files during prediction.

Chapter Four

4 MODEL

4.1 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.

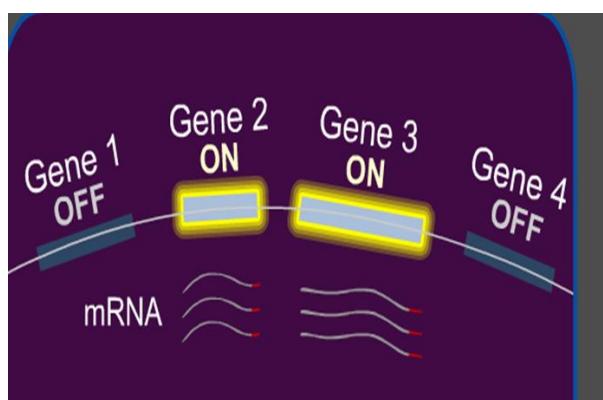


Figure 4_1: Genomic expression

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T			
1	LASP1	HOXA11	CREBBP	ETV1	GAS1	CD79B	PAX7	BTK	BRCA1	WAS	WWTR1	CD74	BIRC3	FAS	BCLAF1	ANK1	RABEP1	ZCHCB8				
2	ACH-000828	9.393476266	0.042644	9.39546	0.871844	0.070389	0.084064	0.055684	3.339137	0.15056	2.731183	0.378512	0.263034	1.226509	5.836934	0.137504	5.53512	4.162693	5.486714			
3	ACH-000568	7.630873837	0.056584	3.42760	2.01634	1.794936	0.739848	0.042644	0.333424	3.193772	0.185576	2.65257	0.405992	0.268861	0.62293	5.306335	0.157023	4.406673	5.243364	5.697107		
4	ACH-000560	5.728192635	0.11302	0.53254	0.51082	0.432959	0.250962	0	0.263034	4.678635	1.292782	0.575312	1.735522	0.545966	1.257044	0.04644	0.26646	0.506235	0.376478	5.82171		
5	ACH-000561	6.03716255	1.565597	4.262283	0.790772	1.257011	0.028569	0.056584	0.042644	3.44228	0.288681	4.84549	0.695994	0.250962	0.9855	0.018367	0.650765	4.19456	4.370164	5.28762		
6	ACH-000562	7.050501697	0.482654	3.08936	0.879704	0.046804	0.175304	0	0.042644	4.939227	0.286881	5.040382	3.080391	5.848247	0.482391	0.610877	0.09771	4.733648	5.415639	5.649061		
7	ACH-000563	7.795195491	0.014355	4.270529	4.14323	0.189034	0.189034	0.028569	0.214245	5.600495	0.070389	7.686837	7.485507	3.252476	3.392317	5.975447	0.084064	4.873349	5.088311	5.506843		
8	ACH-000565	6.716715692	2.784504	3.60058	1.063503	0.895303	0.298658	0.042644	1.63499	3.997285	0.562326	9.181386	4.012569	2.93546	1.891419	5.428946	0.084064	3.544773	4.596339	5.492494		
9	ACH-000566	5.708507884	1.646163	4.249445	2.74416	0.238781	1.124328	0	0.263034	4.140779	0.879706	6.730504	5.950702	3.134044	3.319084	5.180819	2.849384	4.409391	3.4397	4.741112		
10	ACH-000567	5.441616269	0.056584	5.025029	0.787512	2.179511	0.762183	0.028569	5.76914	0.940426	5.672078	1.09711	10.79401	3.611172	2.339317	6.895871	0.695994	5.878321	5.412104	5.649911		
11	ACH-000228	7.16510785	1.298569	4.714167	0.58144	1.077243	0.464624	1.459432	0.097611	3.578939	0.31034	4.020591	6.419395	5.548437	3.325317	0.01221	1.83777	3.871844	4.140779	5.38405		
12	ACH-000768	8.157447996	0.028569	3.849994	4.864434	0.189034	0.15056	0	0.176323	4.059243	0.9854	5.744336	7.210263	4.163494	1.984601	5.57622	2.933573	4.762349	4.341274	5.301953		
13	ACH-000197	6.1598171337	0.044604	4.791209	0.21425	1.506078	0.87525	0.028569	4.329529	5.296824	0.6080231	0	0.042647	2.074827	3.275735	4.385434	0.262509	7.00304	0.050684	4.863938	4.190159	5.839209
14	ACH-000196	5.199279721	0.9855	4.526695	0.15056	0.263034	1.275007	0	0.163499	4.559492	0.275007	6.141392	0.130562	0.048601	1.327687	6.429114	1.88034	4.135863	5.154616	5.286142		
15	ACH-000191	6.08512718	0.3	3.372952	3.715893	0.356144	0.464668	0	0.124238	3.099295	0.261588	5.576824	5.719113	2.723739	0.606047	3.04166	0.174398	5.781529	5.175525	5.175525		
16	ACH-000190	5.632323076	0.432959	3.341986	5.257011	0.124238	0.137504	0	0.124238	4.308885	0.060471	6.994014	6.982415	4.773996	4.280956	0.269033	0.049761	4.787384	4.028569	5.751555		
17	ACH-000198	4.95930649	2.111031	4.017922	0.15056	4.527321	1.85997	0	0.685997	4.647279	0.712457	10.6749	0.298658	10.6749	0.298658	15.45411	0.15056	4.855491	0.527685	5.209453		
18	ACH-000222	7.23112518	2.753894	4.748372	4.286881	0.080464	0.080464	0.014355	0.028569	2.891976	0.214255	4.103961	1.104337	4.574102	3.658783	5.469105	2.126173	4.043519	4.600508	6.216206		
19	ACH-000298	6.468583317	0	4.719188	1.531069	0.014355	0.084064	0.014355	3.13104	4.690417	0.176322	2.803227	1.618239	6.117072	2.849994	4.387441	5.562853	8.717367	5.287624			
20	ACH-000294	6.798569004	8.213525	3.60288	0.042644	3.31794	3.872829	0.014355	5.964861	4.755955	5.26791	0.434607	1.367371	6.136352	7.265735	0.245670	0.07854	4.75942	0.082788	6.234193		
21	ACH-000295	7.0745695697	0.056057	3.894856	0.545968	3.755956	0.536053	0	5.93121	2.625352	0.505165	0.411426	1.405992	0.040533	2.996389	7.048759	0.014355	4.393691	4.530563	5.552746		
22	ACH-000292	6.231701199	1.20571	3.642702	2.805292	0.344828	0.014355	0	0.042644	4.509661	0.078196	2.845992	6.322686	2.788688	2.797013	6.669504	0.15056	5.539662	4.767977	6.114575		
23	ACH-000290	5.466953337	1.823749	1.847997	2.475085	0.028569	0.411426	1.634994	0.042644	3.052404	0.575131	1.116795	0.584963	2.382787	0.042644	6.05398	4.862947	5.682011	5.139551	6.873075		
24	ACH-000291	7.254754198	0.59955	3.702658	1.144046	0.15056	0.124238	0	0.250962	3.529821	0.15056	4.279795	0.495695	0.485427	1.763323	4.038775	0.07203	4.114367	4.320485	5.771093		
25	ACH-000667	6.074676686	1.704872	3.347666	2.0695	0.111031	0.056584	0	0	3.049631	0.137054	5.074249	7.209063	3.992768	1.769772	6.101018	0.378512	4.522307	3.966246	5.062208		
26	ACH-000665	7.525207571	0.536053	3.674687	4.849949	0.014355	0.084064	0	0.084064	4.314697	0.124238	6.510487	5.735795	5.093814	3.214125	5.74086	4.659952	3.729009	4.294253	5.979111		
27	ACH-000664	4.503384739	0.14355	3.43844	3.98008	0.060288	0.304593	0.014355	5.360633	6.473754	3.87576	2.98658	9.100308	4.313971	3.351951	6.359361	0.479129	4.539361	4.791291	6.257386		

Figure 4_2: Genomic expression dataset

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

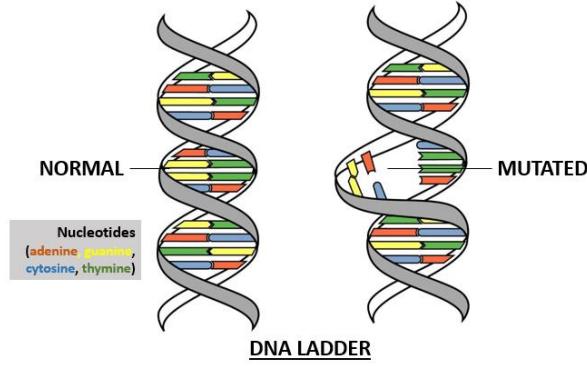


Figure 4_3: Genomic Mutation

Figure 4_4: Genomic Mutation dataset

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).

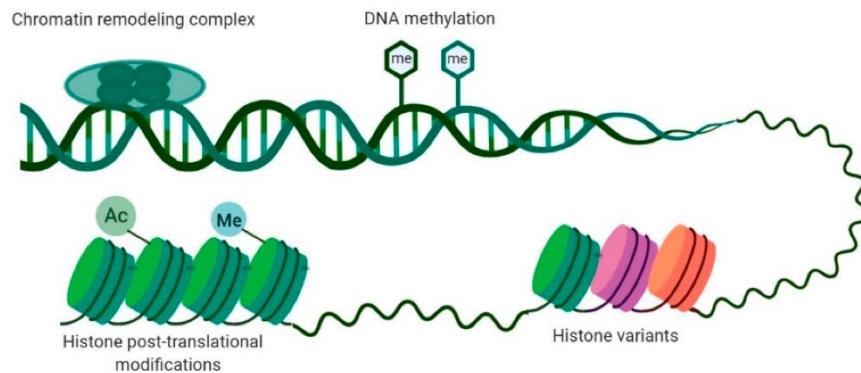


Figure 4_5: Genomic Methylation

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U		
1	SKI_1_215	TNFRSF14	PRDM16	RPL22_1_(CAMTA1_`MTOR_1_`PRDM2_1	PRDM2_1	CASP9_1_1585	CASP9_1_1585	SPEN_1_1	SDHB_1_1	ARHGEF10	ARHGEF10	PAX7_1_1	MDS2_1_2	ARID1A_1	LCK_1_32	THRAP3_1	THRAP3_1					
2	ACH-000698	0	0.00739	0.04799	0.14687	0.00553	0.54138	0.0716	0.0004	0.01836	0.03461	0	0	0.14391	0.38605	0.00223	0.5833	0.00355	0.7658	0	0	
3	ACH-000489	0	0.00363	0.64891	1	0.00538	0	0	0	0	0	0.0002	0	0.10881	0.53469	0.51547	0.00534	0	1	0	0	
4	ACH-000522	0.00143	0.47465	0.27715	0.19441	0.00858	0.40555	0.21625	0.01367	0.00338	0.00338	0.00656	0	0.20194	0.88821	0.61515	0.94318	0.00664	0.973	0.0195	0.0195	
5	ACH-000613	0	0	0.05112	0.10686	0.00268	0.28618	0.07706	0	0	0	0.00383	0	0.38074	0.86867	0.9214	0.01066	0.00686	0.9792	0	0	
6	ACH-000614	0.00118	0.00471	0.15165	0.07407	0.0105	0.51629	0.03292	0.00405	0	0	0.0037	0	0.06404	0.9174	0.32482	0.01795	0.00831	0.9815	0.0238	0.0238	
7	ACH-000691	0.02257	0.5525	0.96289	0.16461	0.00295	0.02677	0.07717	0.00167	0.01129	0.00905	0.01258	0.04858	0.79532	0.91827	1	0.00354	0.8491	0.00303	0.00303		
8	ACH-000610	0.00837	0.43893	0.01333	0.07799	0.0025	0.08723	0.00792	0	0.00329	0.00262	0.00214	0	0.01527	0.52772	0.0501	0.73431	0.00162	0.9344	0.0057	0.0057	
9	ACH-000420	0	0.34969	0.08986	0.14393	0.00233	0.02264	0.36857	0.0693	0.00774	0.00774	0.00391	0	0.15517	0.92885	0.27456	0.93941	0.00871	0.987	0.0245	0.0245	
10	ACH-000648	0	0.25533	0.35047	0.24272	0.00314	0.19533	0.77779	0.02689	0.00762	0.00762	0.0079	0.00489	0.30207	0.54446	0.44309	1	0.02019	1	0.04491		
11	ACH-000091	0.02507	0.04126	0.16667	0.01028	0.01784	0.0093	0	0.01088	0.01088	0.00354	0.03125	0.12944	0.90878	0.35063	0.01296	0.00198	1	0.0283	0.0283		
12	ACH-000595	0.00205	0.04518	0.21184	0.31788	0.00119	0.28052	0.63486	0.00202	0	0	0.00147	0	0.046	0.74267	0.06575	0.8025	0.00128	0.975	0	0	
13	ACH-000770	0	0.30797	0.29483	0.17019	0.00302	0.05504	0.09624	0.01518	0	0.1076	0.01184	0.00407	0.28346	0.86883	0.52583	0.38142	0.03041	1	0.01772	0.01772	
14	ACH-000505	0.00085	0.46132	0.27549	0.13298	0.04962	0.2014	0.02507	0.03067	0	0	0.00737	0	0.11787	0.94261	0.41167	1	0.00296	0.9348	0.19809	0.19809	
15	ACH-000609	0	0.14395	0.23674	0.07692	0.01103	0.3109	0.45678	0.09645	0	0	0	0	0.1507	0.98496	0.44123	0.04858	0	1	0.01558	0.01558	
16	ACH-000974	0	0.56128	0.07978	0.37318	0.00759	0.30233	0.31681	0.41011	0.06644	0.06644	0.01777	0	0.40952	0.5627	0.26797	0.05705	0.0073	1	0.0076	0.0076	
17	ACH-000875	0.00127	0.01471	0.18349	0.16667	0.01665	0.01116	0	0	0	0	0	0	0.17448	0.97222	0.6843	0.05979	0.00808	0.9516	0.14595	0.14595	
18	ACH-000155	0.00036	0.00456	0.04571	0.25755	0.00442	0.14718	0.05354	0	0	0	0.00186	0	0.32723	0.85463	0.49496	0	0.0129	1	0	0	
19	ACH-000135	0	0.03566	0.02906	0.16058	0	0.30612	0.01044	0	0.02095	0.02095	0	0	0.0349	0.67022	0.10244	0.00757	0.0004	0.9294	0	0	
20	ACH-000144	0.01392	0.18956	0.03746	0.25773	0.00525	0.07063	0	0	0	0.0012	0	0.04559	1	0.89174	0.0125	0.06024	1	0	0	0	
21	ACH-000054	0	0.18786	0.39127	0.17483	0.01235	0.36029	0.09712	0.00226	0	0	0.01551	0	0.17148	0.84776	0.01267	0.01265	0	1	0	0	
22	ACH-000841	0.07678	0.2357	0.29994	0.28458	0.06658	0.56756	0.03621	0.05507	0.01206	0.01206	0.01601	0.0562	0.20662	0.9734	0.835	0.42106	0.05864	1	0.09685	0.09685	
23	ACH-000451	0	0.08899	0.02953	0.30052	0	0	0	0	0	0	0	0	0.43718	0.00527	0.04407	0.5952	0	0	0	0	
24	ACH-000871	0.00218	0.08585	0.03458	0.05611	0.00228	0.00522	0.00136	0.01054	0.01092	0.01092	0.02253	0.03261	0.13796	0.80295	0.0202	0.76692	0.01458	0.7778	0	0	
25	ACH-000433	0	0.06421	0.22346	0.9091	0	0.31815	0.28733	0.00389	0	0	0.00283	0	0.08534	0.71592	0.13376	1	0.00264	0.00762	0	0	
26	ACH-000491	0.00113	0.22859	0.13157	0.06367	0.00098	0.03005	0.00786	0.02947	0	0	0.00225	0	0.17382	0.08595	0.06473	0.8333	0.01135	1	0.03656	0.03656	
27	ACH-000727	0	0.05955	0.0459	0.25786	0.00718	0.05847	0.00329	0.03231	0.02982	0.02982	0.00406	0	0.08249	0.61041	0.25912	0.01016	0.00548	1	0.2521	0.2521	

Figure 4_6: Genomic methylation dataset

GDSC database:

