

# *Introducing A Cost-Effective Cancer Registry System For Bangladesh*

by

Md. Tareq Hassan

18301124

Rahmat Ullah Sujan

18201081

Tushar Imran Sarkar

18201060

Gazi Sharin Mostafiz

18301123

Mohammad Jonayed Talukder

18301186

A thesis submitted to the Department of Computer Science and Engineering  
in partial fulfillment of the requirements for the degree of  
B.Sc. in Computer Science

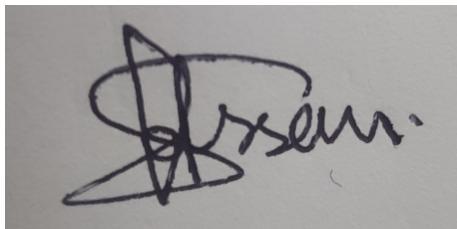
Department of Computer Science and Engineering  
Brac University  
September 2022

## Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

**Student's Full Name & Signature:**



---

Md. Tareq Hassan  
18301124



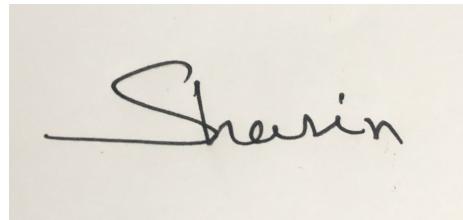
---

Rahmat Ullah Sujan  
18201081



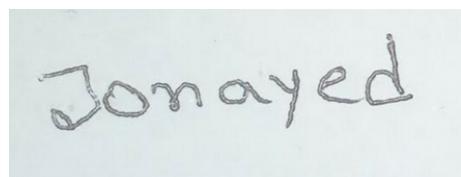
---

Tushar Imran Sarkar  
18201060



---

Gazi Sharin Mostafiz  
18301123



---

Mohammad Jonayed Talukder  
18301186

# Approval

The thesis titled “*Introducing A Cost-Effective Cancer Registry System For Bangladesh*” submitted by

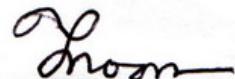
1. Md. Tareq Hassan (18301124)
2. Rahmat Ullah Sujan (18201081)
3. Tushar Imran Sarkar (18201060)
4. Gazi Sharin Mostafiz (18301123)
5. Mohammad Jonayed Talukder (18301186)

Of Summer, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 20, 2022.

## Examining Committee:

Supervisor:

Jannatun Noor



---

Jannatun Noor  
Senior Lecturer

Department of Computer Science and Engineering  
BRAC University

Thesis Coordinator:

Md. Golam Rabiul Alam, PhD

---

Md. Golam Rabiul Alam, PhD  
Professor

Department of Computer Science and Engineering  
BRAC University

Head of Department:

Sadia Hamid Kazi

---

Sadia Hamid Kazi, PhD  
Chairperson and Associate Professor  
Department of Computer Science and Engineering  
Brac University

## **Abstract**

Cancer is an interneccine disease caused by the uncontrollable growth of cells which spreads to the other parts of the body. There are a lots of variant of cancer invented where some of them are extremely severe and untreatable. In Bangladesh, there are almost thirteen to fifteen lacs patients suffering from several cancers and about two lac patients are newly diagnosed with cancer each year. A central cancer registry system is essential to provide these patients with more accurate diagnoses and treatments as well as to preserve their data for research purposes. This system should compile data on the basis of demographics, diagnoses, tumor histology, treatment, and outcomes into one central database. Moreover, some dispersed hospital-based registry systems have been introduced already but there is no central database for treatment and diagnosis information of patients which can be accessed by registered users. For this reason, we made the decision to conduct research on a cancer registry system that is both affordable and effective because it will give patients and doctors the information they require. We claim it as a user friendly system because before building the system we did a survey from both doctor and patients to know how they are comfortable to use this system. Moreover, we have intended to implement OCR in our system, which can extract data from any types of documents. In addition, for a developing country like Bangladesh a digitalized cancer registry system is only practical when it is cost-effective. Also, using individual cards with distinctive QR codes for patient login and retrieval of data from the main database is an ideal solution for Bangladesh that can be put into use practically with the help of the government.

## **Acknowledgement**

This work was conducted at and thoroughly supported by BRAC University. To begin with, all praise to the Great Allah for keeping us safe and healthy throughout our whole research time. We would take this opportunity to thank our supervisor Jannatun Noor ma'am for her efforts to guide us and giving us the opportunity to work under her. We would like to express our gratitude towards Dr. Md. Golam Zel Asmaul Husna (Resident, Radiation oncology, National Institute of Cancer Research and Hospital) and Shahnewaz Rifat(Solutions Architect at Virtasant) Sir for guiding us along with providing advice throughout the research. Also, they give us tremendous support from the very beginning. And finally, to our parents for their kind support and prayers.

# Table of Contents

<b>Declaration</b>	i
<b>Approval</b>	ii
<b>Abstract</b>	iii
<b>Acknowledgment</b>	iv
<b>Table of Contents</b>	v
<b>List of Figures</b>	vii
<b>List of Tables</b>	1
<b>1 Introduction</b>	2
1.1 Thoughts Behind the Scene . . . . .	2
1.2 Research Objectives . . . . .	3
1.3 History . . . . .	5
1.4 Research Problem . . . . .	7
1.5 Research Contribution . . . . .	8
1.6 Thesis Outline . . . . .	10
<b>2 Related Works</b>	11
<b>3 Methodology</b>	17
3.1 Conducting the Survey . . . . .	17
3.1.1 Survey Questionnaire Design . . . . .	17
3.1.2 Approaching the Target Audiences . . . . .	19
3.1.3 Analyzing the Survey Result . . . . .	20
3.2 Designing the Prototype . . . . .	20
3.2.1 Scan QR Code . . . . .	21
3.2.2 Patient's Page . . . . .	21
3.2.3 Convert Image to Text . . . . .	22
3.2.4 Find Necessary Data From .txt File . . . . .	27
3.2.5 Store Data to Database . . . . .	27
<b>4 Experimental Evaluation</b>	29
4.1 Analysis On Survey Results . . . . .	29
4.1.1 Outcome From the Interviews With Patients . . . . .	29
4.1.2 Outcome From the Interviews With Doctors . . . . .	32

4.1.3	Quantitative Data Analysis From Survey . . . . .	33
4.2	Analysis on System Output . . . . .	37
4.2.1	Quantitative Analysis on Data-sets . . . . .	37
4.3	Estimated Cost of the Proposed System . . . . .	40
<b>5</b>	<b>Design &amp; Implementation</b>	<b>42</b>
5.1	Design of Prototype . . . . .	42
5.1.1	Doctor Panel . . . . .	43
5.1.2	Patient Panel . . . . .	46
5.1.3	Admin Panel . . . . .	48
5.2	Implementation of the Prototype . . . . .	50
<b>6</b>	<b>Discussion</b>	<b>52</b>
6.1	Understanding the Context of the Complications of Patients . . . . .	52
6.2	Designing According to the Requirements . . . . .	53
6.3	Comparison With Existing Works . . . . .	54
6.4	Limitation . . . . .	57
<b>7</b>	<b>Conclusion And Future Work</b>	<b>58</b>
<b>References</b>		<b>63</b>
<b>APPENDIX A: Survey Questionnaire For Patients</b>		<b>64</b>
<b>APPENDIX B: Survey Questionnaire For Doctors</b>		<b>67</b>