A	B	C	D	E	F	G	H
drug_id	Name	Synonyms	Targets	Target pathway	PubCHEM	Sample Siz	Count
1	1242 (S2)-7-Oxozaenol	S2-7-Oxozaenol, LL-Z1640-2	TAK1	Other, kinases	9863776	945	266
2	179 5-Fluorouracil	5-FU	Antimetabolite (DNA & RNA)	Other	3385	968	266
4	86 A-443654	KIN001-139	AKT1, AKT2, AKT3	PI3K/MTOR signaling	10172943	425	266
5	55 A-770041	KIN001-111	LCK, FYN	Other, kinases	9549184	426	266
6	1001 AICA Ribonucleotide	AICAR, N1-(b-D-Ribofuranosyl)-5'-AMPK agonist		Metabolism	65110	872	266
7	171 AKT inhibitor VIII	Akti-1/2	AKT1, AKT2, AKT3	PI3K/MTOR signaling	10196499	934	266
8	228 AKT inhibitor VIII	Akti-1/2, KIN001-102	AKT1, AKT2, AKT3	PI3K/MTOR signaling	10196499	979	266
9	272 AR-42	HDAC-42, AR 42, AR42	HDAC1	Chromatin histone acetylation	6918848	965	266
10	207 AS601245		JNK1, JNK2, JNK2	JNK and p38 signaling	10109823	925	266
11	224 AS605240	KIN001-173, AS-605240	PI3Kgamma	PI3K/MTOR signaling	5289247	975	266
12	219 AT-7519	AT7519	CDK1, CDK2, CDK4, CDK6, CDK9	Cell cycle	11338033	976	266
13	29 AZ628	AZ-628, AZ 628	BRaf	ERK MAPK signaling	11676786	428	266
14	156 AZD6482	AZD 6482, AZD-6482, AK-55404	PI3Kbeta	PI3K/MTOR signaling	44137675	933	266
15	1066 AZD6482	AZD 6482, AZD-6482, AK-55409	PI3Kbeta	PI3K/MTOR signaling	44137675	931	266
16	1022 AZD7762	AZD-7762, AZD 7762	CHEK1, CHEK2	Cell cycle	11152667	881	266
17	1059 AZD8055	AZD-8055	MTORC1, MTORC2	PI3K/MTOR signaling	25262965	870	266
18	1032 Afatinib	BIBW2992, Tovok, Gilotrif	ERBB2, EGFR	EGFR signaling	10184653	881	266
19	1377 Afatinib	BIBW2992, Tovok, Gilotrif	ERBB2, EGFR	EGFR signaling	10184653	944	266
20	281 Alectinib	CH542802, CH 542802, Alecensa	ALK	RTK signaling	49806720	974	266
21	293 Amuvatinib	MP470, MP 470, MP-470	KIT, PDGFR, FLT3	Other, kinases	11282283	971	266
22	205 Avagacestat	BMS-708163, BMS 708163	Amyloid beta20, Amyloid beta40	Other	46883536	977	266
23	1072 Avagacestat	BMS-708163, BMS 708163	Amyloid beta20, Amyloid beta40	Other	46883536	957	266
24	1021 Axitinib	AG-13736, Inlyta	PDGFR, KIT, VEGFR	RTK signaling	6450551	877	266
25	178 BAY-61-3606	Syk Inhibitor, BAY-613606	SYK	Other, kinases	10200390	925	266
26	60 BI-2536		PLK1, PLK2, PLK3	Cell cycle	11364421	424	266
27	279 BIX02189	BIX 02189	MEK5, ERK5	ERK MAPK signaling	46931012	975	266

Figure 4_7: Drugs dataset

4.2 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).

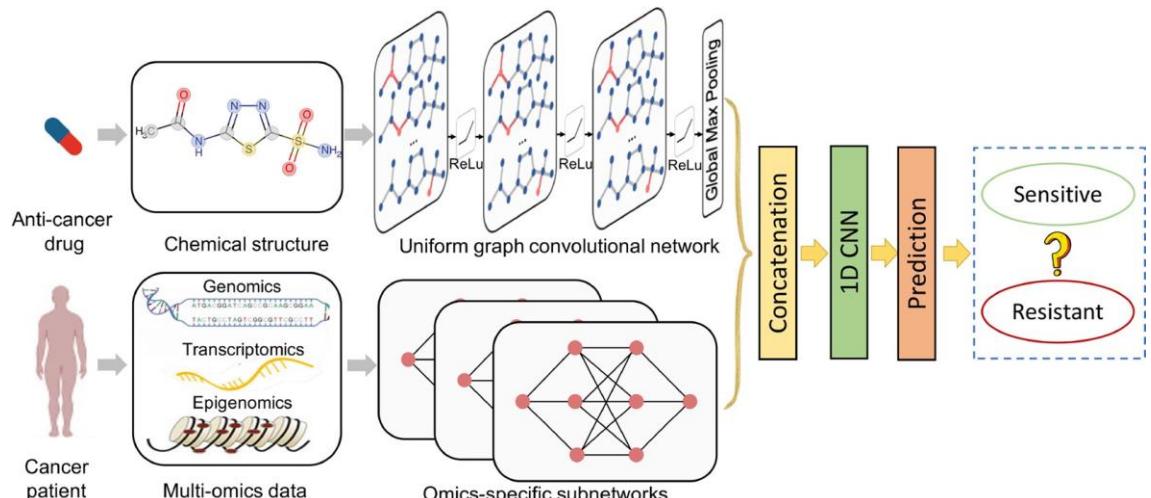


Figure 4_8: Model Framework

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).

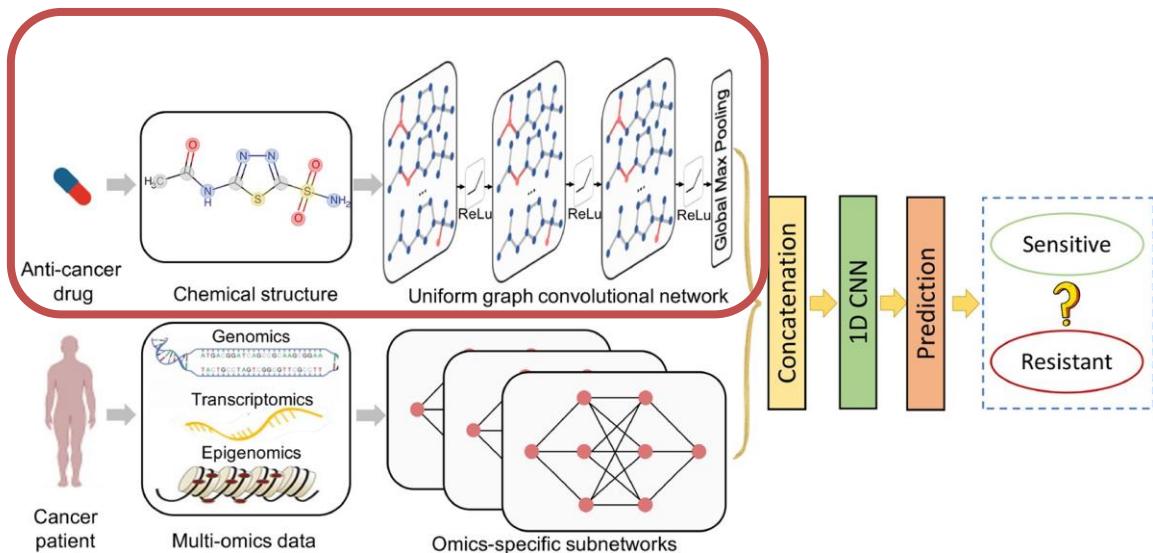


Figure 4_9: Model framework (first part)

The first part: Drug

Drug is represented as a graph based on the chemical structure.

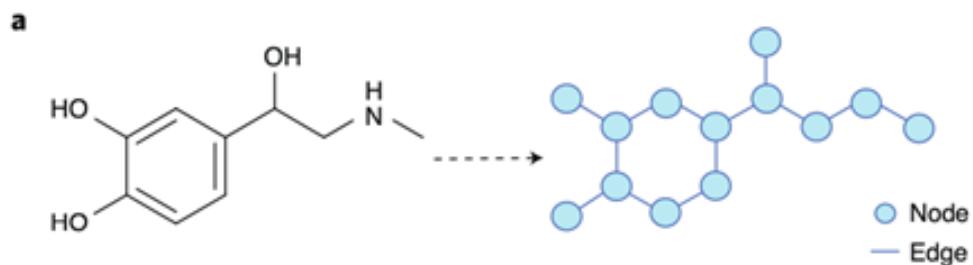


Figure 4_10: Drug Representation

4.3 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 $G = (V, E)$.

Learn the relationship between the molecular features of drugs such as (

Chemical features: These are molecular descriptors that capture the chemical properties of a drug, such as its size, shape, and functional groups. Chemical features can be represented as a matrix where each row corresponds to a drug and each column corresponds to a specific chemical feature.

Structural features: These are features that capture the structural properties of a drug, such as its substructures or its 3D structure. Structural features can be represented as a matrix where each row corresponds to a drug and each column corresponds to a specific structural feature.

Pharmacological features: These are features that capture the pharmacological properties of a drug, such as its target proteins or its side effects. Pharmacological features can be represented as a matrix where each row corresponds to a drug and each column corresponds to a specific pharmacological feature), and their neighboring molecules, The adjacency matrix of the graph is used to represents the connections between the nodes in a graph. The rows and columns of the matrix represent the nodes in the graph, and the elements of the matrix indicate whether there is an edge between two nodes. If there is an edge between two nodes, the corresponding element in the matrix is set to 1, otherwise it is set to 0.

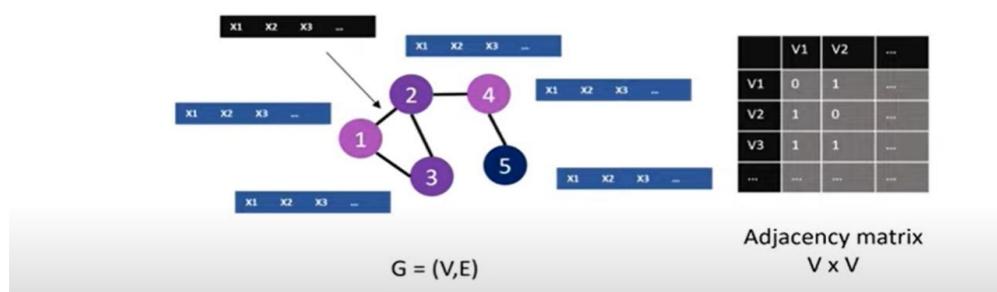


Figure 4_11: GCN

GCN Layers:

Input layer: This layer receives the input graph data, which typically includes the node features and the adjacency matrix of the graph.

Graph Convolutional Layer: This layer performs the graph convolution operation on the molecular features of drugs, using the adjacency matrix of the drug

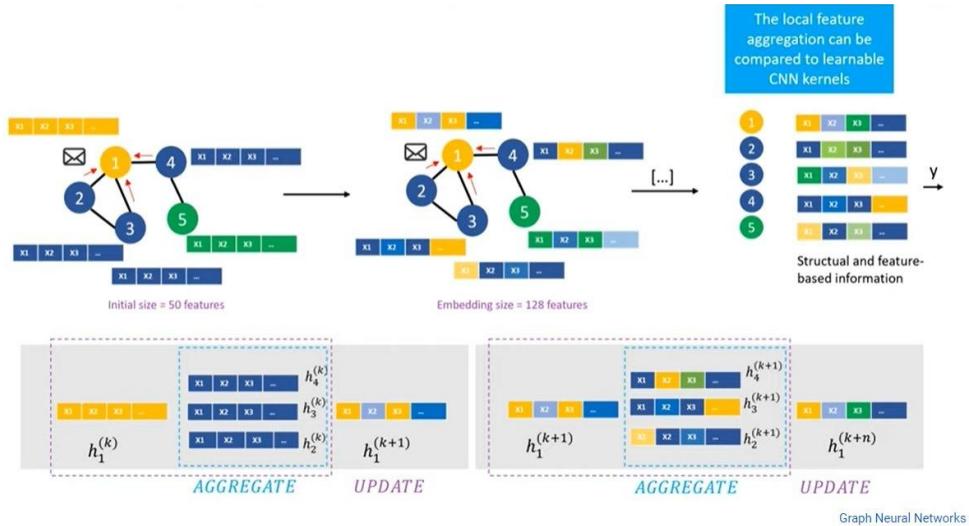


Figure 4_12: Convolution

For each node, we get the feature information from all its neighbours. Each node takes messages from other nodes in more than round which leads to all nodes know each other.

Graph convolution operation:

- 1) Normalize the adjacency matrix of the graph by adding the identity matrix to ensure that the graph is fully connected, and then calculating the degree matrix of the normalized adjacency matrix.
- 2) Multiply the input feature matrix (i.e., molecular features of drugs) by the weight matrix to obtain a new feature matrix.
- 3) Compute the product of the normalized adjacency matrix and the new feature matrix to obtain the output feature matrix.

Activation layer: This layer applies an activation function to the output of the GraphConv layer

RELU: $F(X) = \text{MAX}(0, X)$ returns the input value if it is positive, and returns 0 if the input value is negative.

Batch Normalization layer: This layer normalizes the output of the previous layer to improve the stability and performance of the model.

$$y = \text{gamma} * (x - \text{mean}) / \sqrt{\text{variance} + \text{epsilon}} + \text{beta}$$

where x is the input to the batch normalization layer, gamma and beta are learned scaling and shifting parameters, mean and variance are the mean and variance of the input batch, and epsilon is a small value added for numerical stability.

Dropout layer: This layer randomly drops out a portion of the output of the previous layer during training to prevent overfitting.

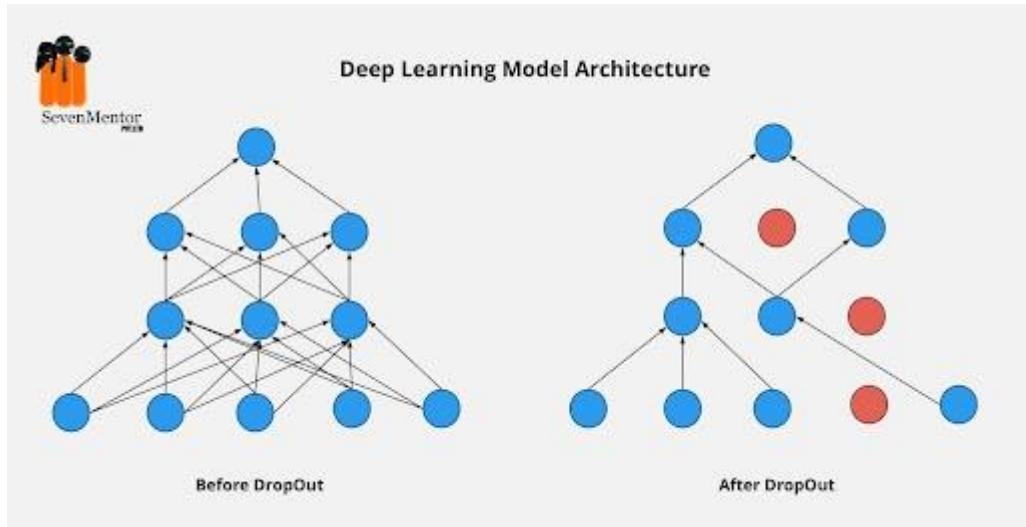


Figure 4_13: Dropout

Randomly selects a subset of neurons in the network and sets their output to zero during training, effectively "dropping out" these neurons from the network for that particular iteration. This helps to prevent the network from relying too heavily on any one particular set of neurons, and encourages the network to learn more robust and generalizable features.

Pooling Layer: responsible for reduce the dimensionality of the feature matrix and improve the computational efficiency of the model by reducing the spatial size of the Convolved Feature.

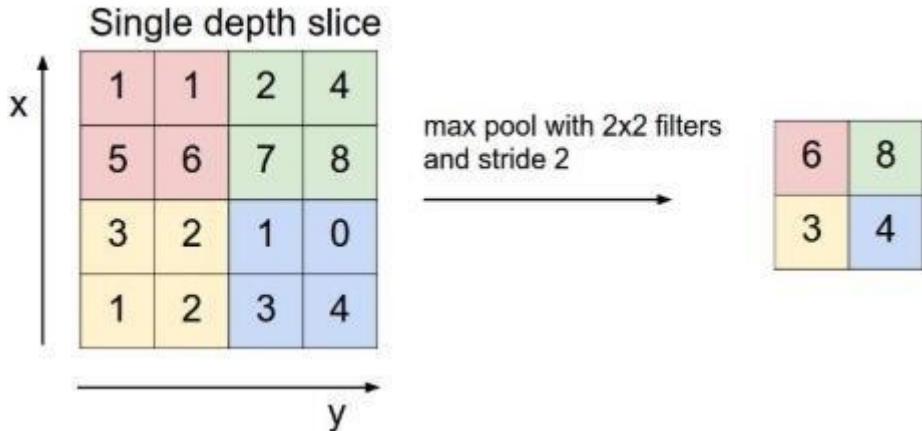


Figure 4_14: Pooling Layer.

It involves dividing the input image into small non-overlapping rectangular regions, and taking the maximum value within each region to obtain a new output image with reduced spatial resolution.

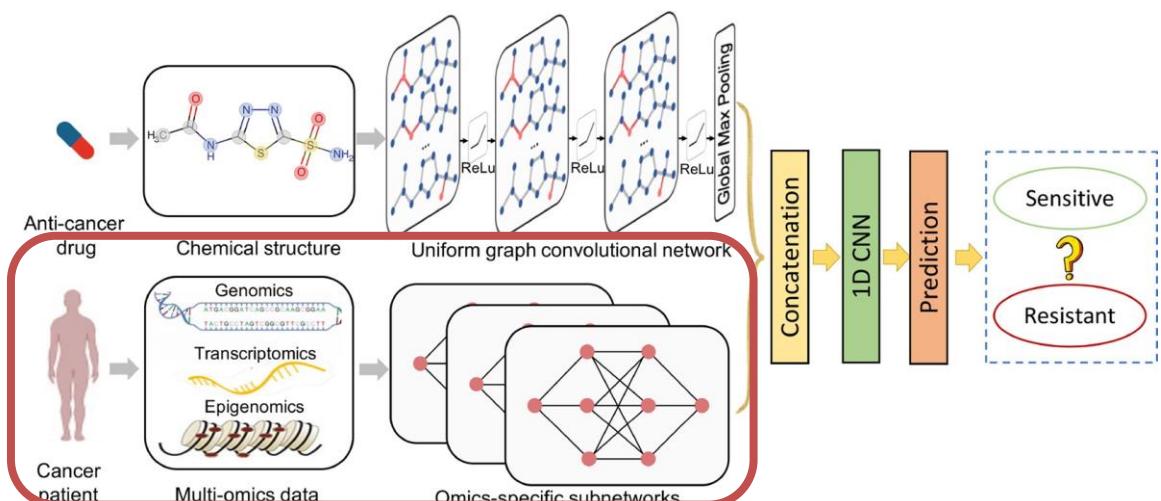


Figure 4_15: Model framework (second part)

The second part:

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

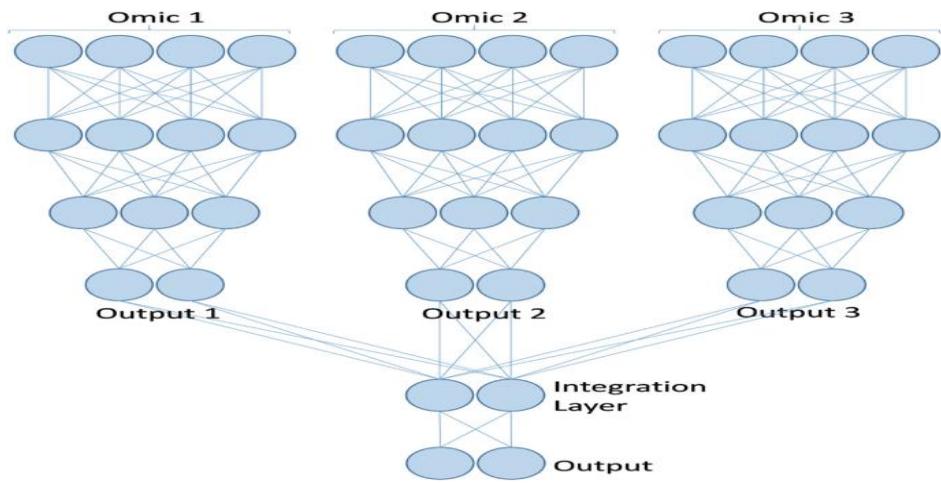


Figure 4_16: Omics Representation

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

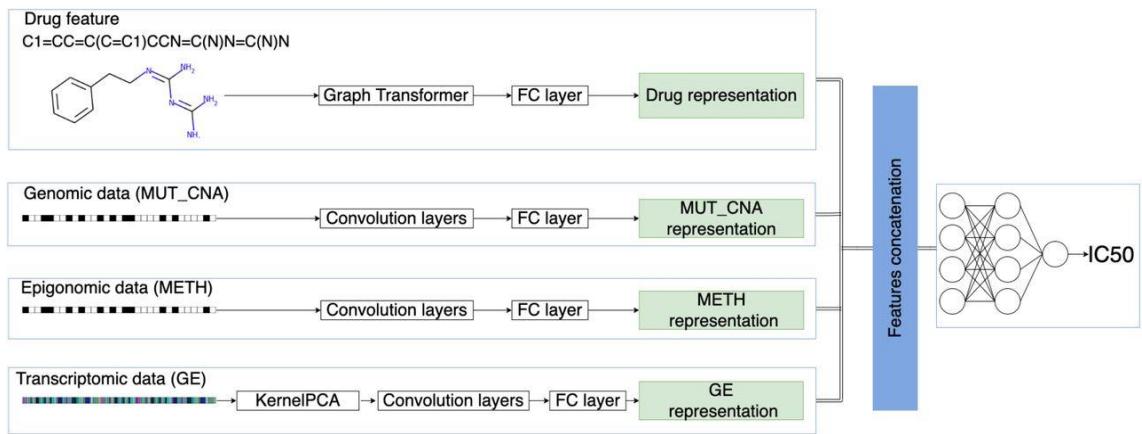


Figure 4_17: Data Representation

3.4 CNN

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

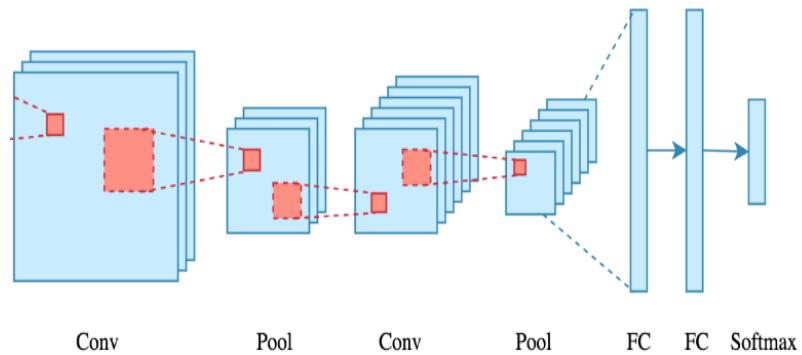


Figure 4_18: CNN Layers

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.

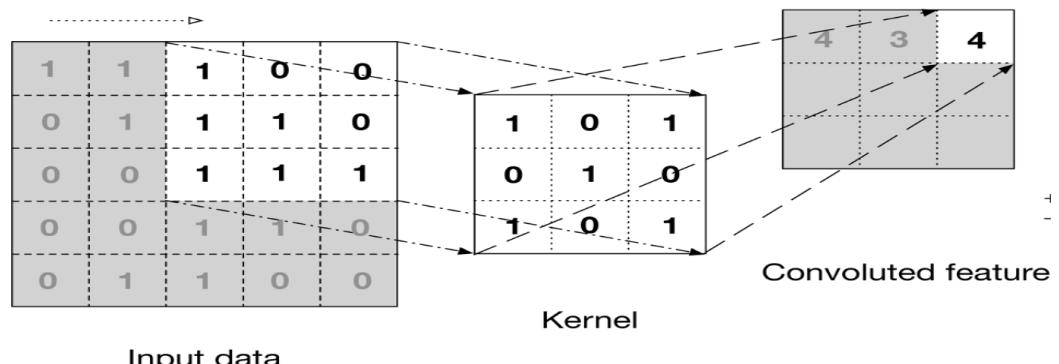


Figure 4_19: Convolution layer

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.

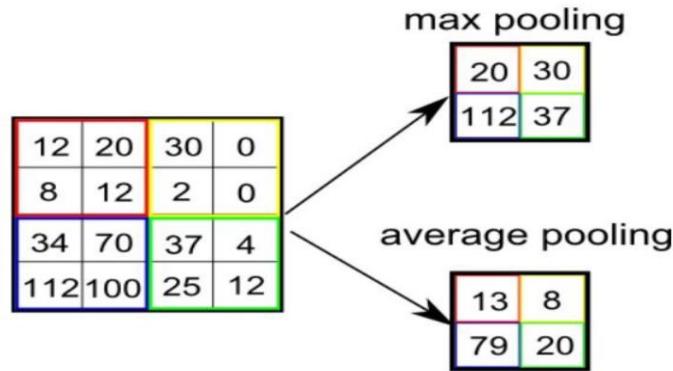


Figure 4_20: Pooling layer

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.

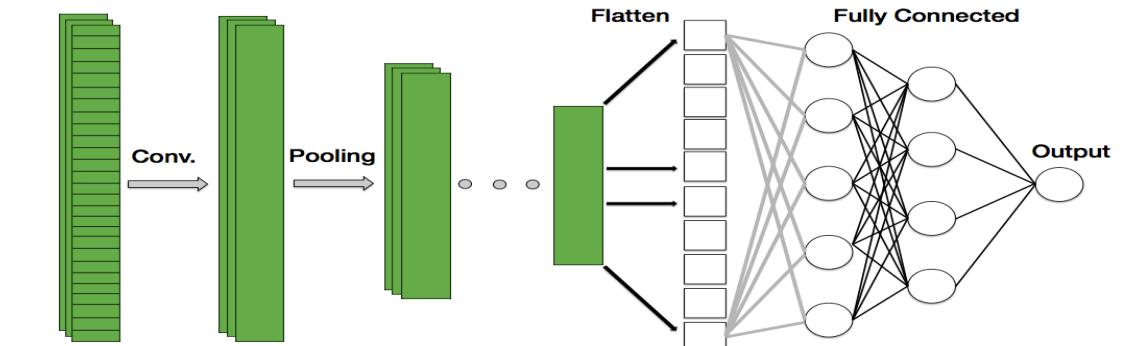


Figure 4_21: FCI layer

Finally the output is either a regression value (the drug sensitivity (IC50) or a binary classification probability (sensitive or resistant).

Chapter Five

5 RESULTS

Previous Results:

Previous accuracy for classification model was 0.84 and Pearson's correlation of regression model was 0.9200.

Our Work:

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: 0.8440, precision of regression: 0.9200. Then we used k-fold and found an increase in Accuracy classification: 1.0000, precision of regression: 0.9855.

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.

With regression we evaluate model with Pearson's correlation which is calculated as the covariance between the two variables divided by the product of their standard deviations. Specifically, the formula is:

$$r = \text{cov}(X, Y) / (\text{std}(X) * \text{std}(Y))$$

We applied K_Fold which gives us Pearson's correlation 0.9200, then we apply Linear Regression model with K_Fold which gives us Pearson's correlation 0.9965, and Decision Tree model with K_Fold which give us Pearson's correlation 0.9967.

Results of regression model :

IDE	Run Time	Model	Person Correlation
Colab pro	25m	GCN(Train_Test_Split)	0.9200
Colab pro	regression: 25m,	GCN(K_Fold)	0.9996
Colab	8.26 m	Linear_regression(K_Fold)	0.9965
Colab	7.25 m	Decision_Tree(K_Fold)	0.9967

With Classification: We evaluate accuracy using `roc_auc_score` function in Python's `sklearn.metrics` module which is a function that calculates the AUC score for a given binary classification model. The function takes two input arrays:

`y_true`: an array of true binary labels (0 or 1) for the test set

`y_score`: an array of predicted probabilities of the positive class for the test set

The `roc_auc_score` function first uses the `roc_curve` function to compute the TPR and FPR values for different threshold values based on the `y_true` and `y_score` inputs. It then computes the AUC score as the area under the ROC curve using the trapezoidal rule.

$$\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$$

$$\text{FPR} = \text{FP}/(\text{TN} + \text{FP})$$

The AUC score is a value between 0 and 1, where 1 indicates a perfect classifier and 0.5 indicates a random classifier.

GCN with Train_Test_Split gives us accuracy 0.8440, we apply K_Fold With GCN which give us accuracy 1.0000, we also applied Random Forest model which gives us accuracy 0.8762 with train_test_split and 0.9996 with K_Fold, and we applied Logistic Regression model which gives us accuracy 0.8361 with Train_Test_Split and 1.0000 with K_Fold.

Results of Classification model.

IDE	Run Time	Model	Accuracy
Colab Pro	20m	GCN(Train_Test_Split)	0.8440
Colab Pro	20m	GCN(K_Fold)	1.0000
Colab	12m	Random_Forest with GCN (Train_Test_Split)	0.8762
Colab	14m	Random_Forest with GCN (Train_Test_Split)	0.9996
Colab	9m	Logistic with GCN(Train _Test Split)	0.8361
Colab	8.30 m	Logistic with GCN(K_Fold)	1.0000

At the end, it became clear to us from the previous that when we used. -k-fold We got the best result (Accuracy classification: 1.0000, Pearson's correlation: 0.9996)

Chapter Six

6 SYSTEM IMPLEMENTATION

6.1 SYSTEM OVERVIEW:

We design application to help solving some problems of patients, doctors and management. In our system we facilitate the process for patient, admin and doctor.

- **For patient**

He can make a registration or login if he had an email before, the system then open and the first page that patient will see is home page which contain our system services (medical rays, medical tests, medical clinics and booking an appointment)

- **For admin**

Each admin will have a code and password to login with.

The system then open and the first page will appear is dashboard which help him navigate to pages and contain bar chart of doctor specializations and pie chart of requests. He will have a profile page contain his information, request page that contain requests made by patient and admin can accept them or canceled, doctor page which contain names of our doctors and when click on any of them will move to the page of doctor details and admin can update this information and in doctor page admin can add new doctor. In each page there is a navigation bar that will help him moving over pages.

- **For doctor**

Each doctor will have an email and password provided by the admin.

The system then open and the first page will appear is dashboard which help him navigate to pages. The doctor will have a profile page containing his information, appointments page that contain his appointments that the admin organized, model page which uses to help him predict the suitable treatment containing upload files and predict, when choose upload files will open other page containing name of patient and labels to upload files such as gene expression file, genomic mutation file, methylation file (data of patient).when choosing predict will open the page

of drug effectiveness containing patient name and drug and then click on predict button will show the result.

Patient, admin and doctor can log out from the system.

6.2 INTRODUCTION PAGES:

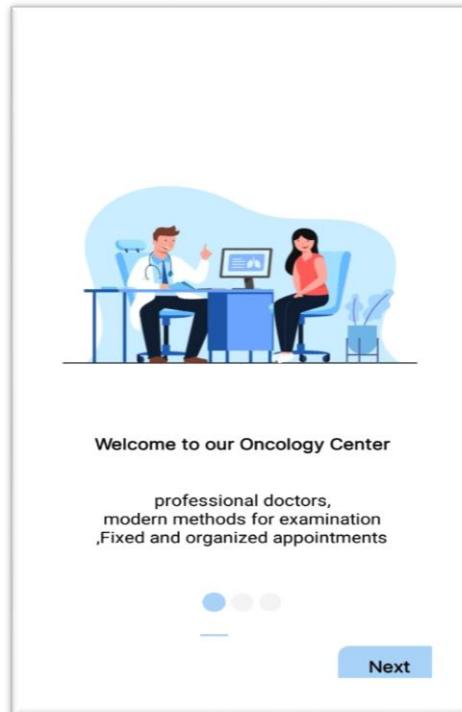
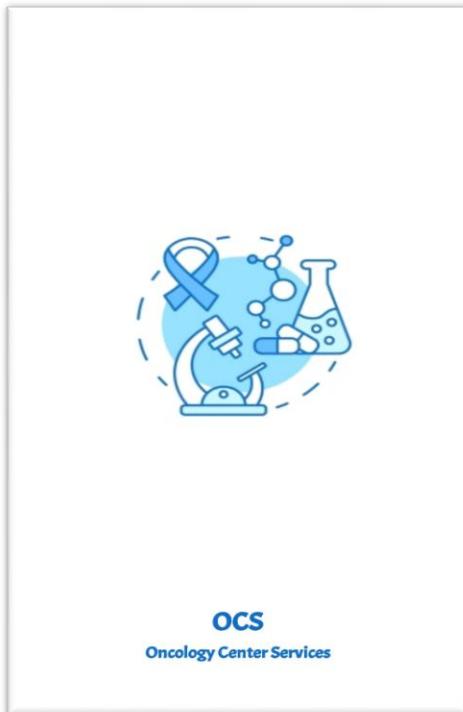


Figure 6_1: Logo

Figure 6_2: Welcome Page

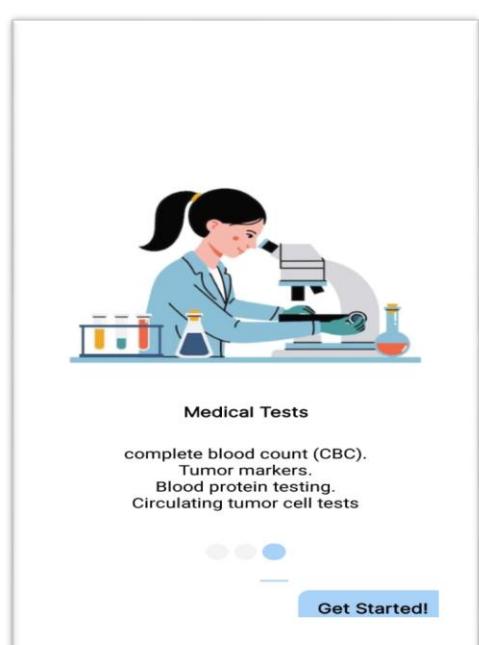
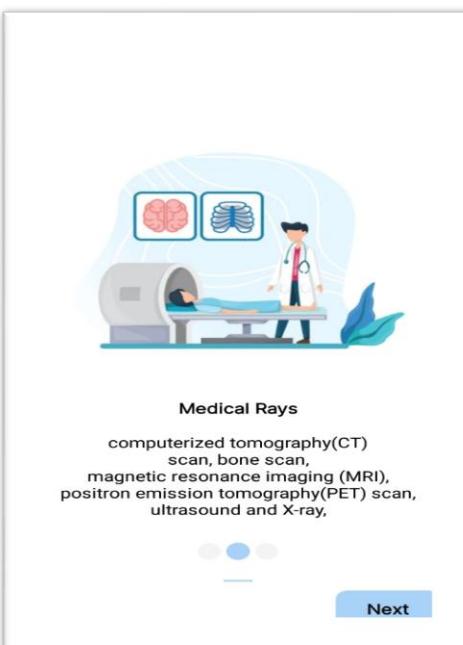


Figure 6_3: Medical Rays

Figure 6_4: Medical Tests

These pages appear when user open application, it welcomes the user and explains some of the services in the system.