# List of Figures

1.1	Worldwide cancer registries [47] . . . . .	4
2.1	How API works. . . . .	13
3.1	Flowchart of methodology . . . . .	21
3.2	Flow Chart of OCR Process. . . . .	22
3.3	Binarization of an image. . . . .	23
3.4	A shows before De Skew and B shows after De Skew. . . . .	24
3.5	Effect of De Speckle. . . . .	24
3.6	Before And after Line Removal. . . . .	25
3.7	Pattern Recognition. . . . .	26
3.8	Flow chart of the proposed system . . . . .	28
4.1	Frequency of patients for getting diagnosed. . . . .	30
4.2	Opinion of patients on difficulty to manage reports. . . . .	31
4.3	Opinion of patients for a central cancer registry system. . . . .	31
4.4	Time spent on tech devices. . . . .	32
4.5	Average Difference between two data-sets. . . . .	38
4.6	Closeness Between Two Datasets in Percentage. . . . .	38
4.7	Exactness Between Two Datasets in Percentage. . . . .	39
5.1	UI of Doctor Panel. . . . .	43
5.2	UI of Diagnosis Panel. . . . .	44
5.3	UI of Blood Test Report Page. . . . .	45
5.4	UI of CT Scane Page. . . . .	46
5.5	Patient's ID Card. . . . .	46
5.6	UI of Patient Profile Page. . . . .	47
5.7	UI of Adding Blood Test Page. . . . .	47
5.8	UI of Medicine Page. . . . .	48
5.9	UI of Admin Dashboard. . . . .	49
5.10	UI of Doctor List Page. . . . .	49
5.11	UI of Patient List Page. . . . .	50
5.12	Homepage Of Webpage. . . . .	50

# List of Tables

4.1	Observed Data for dea about card-based registry system VS opinion on using ID card for easier accessibility. . . . .	34
4.2	Expected Data for dea about card-based registry system VS opinion on using ID card for easier accessibility. . . . .	34
4.3	Chi-square calculation . . . . .	35
4.4	Critical value calculation for Idea about card-based registry system VS opinion on using ID card for easier accessibility. . . . .	35
4.5	Opinion on the difficulty of managing hard copies of report VS A system storing all diagnosis data is helpful. . . . .	35
4.6	Opinion on the difficulty of managing hard copies of report VS A system storing all diagnosis data is helpful. . . . .	36
4.7	Critical value calculation for Opinion on the difficulty of managing hard copies of report VS A system storing all diagnosis data is helpful. . . . .	36
4.8	Observed Data Idea about card-based registry system VS opinion on using a patient's ID card for easier accessibility. . . . .	36
4.9	Expected Data Idea about card-based registry system VS opinion on using a patient's ID card for easier accessibility. . . . .	36
4.10	Critical value calculation for Idea about card-based registry system VS opinion on using a patient's ID card for easier accessibility. . . . .	37
6.1	Comparison of the proposed system with existing works[18], [39], [42], [50], [52]. . . . .	56

# Chapter 1

## Introduction

### 1.1 Thoughts Behind the Scene

There has been research done by ‘Globocan’ regarding cancer that in 2020, 156,775 affected by cancer in Bangladesh. Among them 108,990 people became died. Moreover, by doing research we got to know that there are few cancers registry system in Bangladesh and most of them are owned by private organizations, only one registry system is owned by government which is not that much enriched. Furthermore, general people don’t have access in those registry system and they only kept the personal information of patient like name, age, gender, occupation etc. For this reason, we decided to do research on the cost-effective cancer registry system which will provide the necessary data both doctor and patients according to their needs. In many nations, cost-effectiveness analysis is used to evaluate public healthcare systems. Cost-effectiveness analysis method was used to calculate the lifetime cost of cancer per quality-adjusted life year for various cancer locations and chose patients under healthcare as a reference group in order to perform a semi-incremental cost-effectiveness ratio. As a comparison, findings were compared to those calculated from patients with end-stage renal disease and those requiring 4 long-term mechanical ventilation.

The primary function of a cancer registry is to keep track of cancer patients and their treatment progress [24]. Systems that manage and analyze data on cancer patients and survivors are known as cancer registries. In order to provide health officials with reliable and timely data on cancer incidence, treatment, and survivorship, cancer registries are maintained [53]. The three main categories of cancer registries are as follows. They are Population-Based Registries, Special Registries, and Hospital and Ambulatory Surgical Treatment Center (ASTC)-Based Registries. All patients who have been diagnosed and/or treated at a facility are included in the databases of hospital- and ASTC-based cancer registries. All cancer patients in a specific geographic area are included in population-based cancer registries, also known as central registries. These registries could be focused on administration, research, or cancer prevention. Data on one aspect or one type of cancer, such as bone tumors, brain tumors, or pediatric tumors, is kept in special cancer registries. They frequently offer assistance to those who may be affected by cancer as well as educational opportunities for those who wish to learn more about a certain type of cancer [51].

To design a cost effective and user-friendly cancer registry system we divide our registry system into three different panels including two databases. They are Doctor, patient and admin. In the patient panel, a patient can do the registry by filling some relevant question. Besides, a registered user can see his all types of uploaded report like MRI, CT scan, blood test and so on. As like patient panel a doctor can also do registry by providing his/her information. Another feature of doctor panel is, here we have added a QR scan system which will be provided to patients and each patient gets a different QR code card from the admin. By using a QR code a doctor can easily get access to his patient's profile which will save the time and decrease the suffering for both doctor and patients. After that, we have added an OCR system in our design. By using this OCR, when a user uploads any picture of report, it will extract relevant data from that picture and stored in the database. Next, in doctor panel there is a option for searching, by which a doctor can easily search his any patient information and go through it according to his/her needs. Finally, there will be some administrator who will have all the editable access and they will provide the QR code to the patient. Thus, we have tried to design a cost-effective cancer data registry system.

## 1.2 Research Objectives

The purpose of this research is to develop a cost-effective cancer registry system that includes a central database containing advanced structured data on cancer patients' diagnosis and treatment. As we mentioned before, a significant national burden is cancer. It is a great and noble goal to lessen the burden of cancer in the country. Many people are involved in this effort, including medical professionals, epidemiologists, public health planners, legislators, medical students, and others. In their quest to win the 'War on Cancer', each of these individuals' values and relies on cancer data. In order to better understand the causes of cancer and find cancer earlier, doctors need cancer data. This increases the likelihood that a cure will be discovered. Cancer specialists base their treatment decisions on accurate cancer information from sources like pathologist and cytologist reports. Cancer data may point to environmental risk factors or high-risk behaviors, so preventive measures can be taken to reduce the number of cancer cases and resulting deaths. Local, state, and national cancer agencies and cancer control programs also use registry data from defined areas to make important public health decisions that maximize the effectiveness of limited public health funds, such as the implementation of screening programs. The cancer registry's lifetime follow-up program is an additional crucial component. Current patient follow-up provides accurate survival data and serves as a reminder to doctors and patients to schedule routine clinical examinations. In a nutshell, cancer registries are significant because they gather precise and comprehensive cancer data that can be used for cancer control and epidemiological research, public health program planning, and patient care enhancement. Actually, all of these initiatives lessen the impact of cancer. Bangladesh is still a developing country, which is why the system's cost-effectiveness is a primary concern. What determines their role is the type of data that population-based CRS (Cancer Registry Systems) may provide. Additionally, advanced plans can collect information on

diagnostic and treatment delays, treatment methods and medical equipment used, and compliance with clinical care recommendations. As a result, CR'S are critical in. A sub field of epidemiology that examines patterns in cancer incidence, survival, and prevalence rates over time and across geographic regions, socioeconomic groups, and time intervals. A study of the etiology of cancer that examines the relationship between a variety of social and environmental factors and the risk of developing cancer. The state cancer registry collects, compiles and analyzes cancer data. Each year, the country's cancer registries record more than 1.6 million cases. Cancer registries enable researchers to monitor the disease's prevalence, gain insight into its etiology, and develop treatments. The National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER) collects data from regional and state registries [6], [7]. By utilizing these registries, health care providers and researchers can gain a better understanding of cancer trends and causes, as well as racial/ethnic disparities and occupational hazards.