6.3 ADMIN PAGES

6.3.1 Login Page

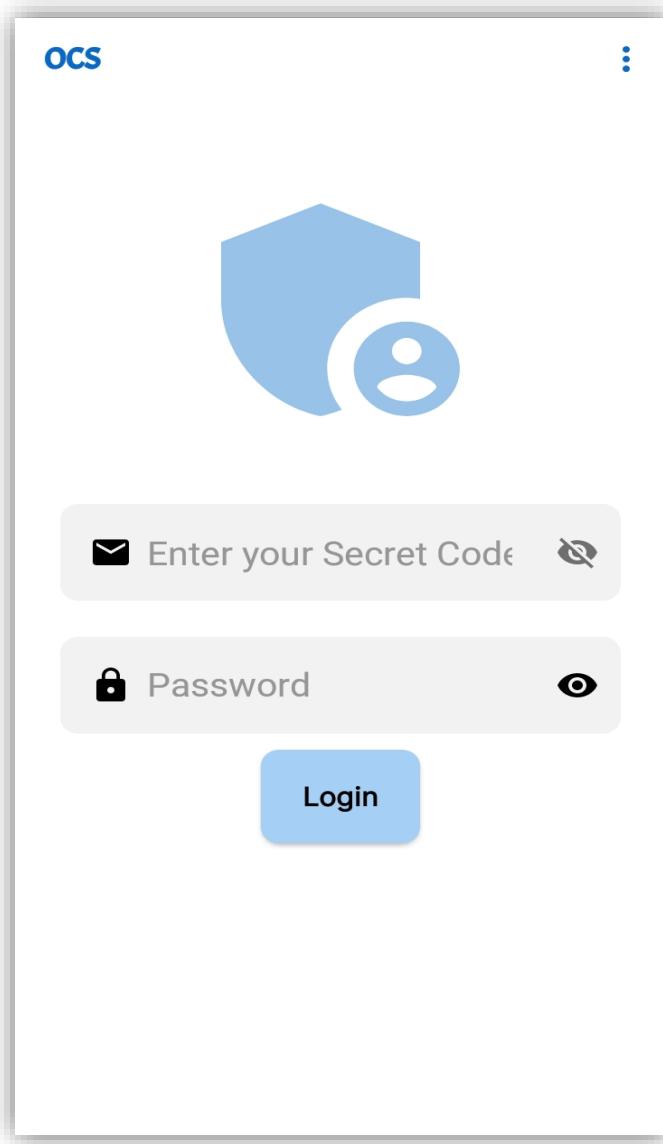


Figure 6_5: Admin Login.

When admin open system, this page appears to him and he should enter his code and password to enter to system and perform his functions.

6.3.2 Dashboard

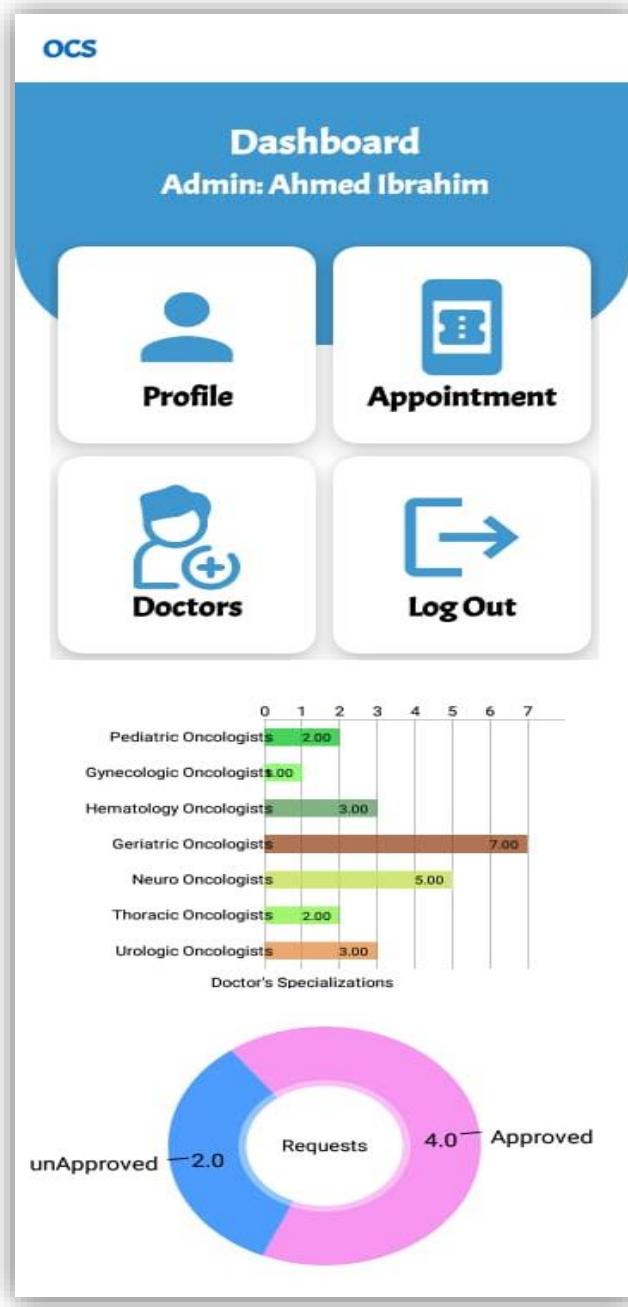


Figure 6_5: Dashboard Page.

After admin login, this page appears to him and from here he can manage doctors and appointments, there are also chart illustrates approved and unapproved requests, and chart illustrates percentage of doctors at each department.

6.3.3 Navigation view

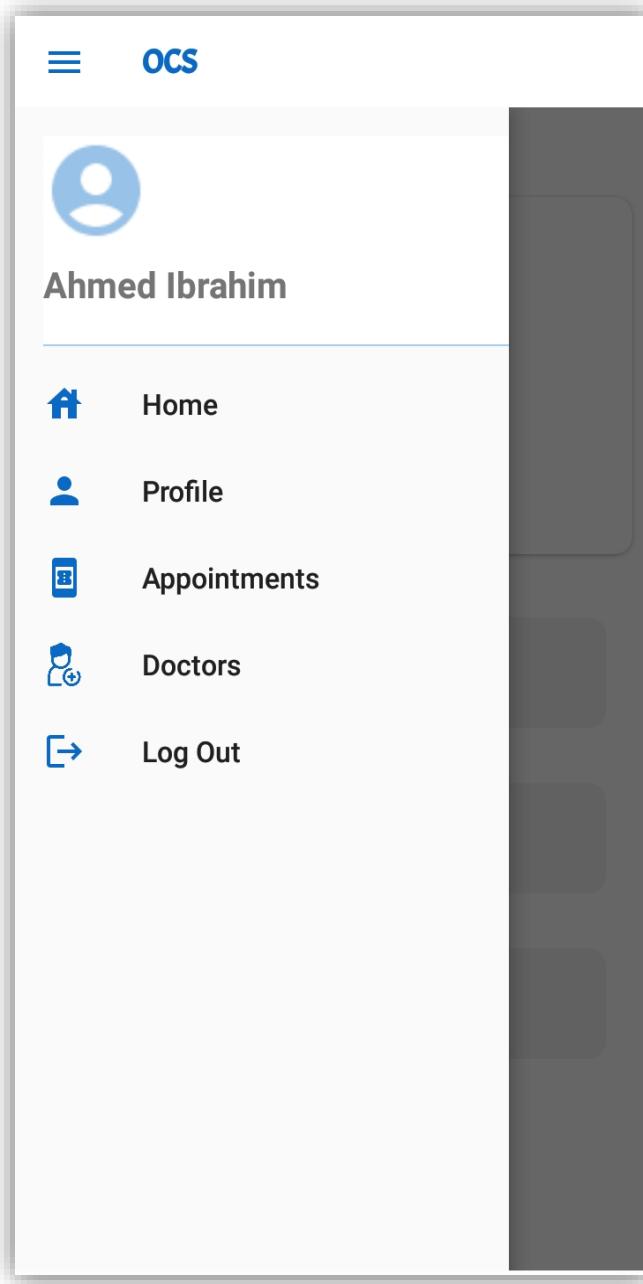


Figure 6_6: Navigation View Page.

This page of navigation view, from this page admin can view his profile, appointments, doctors also he can log out.

6.3.4 Profile

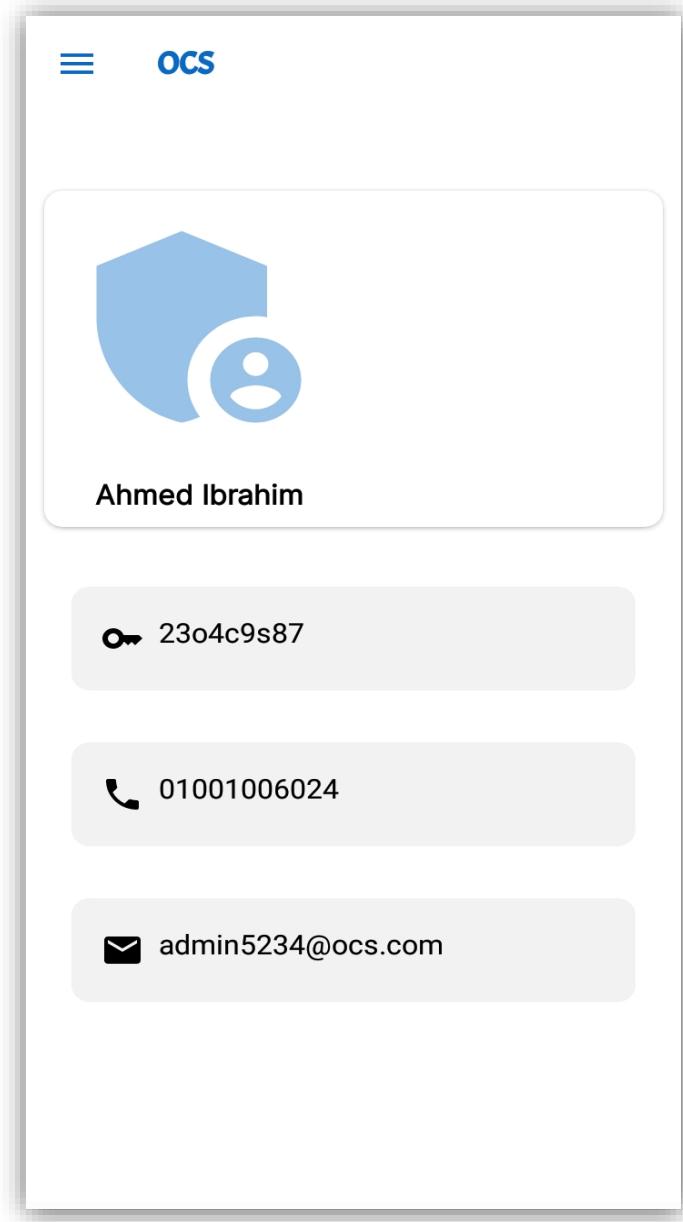


Figure 6_7: Profile Page.

This page is profile of admin, appears when admin clicks on profile on navigation view contains his name, code, phone number and email.

6.3.5 Appointment requests

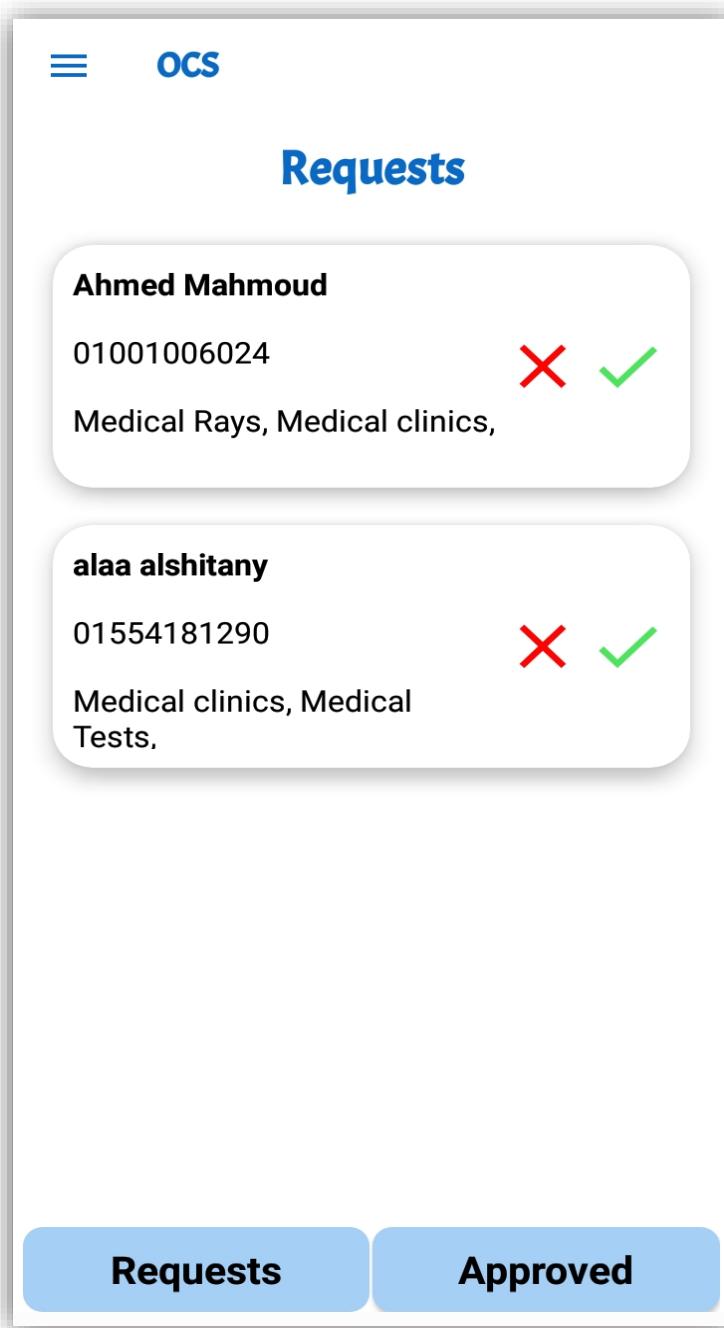


Figure 6_8: Requests Page.

This page appears when admin clicks on appointments on navigation, contains requests of system, admin can approve or unapproved them, when he clicks on false sign request is unapproved and when he clicks on right sign the new page will appears, from it admin choose time and date of each request, we will illustrate this page .

6.3.6 Appointment Confirmation

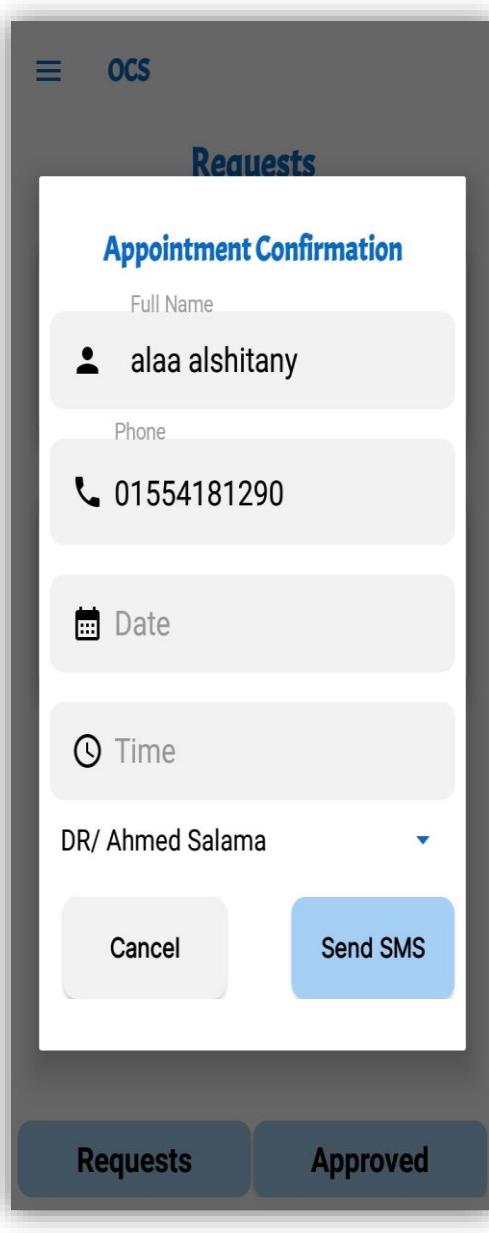


Figure 6_8: Confirmation Page.