A cancer control strategy, including the prioritization of various measures in light of the current and projected cancer burden. A population-based cancer registry is required in Bangladesh because the country currently lacks an enriched cancer data registry system, and we wish to establish one that will assist anyone conducting cancer research in obtaining sufficient data from the registry system. According to 'iccp-portal.org', across the world more than 700 cancer registry systems are available where only 8% of total population of Asia and 21% of world population is covered by the population-based cancer registries [47]. And there is no central cancer registry system in Bangladesh.

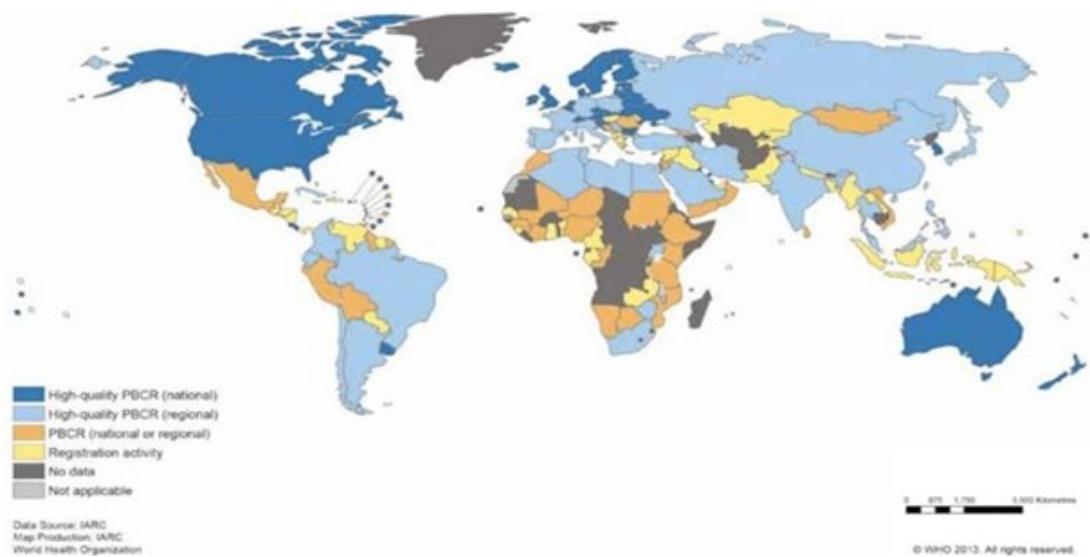


Figure 1.1: Worldwide cancer registries [47].

Developing a cost-effective cancer registry system, which includes a central database containing advanced structured data on cancer patients' diagnosis and treatment, is the goal of this research. Bangladesh is still considered a developing country, and as a result, the cost-effectiveness of the system is of primary concern. The type of data that population-based CRS (Cancer Registry Systems) may provide the role to patients. Additional information collected by advanced plans includes

diagnostic and treatment delays, treatment methods and medical equipment used, and compliance with clinical care recommendations. As a result, CRS is critical for the fight against the epidemiology of cancer. Therefore, the objectives of the cancer registry system are given below:

- To introduce a well-developed cancer registry system for storing cancer data.
- The system will be cost effective because there are so many poor people in Bangladesh, for this reason, we have to make it affordable for them.
- User will get all types of information regarding cancer.
- Hassle free and user-friendly digital system which will reduce people suffering.
- To introduce an easily manageable and convenient ID card containing unique QR code used as a portal for storing and retrieving patient's data.
- To ensure continuous care to patients and provide timely information through dynamic recommendations.
- Determine whether measures to prevent cancer and enhance patient outcomes have a significant impact on cancer incidence and survival. This objective requires the gathering of information about the prevalence and trends of risk variables.
- There will be enough information in this registry system to assist anyone who intends to initiate cancer research of any kind. A large number of cancer researchers in Bangladesh are currently interested in conducting cancer research, but due to a lack of sufficient data, almost all of these researchers are unable to proceed.
- The system will assist in the detection of cancer in its early stages, allowing patients to receive treatment as soon as possible. As we all know, cancer is a major cause of morbidity in almost all cases, owing to the fact that most people only find out about it when they are in the final stages of the disease. Aside from that, people do not have enough knowledge about the disease.

### 1.3 History

Cancer registration is the process of compiling comprehensive data on all types of cancer and cancer patients from a variety of sources. The past history of cancer registry was using a written, organized registry of man's ailments to better understand them is an idea that has been around since at least the late 16th century. The English Crown assigned elderly, epidemic-torn women to scour the countryside for the dead and dying as a result of the plague's seemingly arbitrary geographic ravages. The "Ancient Matrons" tabulated deaths by causes like "the purples" (likely leukemia), "riting of the lights," "consumption" (often a side effect of cancer), and of course, the plague in weekly "Bills of Mortality" for each parish. There is no documentation of the precise usage of this data. Perhaps knowing where the plague was present helped the aristocracy move to a different location. By critically and

mathematically analyzing decades' worth of mortality data, a London businessman named John Graunt created medical history around the year 1665. By publishing a pamphlet with 108 conclusions, he published what can be considered the first work on medical epidemiology and statistics.

Revolutionary findings on the list included the fact that women visited doctors twice as often as men but lived longer and that plague epidemics spread from swampy areas. By sloppily fitting data samples to a logarithmic curve, Graunt was also the first to use mortality statistics to project population survival. John Graunt was the first non-scientist to be appointed to the Royal Academy as a result of his work. The first actuarial tables for a developing French life insurance industry were produced a few years later with the help of Sir Edmund Halley, of comet fame. The foundation of the analysis of cancer outcomes is now the 'life table' idea. Early in the 20th century, personal doctor or institutional projects in the United States or Europe gave rise to the first contemporary case registries for the study of cancer. The American College of Surgeons (ACoS) formally adopted a policy in 1956 to support the creation of hospital-based cancer registries through their Approvals Program. It was thought that by routinely examining the outcomes of cancer treatment regimens, hospitals and doctors might identify flaws in regional patterns of care and ultimately come to a better understanding of the illness and its treatment. Early hospital registries were, for the most part, inaccessible card files of data. Large central registry systems like the NCI SEER program provided the most useful information. When micro-computer registry systems first appeared in the 1980s, a new window of opportunity opened up for using registry data for the benefit of patients. Data from various registries can now be combined as part of projects like the National Cancer Data Base (NCDB) of the Commission on Cancer.

The United States collects data on cancer incidence and types, anatomic location and disease extent at the time of diagnosis, cancer patient treatment, and cancer patient outcomes using this method [24]. A cancer registry is a database that collects information on cancer incidence in the general population. This information is then used for clinical and epidemiological research purposes. Statistical analysis, targeted investigations, regulatory and program evaluation, are all possible uses for the information gathered during the investigation process. As a result, a cancer registry is now considered to be a requirement for modern health information systems. When it comes to gathering information for cancer prevention efforts, the cancer registry is a low-cost and highly effective tool.

More than 250 population-based cancer registries are currently operational in more than 60 countries around the world, representing a significant increase from previous years. Clinical and epidemiological research is increasingly reliant on the use of registries, which are becoming more common [11]. Moreover, in different countries around the world, there are different types of Cancer registry systems applied and used to preserve data. Among them, population-based, semi-population based and hospital-based Cancer registry systems are mostly used. Hospital-based registries are divided into two types. They are a single hospital registry system and collective registry system. Cancer registries located in hospitals keep track of all patients who have been treated for cancer. The clinical and administrative aspects

of cancer care are the primary focus of the hospital-based cancer registry. Patients' care, professional education, administrative data, and clinical research are all aimed at improving through the use of hospital-based registries. For epidemiological and public health purposes, population-based registries keep record of new cases within a defined population (typically a state). Registries can be generic (for all types of cancers) or narrowly focused (restricted to a given site-group or age group). There are some specific aims of population-based registers. They are cancer prevention, Early detection, Cancer rates and trends, care and outcome patterns, Research and Control effort evaluation. We can gain a better understanding of how cancer affects different populations and sub-populations with their assistance, develop more effective strategies for preventing it, better allocate scarce health resources for its treatment, gain a better understanding of how to prevent it over time, and advance the fields of cancer prevention and control research. The data from the population-based registry are used for a broader range of purposes than the data from the hospital-based registry. A population-based cancer registry benefits cancer prevention, treatment, administration, and research. The distribution of late-diagnosed cancer cases can be tracked using data from population-based registries, notably among communities, ethnic groups, ages, and another demographic group.