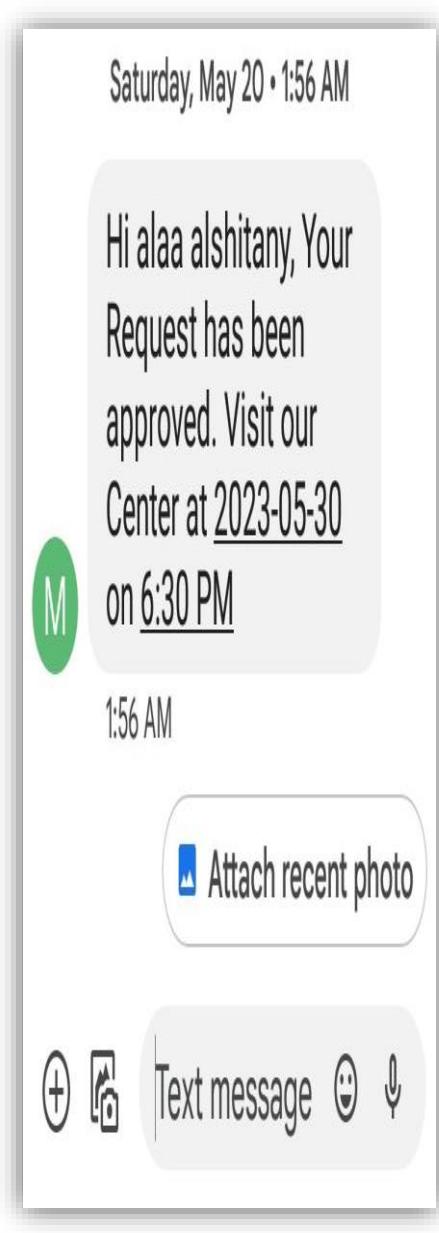


Figure6_9: Confirmation message.

After admin approves request he should define time and date for each patient then sends message to his phone to confirm the booking. Figure 5_9 illustrate the confirmation message sent to patient.

6.3.7 Approved requests

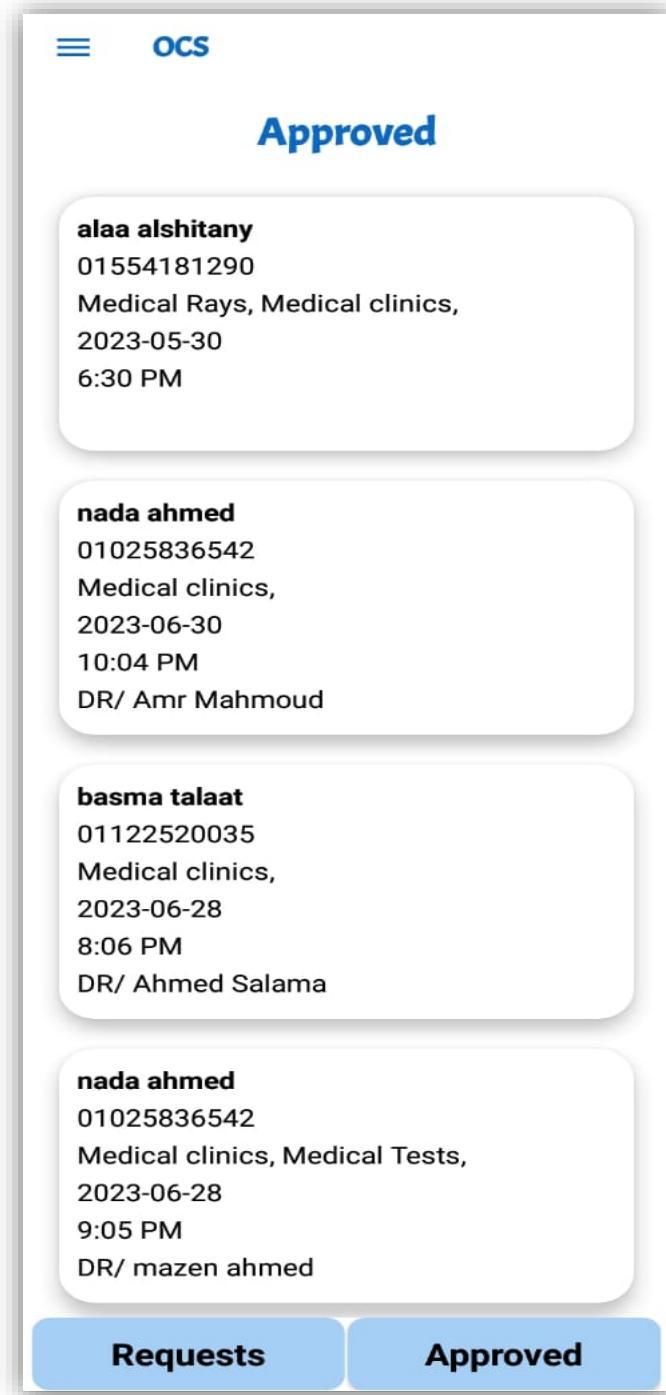


Figure 6_10: Approved Requests Page.

This page contain approved requests which contain patient name, phone number, service type, date, time and doctor name.

6.3.8 Doctors

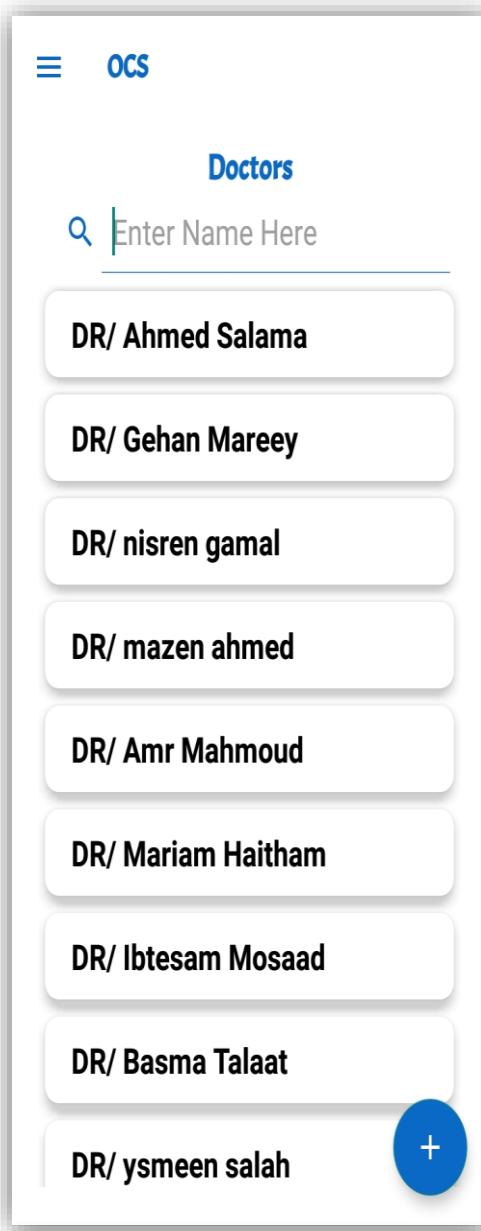


Figure 6_11: Doctors page.

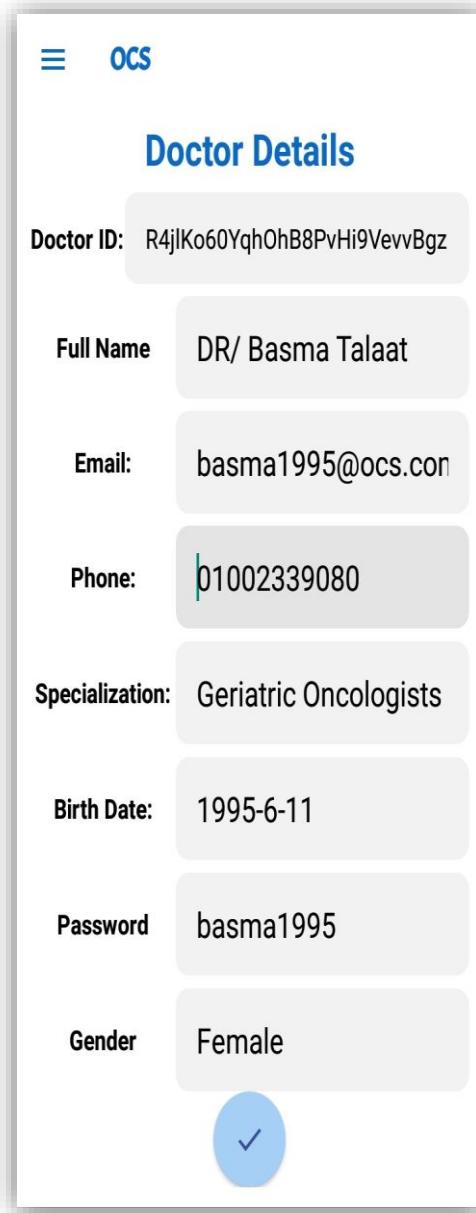


Figure 6_12: Doctors details page.

Figure 5_11 contains doctors name, when click on doctor name, figure 5_12 will appears which contains (doctor ID, full name, email, phone, specialization, birth date, password and gender), admin can adds new doctor by clicking on plus sign.

6.3.9 Add Doctors

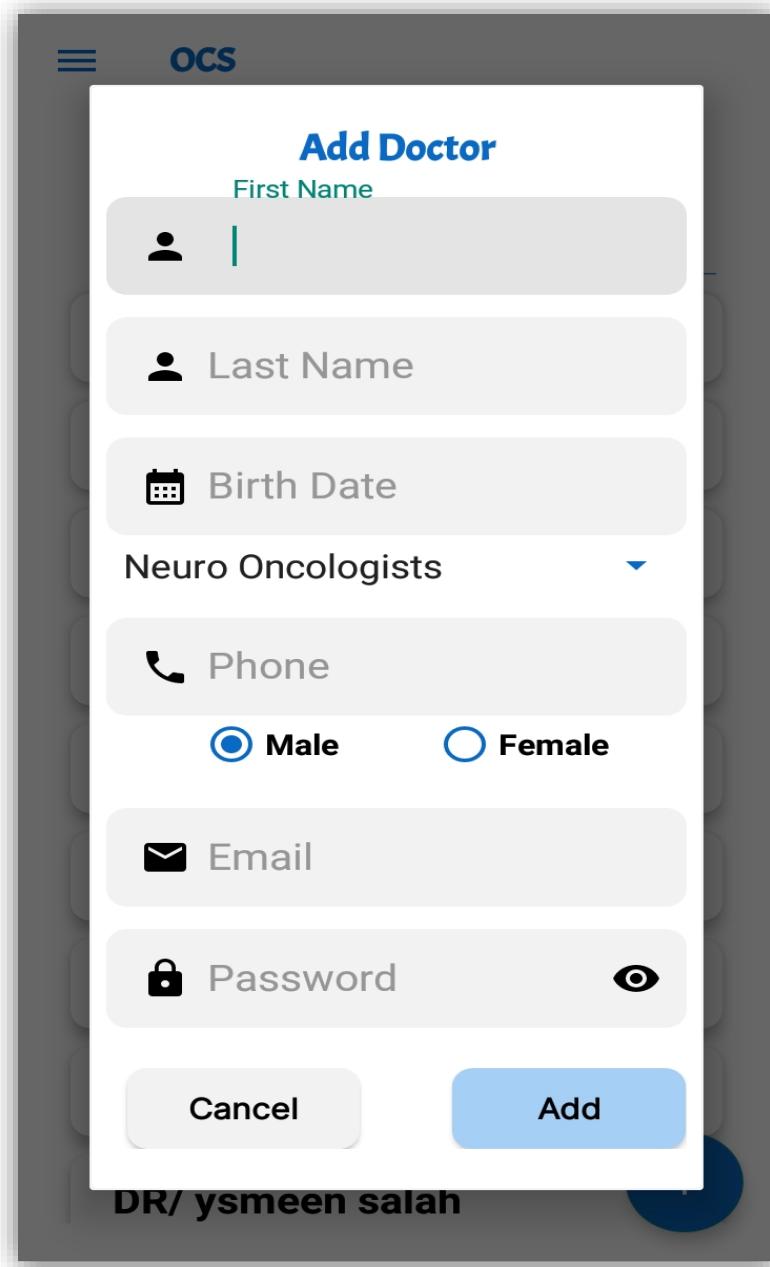


Figure 6_13: Add Doctors page.

This page appears when admin clicks on plus sign, from this page admin can add new doctors by adding information of doctor (name, birth date, specialization, phone, gender, email and password), by clicking on Add new doctor is added successfully.

6.4 PATIENT PAGES

6.4.1 Register Pages:



Register

First Name

Last Name

Email

Phone

Male Female

Continue →



Continue Register

Birth Date

Address

Password

Confirm Password

Register

Already have an account? [Login](#)



Figure 6_14: Register Page 1.

Figure 6_15: Register Page2.

When patient use an application for the first time he should sign up (register) by entering his information such as (First Name, Last Name, Email, Phone, Birth Date, Address, Password, Confirm Password), when patient click on register his data will be saved in storage in firebase and he can login easily.

6.4.2 Login Page

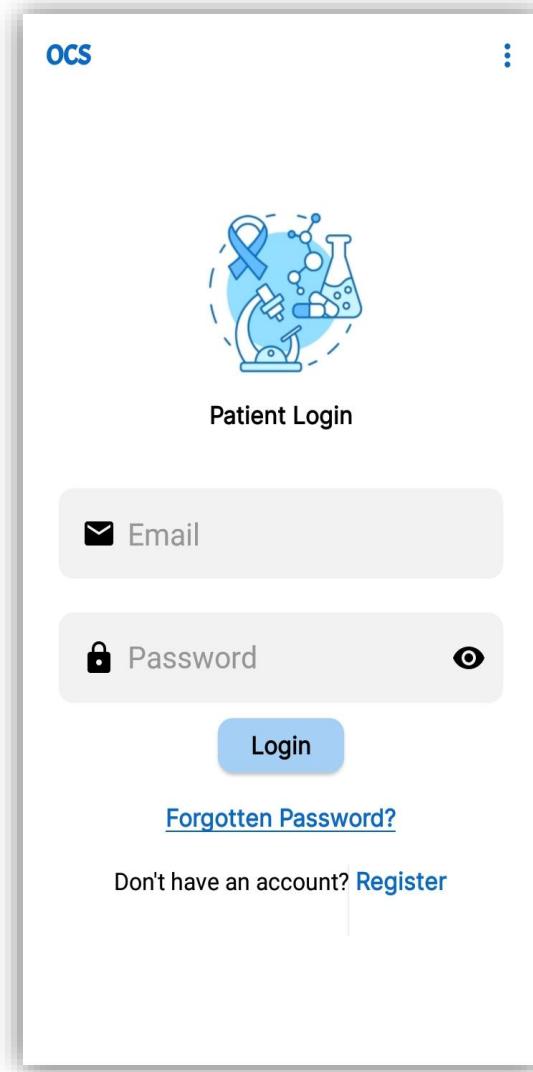


Figure 6_16: Login Page

After patient register he can login to application by entering his Email and Password, after logged in he can view all services of system and use its functions easily.

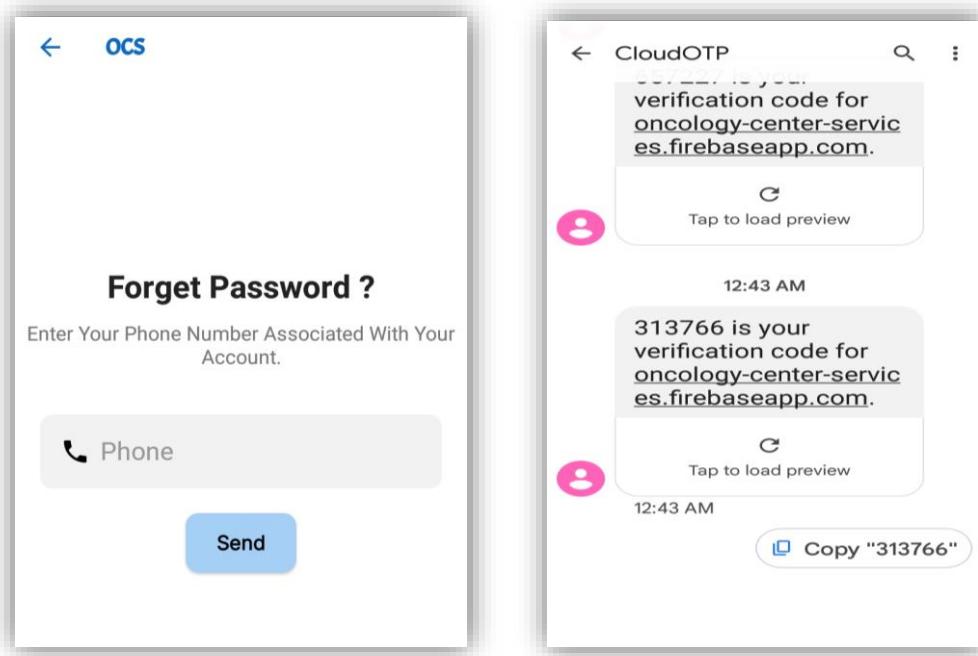


Figure 6_17: Forget Password Page

Figure 6_18: Code Page

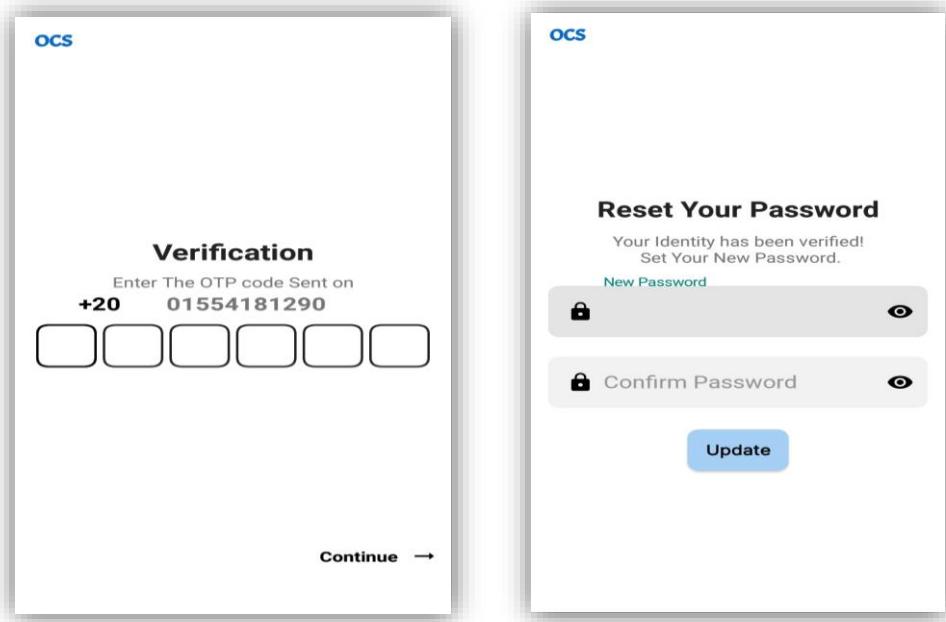


Figure 6_19: Verification Page.