## 1.4 Research Problem

Cancer treatment is a long-term process and expensive. In the present situation of Bangladesh, there are only two-time points for the data collection of the cancer patients which are at diagnosis and at death. There is no central registry system available for capturing the whole picture of the treatment process of cancer. As a result, the cancer treatment process of Bangladesh is in a chaotic situation which is a headache for most of the cancer patients. Besides, all the diagnostic, test and treatment reports are in the form of hard-copy which creates a tenacious state for both patient and doctor to find the exact spying document or information. However, there are some registry systems available in some expensive hospitals but the collected information can be only accessed within that hospital. Because of that the patients get into big trouble when they change hospitals for various reasons.

Moreover, the inefficiency of the process of collecting data is not only the problem. Sometimes the patients are not willing to share their personal information with others for various reasons. Furthermore, the people of rural areas are totally unaware of the registry system. Besides, there are so many people in our country who are not able to find the proper way to get better treatment. For this reason, they have to go through a long-suffering period. As we do not have any previous data regarding Cancer updates. So, at the beginning stages, doctors were unable to detect the cancer symptoms. At the same time, poor patients waste their valuable money by doing different pathological tests. By doing these things they spent their rest of money and when doctors figure out cancer either it's too late or the money for continuing the rest of the treatment remains unsuccessful. As a result, they fall into dying with miserable conditions. According to the journal "Danish Cancer Registry", Medical practitioners and management must be actively involved in learning about the cancer registry's purpose in order to improve the quality of reported cancer care.

In order to achieve quality-care requirements, outpatient treatment documentation and timeliness may be enhanced by informing clinicians about the need of reporting information necessary for hospital registries [16].

The vital need of a central cancer registry system can be perceptible through the misery of the cancer patients. Bangladesh is a developing country where approximately 63 million people are still living under the poverty line and it is estimated that cancer related death will increase to 13% within 2030 [21]. As there is no digitized way to store a patient's diagnosis, test and treatment report so they are amassed in the form of hard-copies which are very difficult to preserve for a long time as cancer treatment is a long-term process. As a result, some patients lose their documents which leads them to suffer more. Because when a doctor wants to check history or a patient changes hospitals for some reason, they may have to do some tests again by increasing the amount of treatment cost which becomes a massive burden for patients. Hence a digitized cancer registry system and the use of less expensive devices for treatment is of frantic concern [32].

In addition, there is no quality of information for researchers to conduct research. They have to look up to the old surveys which sometimes lead them to reach the wrong conclusion in the perspective of the present day. Because the epidemiology of cancer is dependent on many variables which change continuously. Thus, the survey of one period might not be valid or incorrect for another time. And the existing surveys have approximate values as most of them are collected from hospital diagnosis count and death causes. As a result, there is no collection of continuously updating data which will exactly refer to the current situation of this deadly epidemiology. A thesis journal “Quality of Cancer Registry Data” states that, the efficient cancer care is mostly dependent on the quality of data [25].

Furthermore, while in the process of treatment a patient has to undergo several tests and therapies. And there are several instructions needed to be followed from the previous day of the test or therapy. But most of the patients are unknown about those instructions. As a result, they do not follow the procedure that was required which leads them to delaying the treatment process. Hence, cancer care functionality with the cancer register system is a blessing for those patients. According to a paper “MyPath: Investigating Breast Cancer Patients’ Use of Personalized Health Information”, it is necessary to receive and provide timely information through dynamic recommendations. Because, sometimes the treatment may have several side-effects which need to be monitored and nurtured continuously [31].

## 1.5 Research Contribution

Each developed system in today’s world has its own features which makes the system unique compared to others. These features are configured based on the demand and needs of the end users of the system. In our case we also want to develop a fully useful system for our users. For doing this we do our analysis on our user group and based on their feedback we try to develop a good prototype. In our proposed prototype we successfully able to implement some special features which are described

below:

- We maintained an user-friendly UI. Good and standard systems always maintain a user friendly UI. A system is said to be user friendly if it provides an environment or creates an atmosphere to the user by which the functions of that system become easier for the user to use and understand. In our case we divide the full system in three parts: First one, the admin panel where the system administrator can control the whole system activities. Another one is the patient panel where a patient can log in to see and upload his/her updated treatment related information. And the last one is the doctor panel where each doctor can log in to see and upload patients all treatment related information in an arranged way. In this way we maintain the simpler user base for easier accessibility to the system.
- We integrate QR code entry system into our system. We maintained user login flexibility in our system. As we developed our system specially in the context of Bangladesh. So we have to consider the user's capability. In our country a lot of people are illiterate who can not able to write their own name but they also use our system. For their perspective when they go to the doctor and want to access his/her information saved in the database. First thing they have to do is, give their proper user name and password to the system which is not possible for them. Moreover, as each cancer patient has to maintain a lot of treatment reports for a longer time period. So there is a big chance to lose some of the documents and reports. As a result, we implemented this QR code entry system where each user is given a unique QR code printed in a card and by scanning this card a patient can easily access his/her information from the system. As a result, there is no chance to lose any report.
- As pictures consume more space compared to any text document. So we search for a process by which we are able to extract only the data and values of the pictures and remove the chance of consuming extra space in the database. And AWS gives us the opportunity to implement our idea. We implemented OCR, Amazon's API for extracting the data. Actually, OCR helps us to extract data to a text document and save only the data or values displayed in the picture.
- Because of the implementation of OCR we are able to add the search ability feature in our system. Normally it is not possible to search any particular patient's data like hemoglobin from a lot of reports. But as OCR converts the picture to a text and we save only the values in the database. So now doctors can easily search for specific values of a particular patient which is also very helpful and time efficient for a doctor.
- In our system we ensure cost effectiveness from two perspectives. First one is, in terms of test reports when a patient tests his blood report, his report value remains valid for some given days . And between those days if he again needs the blood test report then he can use the saved report from the database. As a result, there is no need to do the same test again and again. And another cost effectiveness is shown in detail in the later chapter.

- Till now, there is no such fully public system available in Bangladesh. Only some private hospitals have patient registry system but those are not available for everyone. Also, the existing public registry system contains only the basic information of the patients. And there are no such options like us to use this data in future. Moreover, in most of the cases a certain portion of any proposed system is shown whereas in our case we are able to have a fully proposed prototype with all the features integrated into it. And this is one of the major contributions of our research.

To conclude, these above features are the main contributions of our proposed prototype system. Though some of these already exist in different systems, our nobility is we are able to do the integration of all these in our system with different methods. As a result, we can say that we have a well developed prototype system.