Figure 6_20: Reset Password Page.

When patient forgets his password he can click on forget password, Figure 5_8 will appear and he should enter his number then system will send code to him as shown in figure 5_9 then he should enter code in figure 5_10 then figure 5_11 will appear and he should enter new password.

6.4.3 Navigation

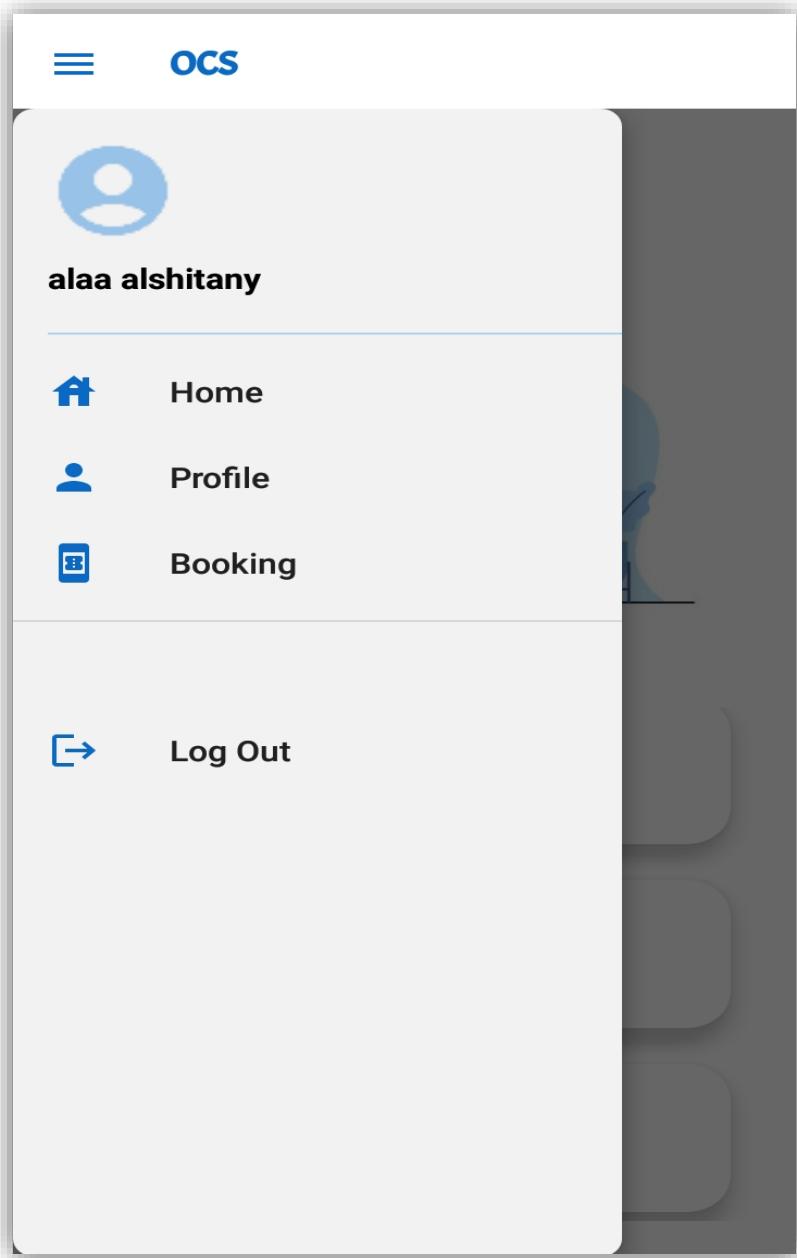


Figure 6_21: Navigation Bar.

This page will appear when patient clicks on navigation bar, it contains home, Profile, and booking. Patient also can log out by clicking on logout, we will explain each of them in the following pages.

6.4.4 System Services (Home)

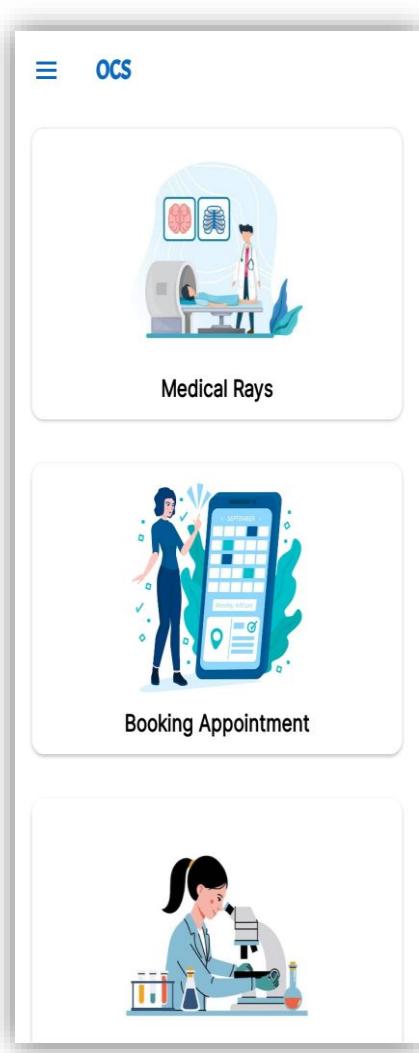


Figure 6_22: Services Page1.

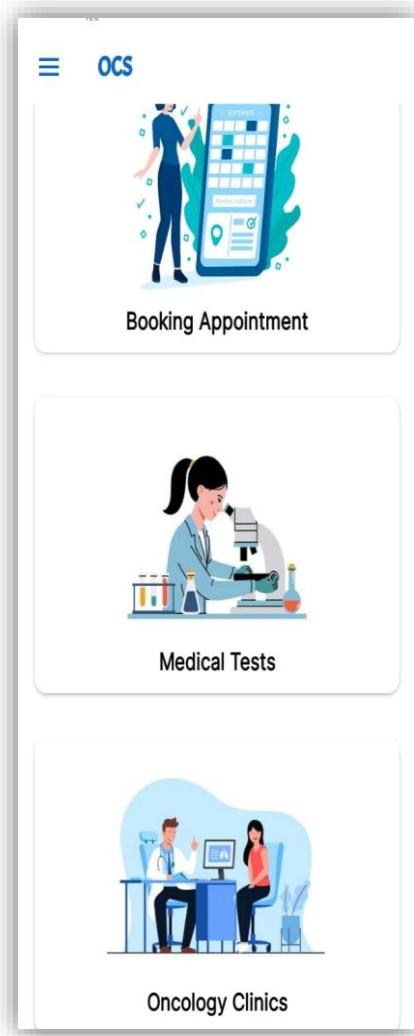


Figure 6_23: Services Page2.

After patient login to system, system services will appear as shown in figure 5_12. Our system provides many services for patients such as (Medical Rays, Booking Appointment, Medical Tests, Oncology Clinics).

When patient click on each of them new page will appear, we will explain in the following figures.

6.4.4.1 Medical Rays

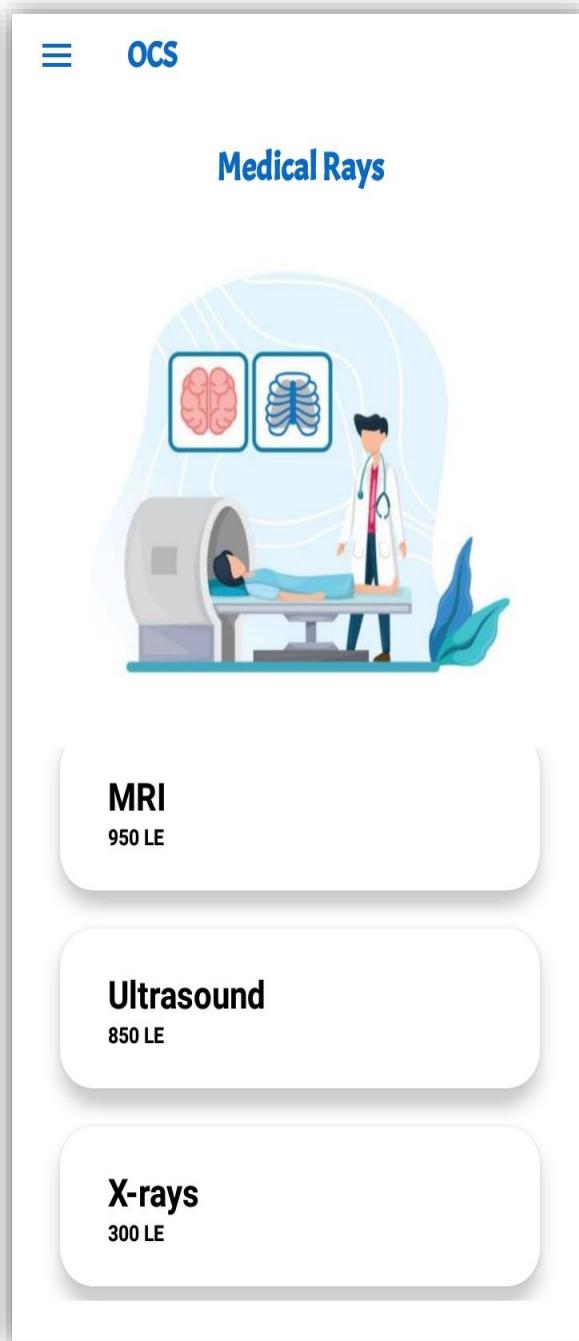


Figure 6_24: Medical Rays.

This page will appear when patient click on Medical Rays on Services page, our system provides MRI, Ultrasound and x_rays.

6.4.4.2 Booking Appointment:

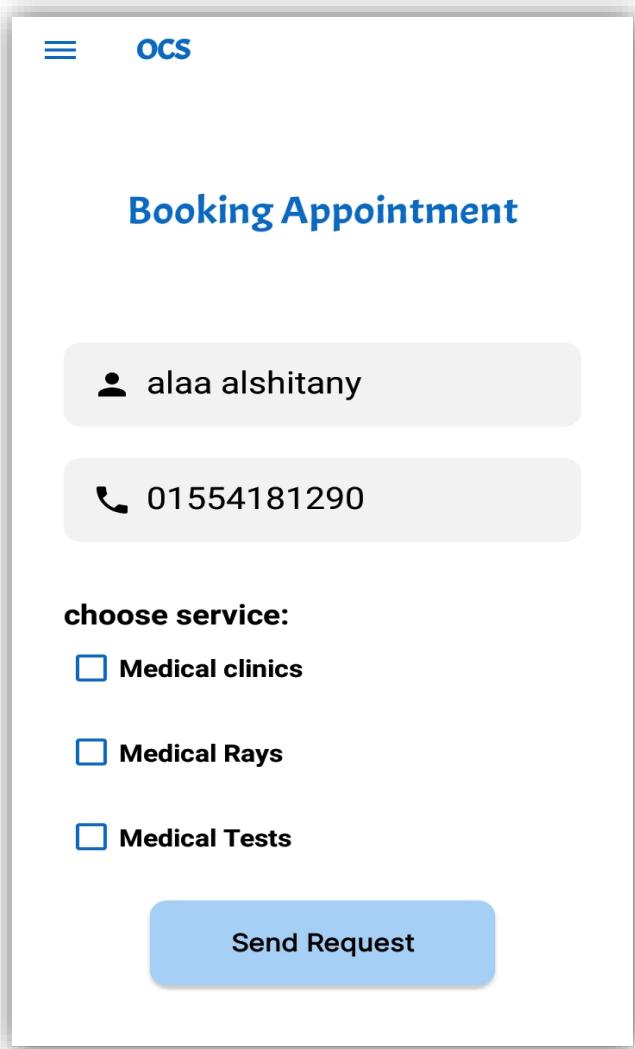


Figure 6_25: Booking appointment.

When patient clicks on booking appointment in services page this page will appear, patient can book appointment by entering his name and phone then choose service he needs(Medical Clinics, Medical Rays, Medical Tests).

6.4.4.3 Medical tests

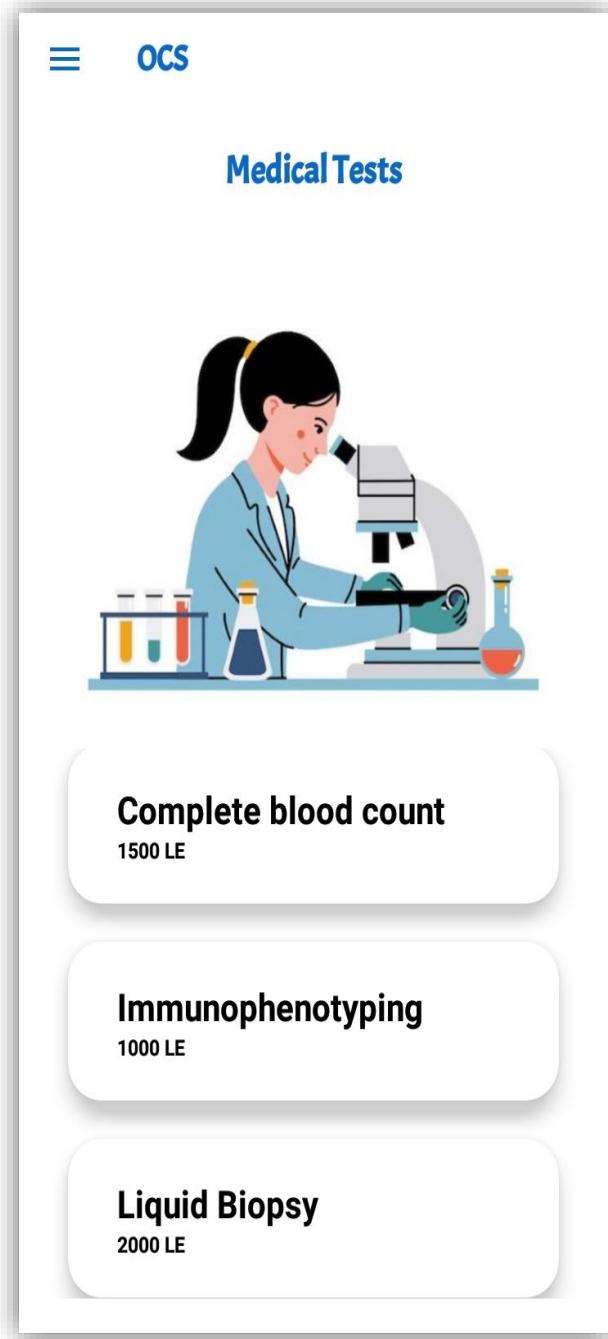


Figure 6_26: Medical Tests.

This page will appear when patient click on Medical Tests on Services page, patient views our system tests (Complete blood count, Immunophenotyping, Liquid Biopsy) and their price.

6.4.4.4 Medical clinics

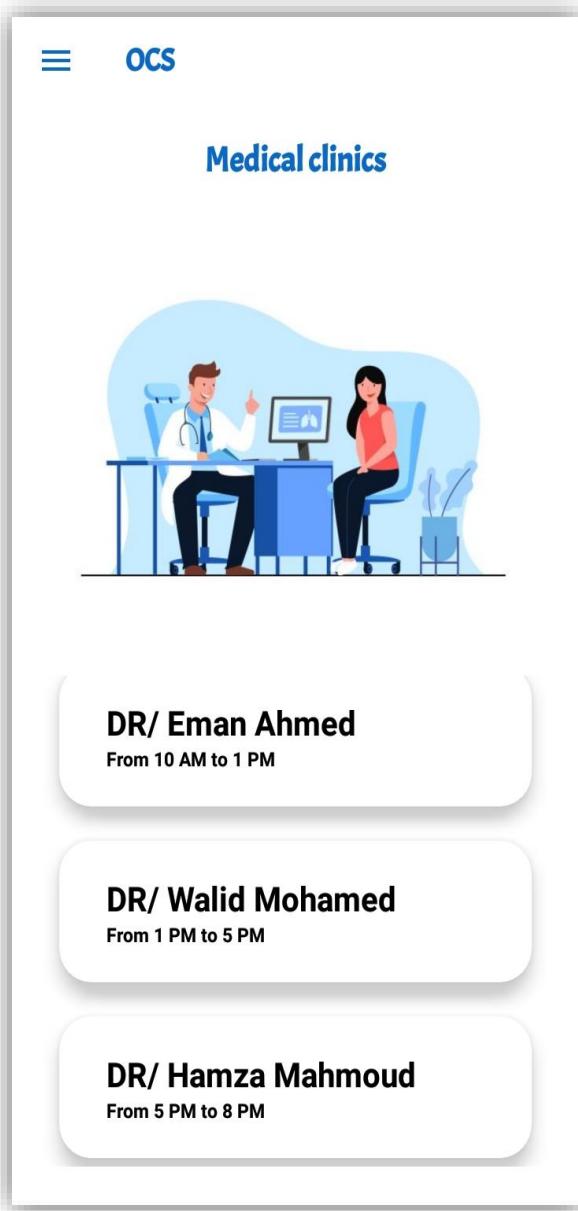


Figure 6_27: Medical Clinics.

This page will appear when patient click on Medical Clinics on Services page, this page contains each doctor's information and appointments.

6.5 DOCYOR PAGES

6.5.1 Doctor Login and Dashboard

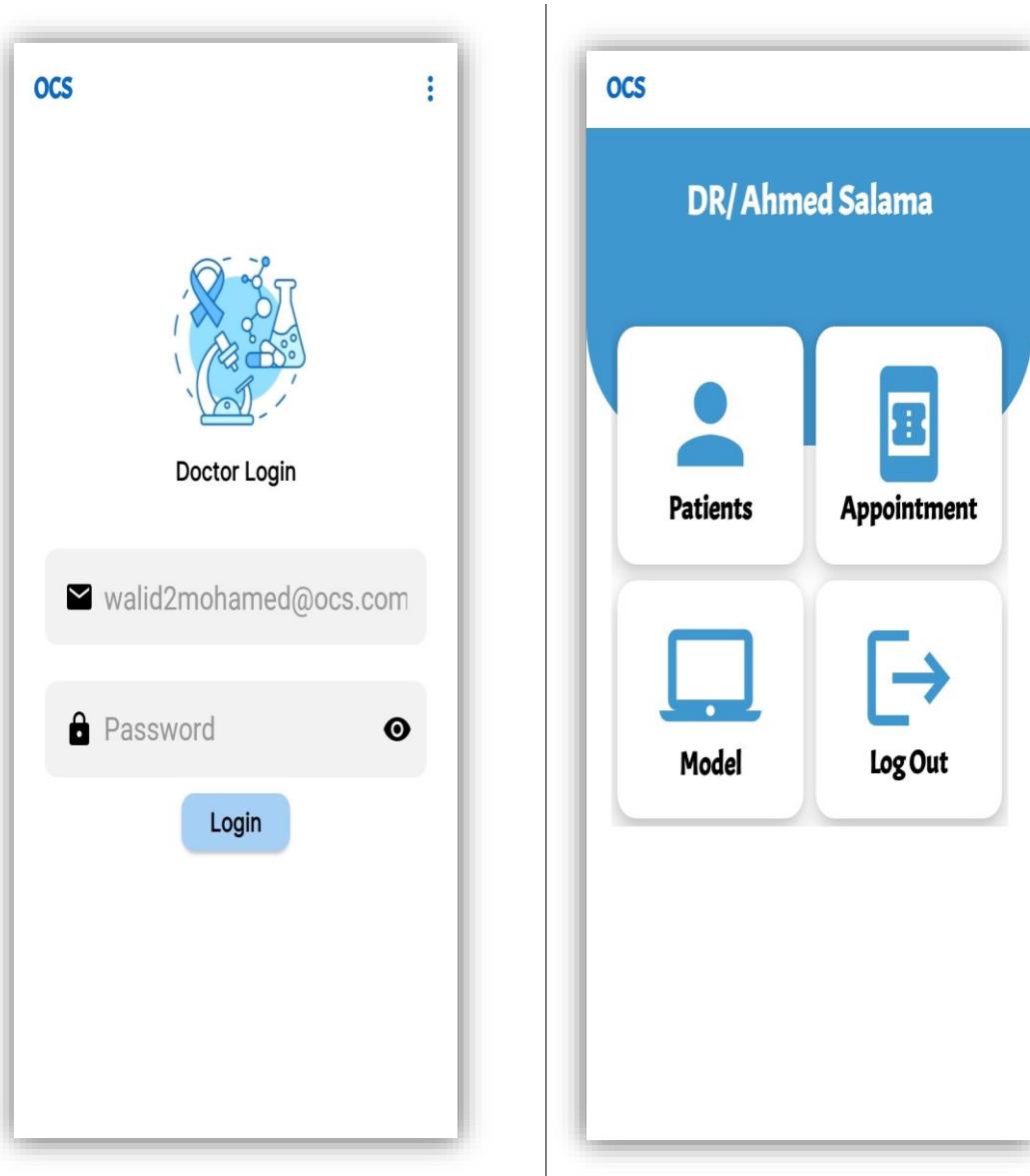


Figure 6_28: Doctor Login.

Figure 6_29: Doctor Dashboard.

When doctor opens the system Figure 5_28 will appear to him and he should enter his email and password which specified by admin, after he logged in dashboard will appear to him as shown in Figure 5_29 from here he can manage patients as he can add or delete patients, he can view appointments, also he can use model to help him in prediction, we will explain each part.

6.5.2 Manage patients

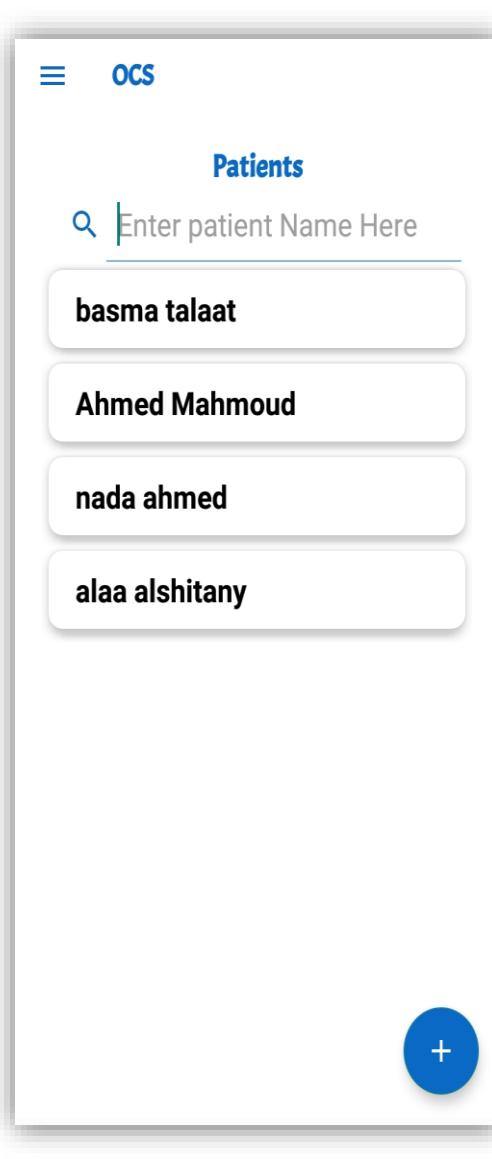


Figure 6_30: Search Patient Page.

A screenshot of a modal dialog titled "Add Patient". It contains fields for "First Name" (with a placeholder " "), "Last Name" (placeholder " "), "Birth Date" (placeholder " "), "Address" (placeholder " "), "Phone" (placeholder " "), and gender selection ("Male" and "Female" radio buttons, with "Male" selected). There are also fields for "Email" (placeholder " ") and "Password" (placeholder " "). At the bottom are "Cancel" and "Add" buttons.

Figure 6_31: Add Patient Page.

Doctor can view Patients' names and search for a specific patient as shown in Figure 5_30, if doctor needs to add patient he can click on plus sign, figure 5_31 will appear, doctor will add patient information as shown in figure.

6.5.3 Appointments

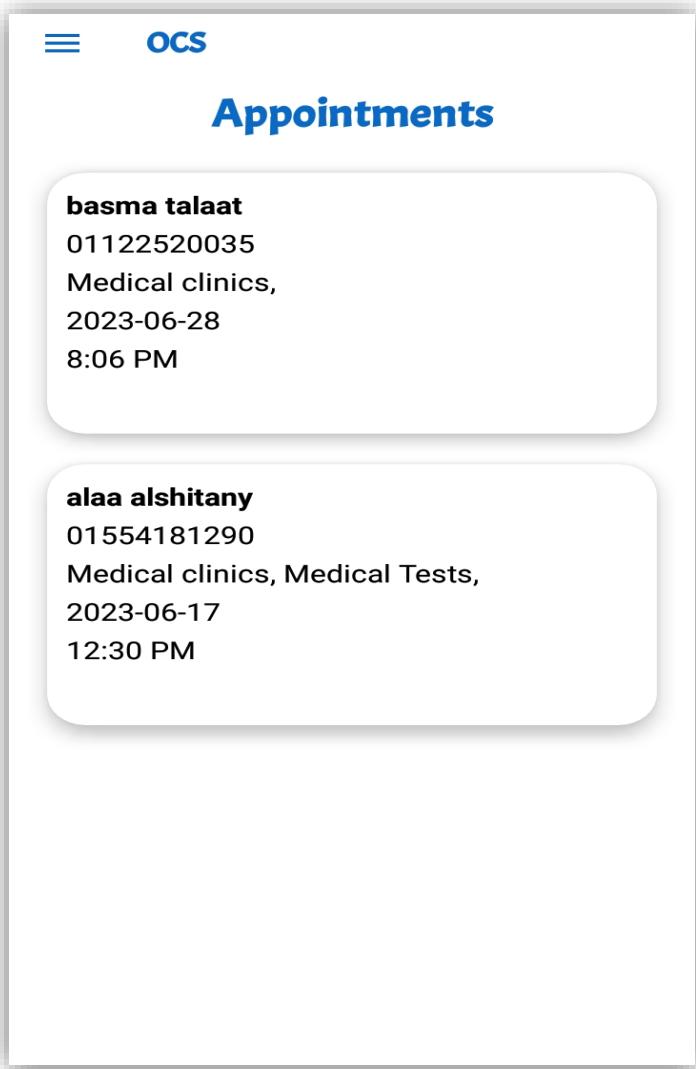


Figure 6_32: Appointments Page.

Doctor can view appointments when clicks on appointments in dashboard as shown in Figure 5_32, this page helps each doctor to his appointments.

6.5.4 Model

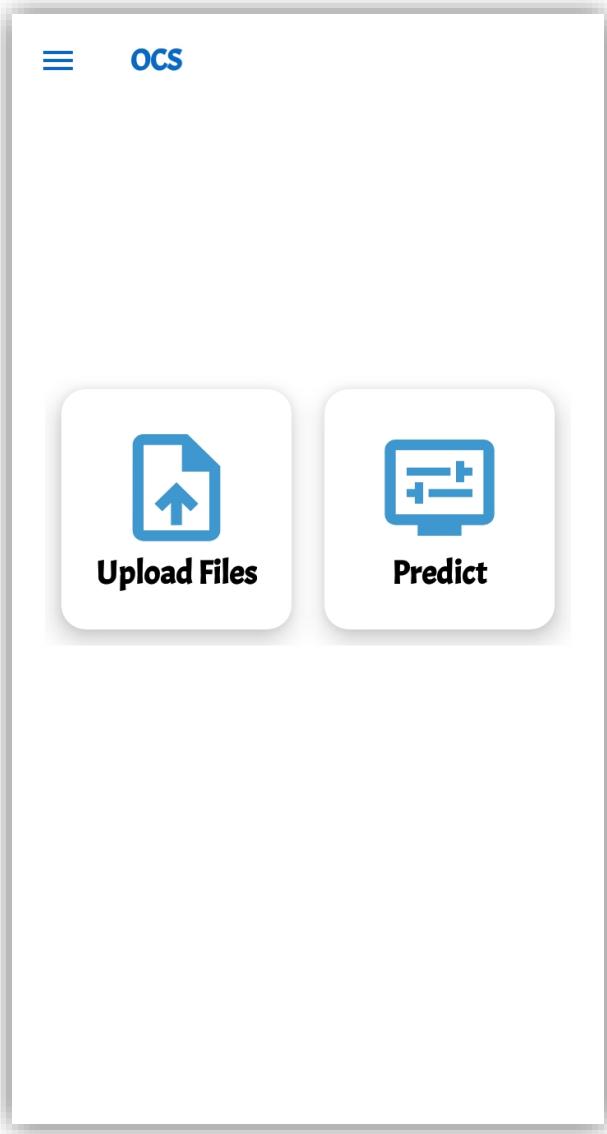


Figure 6_33: Model Page.

This is the most important part in our system as from this page doctor can upload files of patient and file of drug then the model will predict the sensitivity or resistance of drug.

OCS

Upload Files

Choose Patient Name

basma talaat

Upload Gene Express



Upload Genomic Mut:



Upload Methylation F



OCS

Drug Effectiveness

Choose Patient Name

basma talaat

choose Drug

(5Z)-7-Oxozeaenol

predict

Figure 6_34: Upload files Page.

Figure 6_35: Choose drug Page.

In Figure 5_34 doctor will upload files of patients which contain his genomic information (genomic expression file, genomic mutation and genomic methylation), then from next page (Figure 5_35) doctor will choose drug name then click predict.

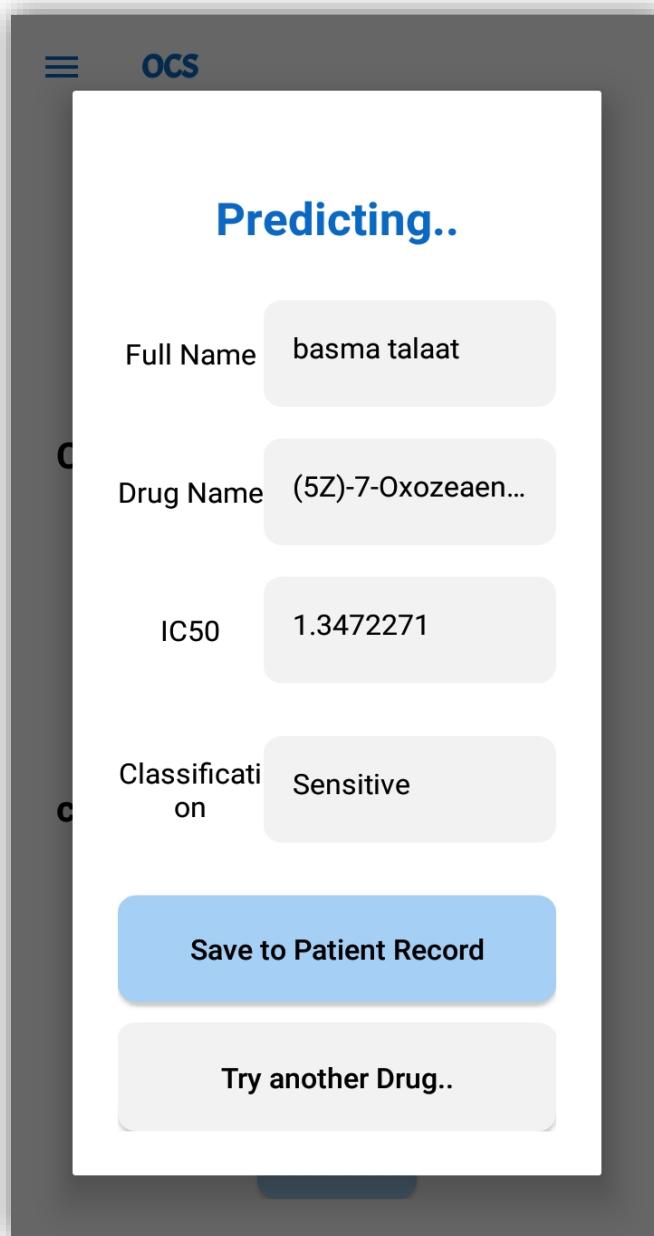


Figure 6_36: Predicting Page.

This page appears after clicking "predict" in previous figure, it contains Patient name, drug name, IC50 value and classification result (sensitive or resistance). Doctor can save prediction result in patient record, if he wants to try another drug he can click on "try another drug" button.