## 1.6 Thesis Outline

Our research's main objective is to create a cancer registry system that is both affordable and easy to use for patients and doctors. Additionally, we have conducted surveys of both doctors and patients to learn more about their preferences and the features they need in the registry system we are developing. The chapters covered in this paper are listed below, along with an overview of each:

1. **Chapter 2:** The Related Works section contrasts, compares, and summarizes the gives insight into the body of knowledge that exists for a particular system or subject. The main focus of this section is a brief discussion of technically published material relates to the work we've proposed.
2. **Chapter 3:** The working method is discussed throughout the research process in the methodology. This section outlines the steps taken to conduct research on a problem and our justification for the particular methods and approaches used to locate, gather, and evaluate data that will aid the reader in comprehending the issue.
3. **Chapter 4:** The section on experimental evaluation is all about the research's findings, which also include fieldwork and a critical evaluation of the data.
4. **Chapter 5:** The system's proposed prototype, which displays the system's present and potential future features, is visualized by the design and implementation of the GUI part. Additionally, it illustrates how the system is currently being used.
5. **Chapter 6:** The discussion section explains the context of our cancer registry system, designs the necessary content, compares our work to previously published works, demonstrates how different or distinctive our research is. Also, it focuses on the limitations which are all about the challenges we face during the research.
6. **Chapter 7:** The paper's conclusion offers some long-term benefits of the research work as well as the paper's conclusion and future work indicates upcoming system improvements.

# Chapter 2

## Related Works

Cancer registration serves as the fundamental aspect for our understanding of the neoplastic disease burden in our communities like at the local, regional, and national levels. Data comparability is important for interpretation, which is dependent on uniformity of technique as well as the diagnostic and other criteria used in this regard [24]. These programs are assessing and adopting new processes which ensure the validity and completeness of cancer data by using information technology and the integration of hospital and community-based practices. Moreover, these data are using for multiple logistic regression [25]. Globally the number of cancer patients is increasing at an alarming rate which depends on many factors. A three-year survey shows that approximately, there were 10.9 million new cases found from which 6.7 million were deaths and around 24.6 million people were cancer survivors. This cancer variation differs from geographical area to area, and it's became a great challenge to prevent this risk factors because of people's unhealthy lifestyle or by the effect of surrounding environment [8]. In terms of developing countries like Bangladesh the situation is apprehensive and it is assumed that within 2050 the cancer diagnosis rate will become 50-60% in the lower income countries [32]. Moreover, India, Bangladesh, Nepal, and Srilanka bear one-third of the world's cervical cancer burden. However, public health authorities in these nations have made no meaningful efforts to implement population-based screening. Now at this time Cervical cancer is in a condition that receives little attention from professional or public health groups in terms of advocacy, screening, and prevention [15]. KCR, like all other cancer registries in the United States and Canada, employs a data dictionary given by the North American Association of Central Cancer Registries (NAACCR) for uniform data entry. They are building mappings from data dictionaries to ontology's using a web-based tool called Interactive Mapping Interface (IMI) [36]. The World Health Organization (WHO) predicts that among Bangladeshis ages 30 and older, there are 49,000 cases of oral 71,000 cases of pharynx and laryngeal and 196,000 cases of lung cancer [10]. Again, a national level survey was done in Bangladesh based on peoples' genetic factors of cancer prevention, medical history and their daily lifestyle. Men are more likely affected by lung cancer whereas women are mostly affected by breast cancer. This survey also shows that 68.14% people were suffered by cancer because of smoking and 67.55% followed by betel leaf [2]. For the same age group, cancers of the oral cavity, larynx, and lungs account for 3.6 percent of hospital admissions [10]. According to World Health Organization (WHO) research, the yearly cost of tobacco related diseases in Bangladesh is US\$ 500 million, while the overall annual

profit from the tobacco industry is US\$ 305 million in tax income [21]. Moreover, the quality of the cancer registry data plays an inevitable role for the efficient usage of the data. In the thesis paper “Quality of Cancer Registry Data”, the author said to provide the quality of cancer care an adequate measure of the efficiency with the available information in the registry is a requirement. Some mentionable methodologies have been introduced here to ensure the quality of data [30]. Different medical record reviews provided these data. Using these data, treatment for colon and rectal tumors was done effectively [7]. For example, searching within the databases, following selection criteria confirming the validation of cancer registry data and the proper usage of rapid quality reporting system (RQRS) [25]. The incidence of cancer differs by age group. According to the NICRH and BSMMU cancer registries, 60% of male cancers and 5% of female cancers are caused by smoking and thus completely avoidable [14]. A hospital-based study found that cervical cancer accounted for roughly one-quarter of all female cancers [2], [4]. Incidence and mortality rates of cervical cancer were 27.6 and 14.8, respectively, per 100,000 women in 2002 [8]. Bangladesh has 13,000 new cases and 6,600 deaths from cervical cancer each year [12]. Moreover, cancer is a fatal disease that can affect any part of the body and is caused by uncontrollable cell growths. It has been reported that the 5-year relative survival rate of breast cancer is 76-82 percent; cervical cancer is 63-79 percent; and colon cancer is 44-60 percent in China, the Republic of Korea, and Singapore. Thai numbers were also lower compared to those of India and Pakistan [2]. It is very obvious that the lack of technical, human, and financial resources is a key hurdle to the implementation of successful preventative and control programs. Accurate data is required to manage the cancer epidemic. “Monitoring and surveillance of chronic noncommunicable diseases: progress and capacity in high-burden countries” cites lack of data comparability as a major issue [15]. Thereby, an efficient national cancer registry system is a great asset for not only assisting the country but also understanding the global cancer situation. A fully digital cancer registry system eases the cost of accumulating patient data [32]. Again, the quality of data determines the registry’s efficiency. “The Danish Cancer Registry” states that the quality of cancer registry data is extremely crucial to the registry’s effectiveness. Actually, cancer registries help establish and maintain cancer incidence reporting systems, provide data for cancer research and management also help to plan and evaluate cancer management strategies [16].

Application programming interface is referred to as API. By allowing automation, APIs have significantly influenced the way that business is done today. In essence, it is a set of guidelines for how two computers should communicate with one another. Application developers may create, connect, and integrate apps fast and at scale thanks to an API’s established communication protocol. API acts like the intermediary between the end user application and the web server. In general, a collection of data will be requested via an API call by the application we utilize in order to present it for the end user to view. The user interface is filled with the desired data when the request is made using the API, which connects to the web server to get it. Actually, API take care of all the obvious abstractions seen in web applications at every level. Any online application may obtain the data it needs to show without having to understand how the web server functions, all it has to know is how to utilize the APIs [43].

When it comes to the large APIs, it can be challenging to understand, which can

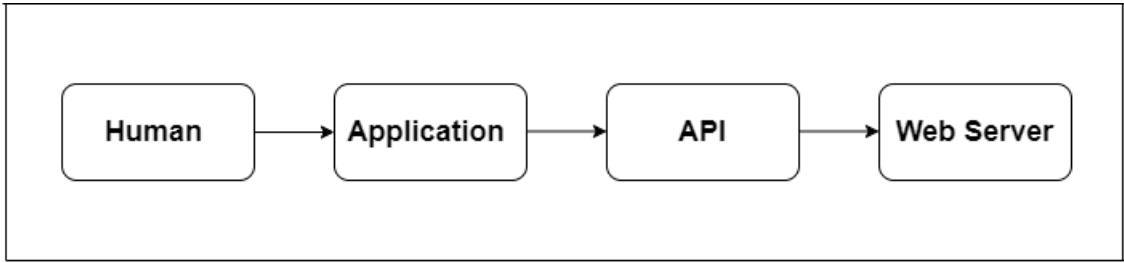


Figure 2.1: How API works.

reduce programmers' productivity. Over 440 experienced developers participated in a study that gathered their thoughts and experiences using a mix of surveys and in-person interviews. According to the report, the documentation and other learning materials were some of the biggest challenges experienced by developers while learning new APIs [17]. Not only that but also application programming interfaces (APIs) are a must for generating current software. APIs can occasionally be used by software developers by simply copying and pasting code samples. But since a lack of thorough understanding of an API's operation can result in errors, make software maintenance more difficult, and restrict what can be expressed via an API [37].