CONCLUSION

Diagnosing and treating cancer is not easy and needs the most effective tools to help doctors and patients. OCS is an app we developed to help in the field of cancer, and we hope it helps. We used deep learning techniques such as CNN and GCN. OCS can be presented as an application to explore drug sensitivity using large-scale, multistage cancer profiles. OCS outperforms several baselines, and our analysis demonstrates how our method can help prioritize therapeutic targets for anticancer drug discovery. And we have targeted Android devices for this application because most of the people in the world are using Android system, so the application will be more widely used and widespread, Which will facilitate the reservation process for the patient and save him a lot of time, and it will also be useful for the doctor, as it will help him in treatment and take the correct action for the patient by knowing whether the medicine will be effective with the patient or not.

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.

APPENDIX

PARTS OF CODE

```
#drug feature with GCN
GCN_layer = GraphConv(units=units_list[0],step_num=1)([drug_feat_input,drug_adj_input])
if use_relu:
    GCN_layer = Activation('relu')(GCN_layer)
else:
    GCN_layer = Activation('tanh')(GCN_layer)
if use_bn:
    GCN_layer = BatchNormalization()(GCN_layer)
GCN_layer = Dropout(0.1)(GCN_layer)

for i in range(len(units_list)-1):
    GCN_layer = GraphConv(units=units_list[i+1],step_num=1)([GCN_layer,drug_adj_input])
    if use_relu:
        GCN_layer = Activation('relu')(GCN_layer)
    else:
        GCN_layer = Activation('tanh')(GCN_layer)
    if use_bn:
        GCN_layer = BatchNormalization()(GCN_layer)
    GCN_layer = Dropout(0.1)(GCN_layer)

GCN_layer = GraphConv(units=100,step_num=1)([GCN_layer,drug_adj_input])
if use_relu:
    GCN_layer = Activation('relu')(GCN_layer)
else:
    GCN_layer = Activation('tanh')(GCN_layer)
if use_bn:
    GCN_layer = BatchNormalization()(GCN_layer)
GCN_layer = Dropout(0.1)(GCN_layer)
#global pooling
if use_GMP:
    x_drug = GlobalMaxPooling1D()(GCN_layer)
else:
    x_drug = GlobalAveragePooling1D()(GCN_layer)
```

Figure 7_1: GCN Layers.

The first GCN layer takes the drug feature input and the adjacency matrix input as inputs and applies a graph convolution with units_list[0] number of hidden units. The output tensor is then passed through an activation function (RELU), a batch normalization layer (if use_bn is True), and a Dropout layer. Subsequent GCN layers take the output tensor of the previous layer and the adjacency matrix input as inputs and apply another graph convolution with units_list[i+1] number of hidden units. The output tensor is then passed through the same sequence of activation, batch normalization, and Dropout layers. The final GCN layer applies a graph convolution with 100 hidden units and follows the

same sequence of activation, batch normalization, and Dropout layers. The output tensor of the final GCN layer is then passed through a global pooling layer (either GlobalMaxPooling1D or GlobalAveragePooling1D, depending on the use_GMP parameter) and concatenated with the other features.

```
def k_fold(data_idx, k = 5):
    data_train_idx,data_test_idx = [], []
    for each_type in TCGA_label_set:
        data_subtype_idx = [item for item in data_idx if item[-1]==each_type]
        fold_size = int(len(data_subtype_idx) / k)
        for i in range(k):
            test_list = data_subtype_idx[i*fold_size:(i+1)*fold_size]
            train_list = [item for item in data_subtype_idx if item not in test_list]
            data_train_idx +=train_list
            data_test_idx +=test_list
        print("split is done")
    return data_train_idx,data_test_idx
```

Figure 7_2:Train_Test_Split

This code defines a function Data Split that splits a list of data indices into training and testing sets for each subtype in a given label set. The function takes two parameters: data_idx, which is a list of data indices, and ratio, which is the proportion of data to use for training (default value is 0.80).

```
def DataSplit(data_idx,ratio = 0.80):
    data_train_idx,data_test_idx = [], []
    for each_type in TCGA_label_set:
        data_subtype_idx = [item for item in data_idx if item[-1]==each_type]
        train_list = random.sample(data_subtype_idx,int(ratio*len(data_subtype_idx)))
        test_list = [item for item in data_subtype_idx if item not in train_list]
        data_train_idx += train_list
        data_test_idx += test_list
    return data_train_idx,data_test_idx
```

Figure 7_3:K_Fold

This code defines a function k_fold that performs k-fold cross-validation on a list of data indices for each subtype in a given label set. The function takes two parameters: data_idx, which is a list of data indices, and k, which is the number of folds to use for cross-validation (default value is 5).

RUN SCREENS

We use Colab Pro to run our model.

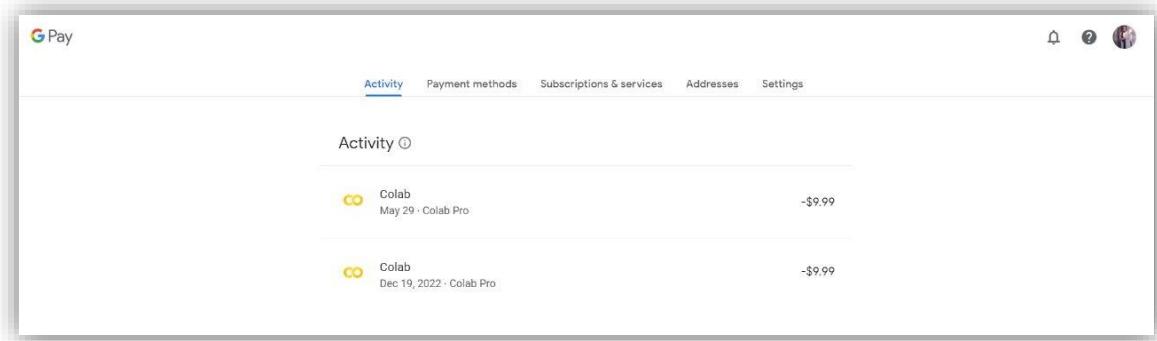


Figure 7_4: Colab Pro



Figure: 7_5 Run of classification (Train_Test_Split).

Appendix.

The screenshot shows a Jupyter Notebook cell with the title "Untitled0.ipynb". The code cell contains a series of training logs for a regression model. The logs show the progress of 148 epochs over 1400 steps. The output includes metrics like loss, mse, and pcc-val, along with execution times (e.g., 1s 5ms/step). The final message states: "The overall Pearson's correlation is 0.9200."

```

pcc-val: 0.92
1400/1400 [=====] - 2s 19ms/step - loss: 1.2052 - mse: 1.2052
Epoch 36/48
148/148 [=====] - 1s 5ms/step
pcc-val: 0.917
1400/1400 [=====] - 2s 19ms/step - loss: 1.2058 - mse: 1.2058
Epoch 37/48
148/148 [=====] - 1s 5ms/step
pcc-val: 0.9192
1400/1400 [=====] - 2s 19ms/step - loss: 1.1888 - mse: 1.1888
Epoch 38/48
148/148 [=====] - 1s 5ms/step
pcc-val: 0.9193
1400/1400 [=====] - 2s 19ms/step - loss: 1.1859 - mse: 1.1859
Epoch 39/48
148/148 [=====] - 1s 5ms/step
pcc-val: 0.9187
1400/1400 [=====] - 2s 19ms/step - loss: 1.1787 - mse: 1.1787
Epoch 40/48
148/148 [=====] - 1s 5ms/step
pcc-val: 0.9186
1400/1400 [=====] - 2s 19ms/step - loss: 1.1711 - mse: 1.1711
148/148 [=====] - 1s 6ms/step
The overall Pearson's correlation is 0.9200.

```

Figure: 7_6 Run of Regression (Train_Test_Split).

The screenshot shows a Jupyter Notebook cell with the title "Copy of Copy of classification(k-fold).ipynb". The code cell displays a series of training logs for a classification model using k-fold cross-validation. The logs show the progress of 196 epochs over 416 steps. Metrics include accuracy, precision, recall, f1_score, and average_precision. The output also includes ETA and warning messages about untraced functions. The final message indicates an overall AUC and auPR of 1.0000.

```

416/416 [=====] - 2s 5ms/step
roc-val: 0.9999, pr-val:1.0000
823/823 [=====] - 19s 23ms/step - loss: 0.0233 - accuracy: 0.9908 - precision: 0.9940 - recall: 0.9948 - f1_score: 0.9944 - average_precision: 0.9999
Epoch 16/20
820/823 [=====] - ETA: 0s - loss: 0.0192 - accuracy: 0.9926 - precision: 0.9950 - recall: 0.9960 - f1_score: 0.9955 - average_precision: 0.9999WARNING:tensor
416/416 [=====] - 2s 5ms/step
roc-val: 0.9999, pr-val:1.0000
823/823 [=====] - 19s 24ms/step - loss: 0.0192 - accuracy: 0.9926 - precision: 0.9951 - recall: 0.9960 - f1_score: 0.9955 - average_precision: 0.9999
Epoch 17/20
820/823 [=====] - ETA: 0s - loss: 0.0172 - accuracy: 0.9940 - precision: 0.9962 - recall: 0.9966 - f1_score: 0.9964 - average_precision: 0.9999WARNING:tensor
416/416 [=====] - 2s 5ms/step
roc-val: 1.0000, pr-val:1.0000
823/823 [=====] - 20s 24ms/step - loss: 0.0172 - accuracy: 0.9941 - precision: 0.9962 - recall: 0.9966 - f1_score: 0.9964 - average_precision: 0.9999
Epoch 18/20
820/823 [=====] - ETA: 0s - loss: 0.0172 - accuracy: 0.9940 - precision: 0.9962 - recall: 0.9966 - f1_score: 0.9964 - average_precision: 0.9999WARNING:tensor
416/416 [=====] - 2s 5ms/step
roc-val: 0.9999, pr-val:1.0000
823/823 [=====] - 20s 24ms/step - loss: 0.0172 - accuracy: 0.9941 - precision: 0.9962 - recall: 0.9966 - f1_score: 0.9964 - average_precision: 0.9999
Epoch 19/20
820/823 [=====] - ETA: 0s - loss: 0.0150 - accuracy: 0.9950 - precision: 0.9965 - recall: 0.9974 - f1_score: 0.9969 - average_precision: 0.9999WARNING:tensor
416/416 [=====] - 2s 5ms/step
roc-val: 1.0000, pr-val:1.0000
823/823 [=====] - 19s 23ms/step - loss: 0.0150 - accuracy: 0.9950 - precision: 0.9965 - recall: 0.9974 - f1_score: 0.9969 - average_precision: 0.9999
Epoch 19/20
820/823 [=====] - ETA: 0s - loss: 0.0164 - accuracy: 0.9943 - precision: 0.9966 - recall: 0.9965 - f1_score: 0.9966 - average_precision: 0.9999WARNING:tensor
416/416 [=====] - 2s 5ms/step
roc-val: 1.0000, pr-val:1.0000
823/823 [=====] - 19s 23ms/step - loss: 0.0164 - accuracy: 0.9943 - precision: 0.9966 - recall: 0.9966 - f1_score: 0.9966 - average_precision: 0.9999
Epoch 20/20
820/823 [=====] - ETA: 0s - loss: 0.0123 - accuracy: 0.9959 - precision: 0.9973 - recall: 0.9978 - f1_score: 0.9975 - average_precision: 1.0000WARNING:tensor
416/416 [=====] - 2s 5ms/step
roc-val: 0.9999, pr-val:1.0000
823/823 [=====] - 19s 23ms/step - loss: 0.0125 - accuracy: 0.9959 - precision: 0.9972 - recall: 0.9978 - f1_score: 0.9975 - average_precision: 1.0000WARNING:absl:Found untraced functions such as _jit_compiled_convolution_op, _jit_compiled_convolution_op, _jit_compiled_convolution_op, _jit_compiled_convolution_op, _jit_compiled_cc
model saved
416/416 [=====] - 2s 5ms/step
The overall AUC and auPR is 1.0000 and 1.0000.

```

Figure: 7_7 Run of classification (K_Fold).

```
Copy of Regression(k-fold).ipynb ☆
File Edit View Insert Runtime Tools Help Cannot save changes
+ Code + Text Copy to Drive
✓ 204/204 [=====] - 1s 5ms/step
pcc-val: 0.9793
411/411 [=====] - 9s 23ms/step - loss: 0.4497 - mse: 0.4497
Epoch 15/20
204/204 [=====] - 1s 5ms/step
pcc-val: 0.9785
411/411 [=====] - 10s 24ms/step - loss: 0.4183 - mse: 0.4183
Epoch 16/20
204/204 [=====] - 1s 5ms/step
pcc-val: 0.9813
411/411 [=====] - 10s 23ms/step - loss: 0.3939 - mse: 0.3939
Epoch 17/20
204/204 [=====] - 1s 5ms/step
pcc-val: 0.9842
411/411 [=====] - 10s 24ms/step - loss: 0.3821 - mse: 0.3821
Epoch 18/20
204/204 [=====] - 1s 5ms/step
pcc-val: 0.9844
411/411 [=====] - 9s 23ms/step - loss: 0.3647 - mse: 0.3647
Epoch 19/20
204/204 [=====] - 1s 5ms/step
pcc-val: 0.9855
411/411 [=====] - 9s 23ms/step - loss: 0.3716 - mse: 0.3716
Epoch 20/20
204/204 [=====] - 1s 5ms/step
pcc-val: 0.9843
411/411 [=====] - 9s 23ms/step - loss: 0.3471 - mse: 0.3471
WARNING:absl:Found untraced functions such as _jit_compiled_convolution_op, _jit_compiled_convolution_op, _jit_compiled_convolution_op, _jit_compiled_convolution_op, _jit_compiled_convolution_op
model saved
204/204 [=====] - 1s 5ms/step
The overall Pearson's correlation is 0.9855.
```

Figure: 7_8 Run of Regression (K_Fold).



كلية الحاسوب والذكاء
الاصطناعي



جامعة بنها

خدمات مركز الأورام وتقنيات التنبؤ لمساعدة في اتخاذ القرار الصحيح

مشروع التخرج المقدم استيفاءً جزئياً لمتطلبات درجة بكالوريوس الحاسوب والذكاء الاصطناعي

قسم المعلوماتية الطبية

فريق المشروع

لاء وليد محمد الشيتاني

ايمان طارق طه علي

بسمه طلعت محمود محمد

عبدالرحمن احمد عبدالهادي

يارا احمد محمد سويسه

تحت اشراف

الاستاذ الدكتور / السيد بدر

المهندس / محمد الحكيم

ملخص المشروع Project Abstract

السرطان هو أي مرض من بين العديد من الأمراض التي تتميز بتطور خلايا غير طبيعية تنقسم بشكل لا يمكن السيطرة عليه وهو السبب الرئيسي لوفاة 10 ملايين شخص ، أو وفاة واحدة تقريباً من بين كل 6 وفيات. من الصعب التنبؤ الدقيق بالاستجابة للأدوية السرطان بسبب عدم التأكيد من فعالية الأدوية وعدم تجانس مرض السرطان. إن إبراز مراكز الأورام والاهتمام بحل المشكلات الموجودة أمر مهم وضروري ، وهذا ما نحاول القيام به في مشروعنا. نحاول تنظيم عملية تحديد المواعيد ومساعدة الأطباء على اختيار الدواء المناسب بكفاءة باستخدام نموذج التعلم العميق.

لقد قمنا بدمج قاعدة بيانات عامتين في دراستنا: CCLE و GDSC. توفر لنا قاعدة بيانات بالنسبة لبيانات فحص الأدوية على نطاق واسع ، عندما تكون قيمة IC₅₀ صغيرة ، فهذا دليل على أن فعالية الدواء عالية. توفر لنا قاعدة بيانات CCLE بياناتجينومية ونسخة وجينية لأكثر من ألف خط من الخلايا السرطانية. بالنسبة إلى مجموعات البيانات الثلاثة ، ركزنا على بيانات الطفرات الجينية وبيانات التعبير الجيني وبيانات مماثلة للحمض النووي. يمكننا الحصول على بيانات omics باستخدام ميكروأري.

الهدف من النموذج هو تحديد ما إذا كان الدواء فعالاً أم لا لمريض السرطان اعتماداً على ملامح الخلايا السرطانية (الطفرة الجينية ، التعبير الجيني وبيانات مماثلة للحمض النووي). في النموذج استخدمنا CNN و GCN. والنتيجة هي أن العملية تصبح أكثر دقة للأطباء لتحديد ما إذا كان الدواء سيكون فعالاً أم لا لمريض معين وفعالية الدواء. من خلال تطبيق أندرويد الذي يقوم بتطويره ، يمكن للأطباء العثور عليه بسهولة لإظهار مواعيدهم بطريقة منتظمة واستخدام أحد التطرق لتشخيص المريض والتعرف على المرض. أيضاً ، يمكن للمرضى تحديد موعد عبر الإنترنت واختيار الخدمات التي يحتاجون إليها مثل الاختبارات والأشعة السينية أو التشخيص والأدوية. فيما يتعلق بالنموذج ، أو لاً ، استخدمنا train_test_split لتقسيم البيانات K-أضعاف. ثانية،كنا KNN مع GCN ، شجرة القرارات k-fold Classification بإستخدام ال regression مع GCN ، الغابة العشوائية مع GCN. دقة 1.0000 هي 0.9996.