Web data extraction is a significant issue that has been researched using a variety of scientific approaches and in several applications. In the area of information extraction, several strategies and algorithms with a high degree of reuse were created. In reality, there are now two primary groups of Web Data Extraction apps: applications for the enterprise and applications for the social web. Web data extraction techniques become a crucial tool for business process re-engineering and data analysis in business and competitive intelligence systems at the enterprise level. Again, in terms of social web level side, data extraction methods enable gathering a significant quantity of structured data that is continually produced and shared by Web 2.0, social media, and online social network users at the level of the social web [22]. The rise of high-powered desktop computer has allowed the development of increasingly sophisticated recognition software. Actually, this reading technology can read a wide range of typical printed typefaces and handwritten texts. Though it is beneficial for data extraction, it remains a difficult challenge to develop an OCR that works in all possible conditions and produces highly accurate results [13]. In OCR a collection of feature extraction and classification algorithms play a critical role in character recognition. Here, feature extraction gives the way for identifying characters uniquely and also provide us a high degree of accuracy. Not only that but also it assists us in determining the form included in the pattern [20]. It also makes a lot of difficult tasks easier to us. Moreover, the accuracy of OCR is dependent on text pre-processing and segmentation techniques. In this process the image quality is one of the most critical variables in improving recognition quality in OCR systems. Images in different formats like .png, .jpg, .jpeg and .gif files can be handled separately or as part of multi-page PDF documents [26]. Again, there are different model present that can display diverse file formats like text files, websites, or spreadsheets and combines different types of documents. Additionally, this varieties of models make it simple to arrange data in a hierarchical fashion, but it is challenging to retrieve raw data for any additional processing or querying. Because

as these models retrieve data so sometimes the actual data get changed and as a result it became very difficult to use this extracted data or values in future [23]. Developing substantial proof policies and interventions requires to access cancer registration data. To ensure the long-term validity of the cancer registry, resource requirements must be quantified and system deficits identified. A web-based cost assessment tool is used to analyze three years of data from 40 National Program of Cancer Registry grantees. Registers were classified as low, medium, or high volume depending on the number of cases they contained. When reporting a case, low-volume cancer registries spent \$93.11 on average, whereas high-volume cancer registries spent \$27.70. According to the study, the most expensive per-care costs were for data collection and abstraction (\$8.33), management (\$6.86), and administration (\$4.99). Low and medium type registries are more expensive than high-volume registries for all critical activities because they have a lower volume of transactions. It is possible that the initial quality of data submitted to registries by reporting sources such as hospitals and pathology laboratories plays a role in the volumetric cost differences observed between the two groups. Increased data collection efficiency, as well as increased automation, may help to reduce overall costs [29].

This part of the report will empathize with the previous remarkable work that we found from different journals, papers and the internet. A cancer registry system has to be efficient as a conservatory of information and an easy data retrieval process is a vital part of the registry system. For achieving the efficiency of the system some techniques and methods have guided us towards the achievement of this goal. The assistance of some previously done journals and papers are undeniable for improving the quality of the proposed system. Below there is a table containing information about different cancer registry systems of different countries which is retrieved from “Astha: Handy and Manageable Cancer Management System for Bangladesh” [32].

Furthermore, there is an existing system named OWise - Breast Cancer Support which is an app to store the diary entries of patients. This app stores treatment notes, voice recordings, photos and recordings in a single place from where they can send those notes to doctors at any time when required throughout their treatment process so that they can get personalized treatment [52]. Moreover, for the study of cancer there is a system named Central Cancer Registry (CCR) which is only for the use of the oncologists. In this system the patients are recorded and only the registered doctors get access to the system by using their username and password where they can create and edit patient profile, create prescriptions for patients and see their treatment history and determine the cancer stage by using TNM staging system which are of two types cType and pType [39]. Again, Cancer, Cancer Control and Bangladesh is another system which is only used for statistical analysis of cancer patients [18]. The system Global Cancer Registry Software is notable for its functionalities of patient care, medical research and clinical studies and for public access to the database [42]. Lastly, Kentucky Cancer Registry is a data entry system which is featured with user friendly and dynamic UI. In this system there are extensive validations and edit checks for single fields, between fields, and between records during data entry and has the capability of storing unfinished abstractions for future completion [50].

A cancer registry system is mainly known for storing data efficiently. According

to the article “Cancer Control in Bangladesh”, the following steps are taken in order to create a cancer registry: Data collection and construction of a cancer atlas for Bangladesh an oncology/radiation therapy department-based cancer registry at all government medical college hospitals NICRH and BSMMU’s Oncology Department will continue to maintain a hospital-based cancer registry. Gazipur has a population-based cancer database. To begin a cancer registry in North Motlab, Chandpur and Mir sari in Chittagong province, Bangladesh. All oncology and radiation departments of the government medical college hospital are starting a population-based cancer registry [21]. Additionally, the accuracy and usefulness of cancer registry data are directly related to their quality. The IKL (Integraal Kankercentrum Limburg) cancer registry conducted a study to assess the data’s quality by comparing clinician-supplied data to registration data. Twenty clinicians re-analyzed data from ten cancer patients diagnosed in 1989 or 1990. Following data coding, it was compared to cancer registry records. There were 190 cases available for comparison. The most frequently encountered major disagreements concerned the following: date of birth, gender, date of incidence, primary site, laterality, histologic type, and behavior code. In general, the clinician’s imprecision in comparison to registration personnel may account for the discrepancies (primary site, laterality). This study established that registration personnel are capable of collecting accurate data [25].

In addition, a cancer management system had been proposed in a conference paper named “Astha: Handy and Manageable Cancer Management System for Bangladesh”. The proposed cancer registry system is a remarkably wonderful hospital-based cancer registry system ‘Astha’ which includes a unique barcode system in smart cards for the cancer patients of Bangladesh through RFID (Radio Frequency Identification) and BC-database. In this system a central cancer database is used in a distributed cloud system which is for the upload and retrieval of cancer information of patients [32]. Moreover, the quality of the cancer registry data plays an inevitable role for the efficient usage of the data. According to the thesis paper “Quality of Cancer Registry Data”, to provide the quality of cancer care an adequate measure of the efficiency with the available information in the registry is a requirement. Some mentionable methodologies have been introduced here to ensure the quality of data. For example, searching within the databases, following selection criteria confirming the validation of cancer registry data and the proper usage of rapid quality reporting system (RQRS) [25].

Again, a lot of surveys are going on at present time. These types of works play an important role in measuring and figuring out the overall situation. In this regard Karachi’s cancer survey is one of them. The aim of this study is to estimate the cancer incidence in Karachi Division from 2017 to 2019. From 2017 to 2019, Karachi Division. With 16.1 million residents in 2017. Data collection began in 2017. The ages (0–14, 15–19, and 20) are included. 16.1 million people in 2017. Every age had ASIRs. KCR had 33,309 malignant cases from 2017 to 2019. 17:3,2216.7; 11:2,233.6. Adults develop breast, mouth, and ovarian cancer (10.89). Cancer of the mouth is the most common cancer in Karachi. Children get bone, leukemia, and brain cancer. Successful programs rely on reliable cancer incidence and risk factor data (CCP). They need ”Cancer Registries” to work. For most cancer patients in Pakistan, pal-

liation and pain relief are their only options. An ethnically and culturally diverse country like Pakistan, the KCR is vital [35].

Additionally, the journal “Taiwan’s Nationwide Cancer Registry System of 40 years: Past, present, and future” states that, Taiwan Society of Cancer Registry is developing and promoting data standards for cancer registration, also auditing the databases for improving the accuracy of data [33]. All cancer cases are reported to the Kentucky Cancer Registry (KCR) within 6 months of diagnosis. Registries in the US and Canada use the North American Association of Central Cancer Registries (NAACCR) data dictionary. The NAACCR Data Dictionary will be linked to the National Cancer Institute Thesaurus (NCIT). Use the Interactive Mapping Interface (IMI) to map NAACCR to NCIT. There’s an ontology library, a mapping interface, and a suggestion It has mapping ontologies. They described IMI’s look and feel. The IMI prototype was used to map NAACCR data elements to NCIT concepts [36]. 272 NCIT concepts per 301 NAACCR data elements. This mapped tree has five branches. Intuitive mapping of data dictionaries to ontologies via the web. Basically, it can map NAACCR data to NCIT concepts [36]. In conclusion, without increasing awareness regarding chronic diseases like cancer, Global leaders’ tangible steps are insufficient to develop and implement cost effective solutions [15].

# Chapter 3

## Methodology

In this section, we will explain the methodologies we will follow to develop the proposed cancer registry system and to conduct the survey regarding our research to understand the research problems and analyse the survey results for the development of the cancer registry system. The methodology is divided into two parts where the first part contains the distinctive procedures and approaches of the proposed cancer registry system and the another part explains the sequence of steps that would be taken to design the survey questionnaires and conduct the survey sessions. Hence the techniques and the approaches are disputed in the following segments.

### 3.1 Conducting the Survey

In this part of the methodology we present a series of approaches and steps for conducting the survey result so that it can serve our research purpose. We planned to do surveys on doctors and patients distinctively for which we have to design two sets of questions each dedicated for the appropriate target audience. It is necessary to keep in mind that doctors and patients are two groups of people having differences in various cases. So, we will maintain the differentiable in the questionnaires to identify the research problem and relations between different facts.

#### 3.1.1 Survey Questionnaire Design

For performing the survey, designing the survey questions is the prerequisite where it is necessary to highlight and keep in mind the research problems, hypothetical solutions and the participants who are relevant to the research. We have divided the target audience or the participants into two categories which are the patients and the doctors. For this reason it is necessary to create two different questionnaires which will state the relevance and co-relation between different facts.

When designing questionnaires it is necessary to follow a standard principle or specification so that the structure and formation remain accurate and unambiguous. The questionnaire should be in an easy and familiar language so that we can grab the attention of the participants and be done with the interview. Additionally, the courtesy of asking one question at a time should be maintained [28]. Moreover, the

purpose is needed to be clear and maintained while designing the questions which lead us to create both open ended and close ended. And we have used a six point standard when designing the questionnaire so that the survey can give a clear idea of the participant's opinion [19].

Firstly, we have designed a questionnaire for the patients by maintaining standard guidelines and principles so that the set of questions can cover our purposes of the survey and get standard outcome for the research studies. We have decided to ask the questions in Bengali language as this is the native language. The design of the questionnaires have been done in English and after that we have translated the questions in Bengali. In fact, the questions will be asked in Bengali and the data collection part will be done in Bengali and then it will be again converted in English for the analysis. The development of the questionnaire has been done in two parts: one is designing questions regarding demographic information and another is the questions regarding research problems. The developed questionnaire is the set of 23 questions where 10 questions cover general information which is also considered as demographic information and the rest of the 13 questions are designed to fulfil the purpose of research studies and understand the problems and relations between them. These 13 questions will help us to get the hypothetical scenarios of the problems. There are 7 Like-rt questions and 2 quantitative questions to get the overall opinions and the appropriate response. Furthermore, the survey session or the interview will be done in-person by going to different hospitals to get the target audiences.

Like the design process of the questionnaire for patients we have also designed a set of questions for doctors which we will ask during the interviews. Though the process of designing the questions is the same, there are slight changes in the question to get the overall idea of the scenarios, correlations between different facts and state hypothesis. We have designed a total of 23 questions for the survey. We have formatted 7 general questions covering demographic information and 16 questions for understanding the problems relevant to the research. Here also we have designed Like-rt questions and quantitative questions. 3 quantitative questions and 9 Like-rt questions have been created. Like the survey for patients we will go to different hospitals and ask doctors to spare some of their valuable time to conduct the survey session.

Lastly, when the formation of the questions are done we will conduct a mock survey session where we will check if the question we have generated can serve our purpose. For doing the mock survey or the mock interview we have strategies to create two different google forms. It is necessary to conduct a mock survey because it is essential to identify the questions that are not appropriate and do not serve our purpose. And once the main survey from patients and doctors are done then we can not change the results and the questions were not appropriate then our research objectives will not be fulfilled. Therefore it is necessary to conduct a mock survey session to identify the problems regarding the survey questionnaires.

### **3.1.2 Approaching the Target Audiences**

The first step in having a good conversation with both the doctor and the patient would be to create separate questionnaires that would be helpful in gathering important pieces of information from them. This can be achieved by scheduling and arranging a meeting with the team supervisors and team members to discuss the questions to be asked, the manner in which they should be asked, and the types of responses that should be taken into account. Likert scale questions, simple yes-or-no questions, own opinions, etc. are a few examples. We can start speaking with the patient and the doctor once the questions are prepared. Whatever the case, it is imperative that we have to avoid upsetting or offending them in any way. In order to understand them better, informal conversations must take place in Bengali. Additionally, it would be a great strategy if we could convey to them a basic understanding of the goal of our research. Last but not least, it would be ideal if the conversation started formally and ended casually. The interviewer and participant would feel more at ease doing so, increasing the likelihood that the patient and doctor would agree to complete the survey provided survey form. While conducting the interview sessions to the patients it is necessary to keep in mind and always assume that they are not in the situation of giving an interview to us. Because they are already sick and when a person is not well in health then he is also not in the mind. So, we can not have the interview session in a desperate way which will disturb them in their sickness. We can only ask them to engage in the interview session in an appealing way so that they can understand that the interview session is for their and other patient's well being. When having an interview session an alternate way is not to ask all the questions if we see that the patient is not feeling comfortable. In that scenario we will only ask the questions regarding research problems and studies. The questions regarding the demographic data can be skipped for some of the patients. And we can not ask for an interview with a patient who is severely ill. Above all we have followed a storytelling manner to conduct the interview session so that the patients engage in and give the data we need for our research studies.

The interview session for the doctors can be lengthy as they are mostly busy professional people serving many patients. So, it is necessary to remember not all doctors will be available for the interview. But we have to convince the doctors in a way that their interview will not only help the doctors in future for their treating purpose but will help many patients. We can not continue requesting for an interview as it will disturb them from their work. As there is no other way to find doctors other than hospitals, first we have to convince the hospital authority to conduct an interview session with the doctors. We can also ask them when the visiting time of the doctors will be over as we can not interrupt during their visits to patients. So, it is very much convenient to ask the doctors for an interview session when their visiting time is over. And it is necessary to get an appointment from a doctor other than their visiting time. Lastly, we have to ask the questions in a manner that the doctors feel comfortable to share valuable information and opinion to our questionnaire.

### **3.1.3 Analyzing the Survey Result**

In this section we talk about the outputs which we will get from our survey. After completing the interview with patients and doctors we will be able to collect a variety of answers from questions prepared for our survey. As we conduct our survey by using google forms as a result we also get the responses on those forms. Moreover, because of using google forms we will take advantage of it. In the return responses of google forms we will also get a variety of summarized diagrams based on the different questions and their corresponding answers. In our survey we set a different set of related questions so we can hope that we will get a mixture of diagrams from it. Normally, we expect the varieties of graphs and diagrams we get. Like, bar graph, pie chart. So, depending on these charts and graphs we will be able to collect a lot of data to analyze the cases and also be able to reach a conclusion. In this way we will be able to design a prototype of the cancer registry system by maintaining the requirements of consumers based on the analysis of the research studies and survey results.

## **3.2 Designing the Prototype**

The proposed cancer registry system functions according to some distinctive methodologies. There are primarily two databases available in this entire system. One of them is a central database, designated database 1, which contains all patient data from all hospitals. Another database is database 2, which is for the database of a specific hospital and contains information about only that hospital's patients. Each hospital collects and submits patient information to the NICRH(National Institute of Cancer Research and Hospital). NICRH has access and the ability to update all data stored in the central database. As a result, it updates these records in the database. When a patient visits a hospital and meets with a doctor for a checkup, the doctor obtains the patient's authorized card. The hospital issues this card to each cancer patient. If a doctor scans this card and it is verified successfully, the card retrieves data from the hospital's database and displays all of the patient's medical information. This time, database 2 of the hospital collects and retrieves data from the central database. All of these are processes that occur between a doctor and a patient. Additionally, we have a website. When a researcher wishes to use cancer statistics data, he or she must first register and verify their identity. As a result, s/he gains access to all the necessary information and documents. The website then retrieves data from the cloud, which is connected to the central database for storage and retrieval. Additionally, this website informs the patient about his or her current treatment and other pertinent information.

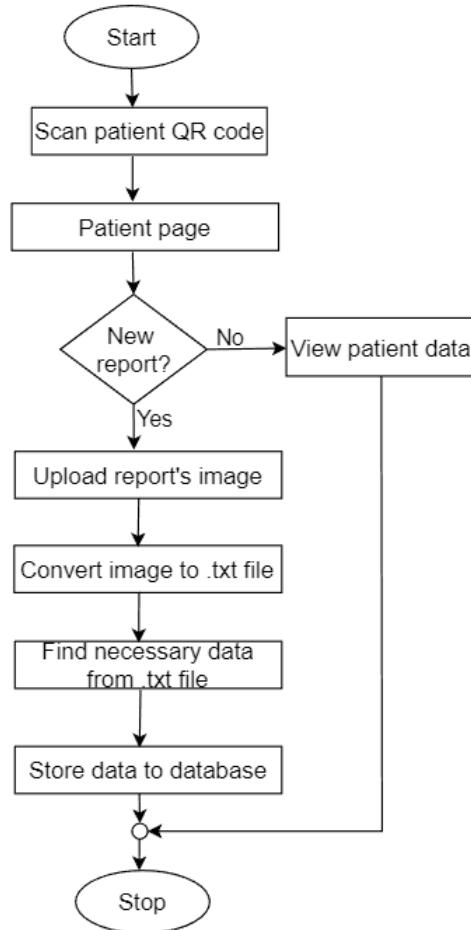


Figure 3.1: Flowchart of methodology

### 3.2.1 Scan QR Code

Every registered patient has an ID with a unique QR code and ID number. In the proposed Cancer Registry System there is a patient page for patient's information which shows all the diagnostic and treatment records of the patient. This page is accessible from the doctor's page and register page. By clicking on the scan QR code option the camera will be functional and capture QR code. Then the QR code will be scanned and the corresponding patients page will appear.

### 3.2.2 Patient's Page

This page shows all information and records of the patient. This page is only accessible by scanning the QR code of the patient's ID card from the doctor's or register page. There are several panels for showing distinctive data and records of patients. For adding new data to this page there is an option to upload images. By clicking that camera will be functional and click the image of that report for extracting necessary information and store them in the database.

### 3.2.3 Convert Image to Text

Text extraction is simpler and more effective if the image is prepared beforehand. Various software applications employ various pre-processing techniques. It is crucial to be able to distinguish text from background pixels when performing OCR. Due to the inaccuracy of the scanning procedure, some characters may have been twisted or inverted. Noise can obscure characters regardless of whether the image is binary or not. For each line of text within the picture, OCR attempts to construct a baseline first. For every single character line is dealt individually. When it comes to programming, comparing tokens and pixels is not a simple task. Using OCR method, once token matching has been completed, the results can be obtained. To prevent absurd outcomes, more deception is required. Each word is compared to a predetermined lexicon, and the closest match is substituted for the originals. OCR approach can be used to identify formulas and jargon from a specific field. It is comparable to the predictive text feature on smartphones. This can lead to writing that is simple to comprehend.

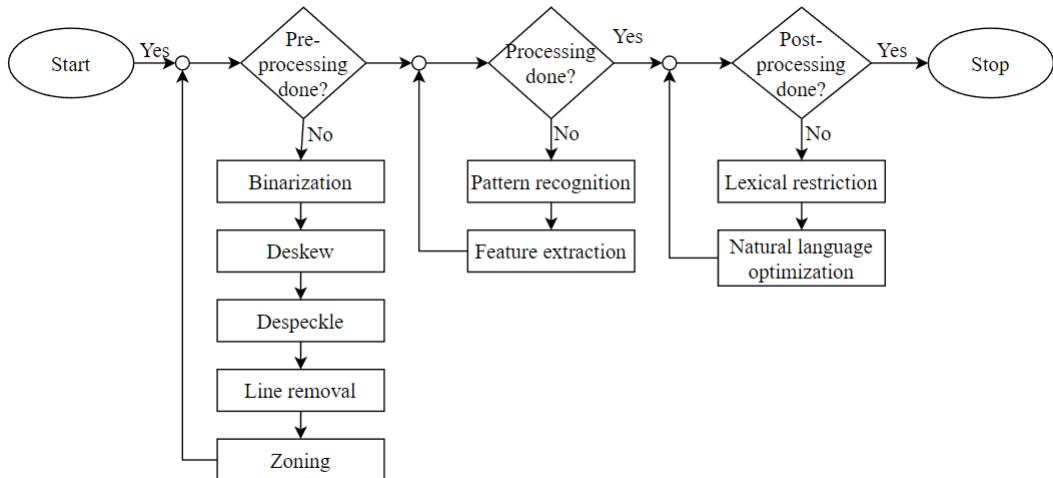


Figure 3.2: Flow Chart of OCR Process.

### Binarization

Image Binarization is the process of converting a document or image into a bi-level image. The number of pixels is divided into two colors which are black and white. An image binarization process is essential at separating text from the current frame [9]. It is a process that divides a picture into its component objects. This is a frequent job when attempting to extract an item from a picture. Therefore, as with so many other images analysis tasks, it is not simple and is totally dependent on the picture's information. In the early 2000s, binarization methods were introduced to preserve historical records in digital form [34]. Obtaining a threshold level in the histogram is just what makes the binarization process function. This threshold is a number that essentially splits the histogram into two segments, with each portion

reflecting one of two different items (or the object and the background). In this particular setting, it is referred to as global threshold [1].

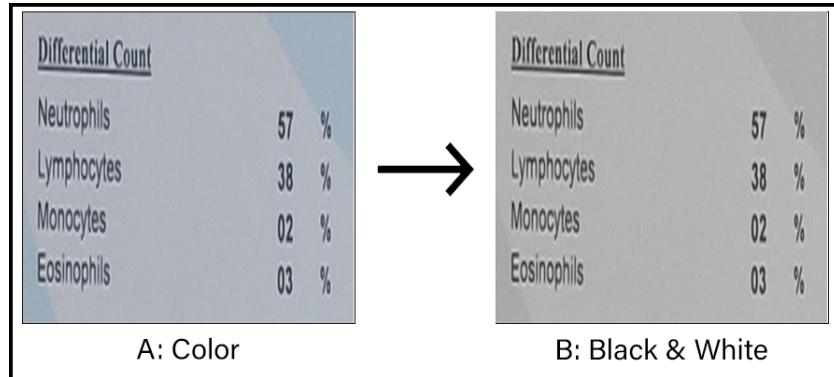


Figure 3.3: Binarization of an image.

In this process, a unique adaptive thresholding strategy is developed in order to binarize historical texts of poor quality and uncover important textual content. The goal of this research is to find meaningful textual information. The concept being recommended is separated into five primary stages. In the first stage, a low-pass Wiener filter is used for the purpose of conducting a denoising method. We make use of an adaptable version of the Wiener approach that is derived from statistics obtained from the immediate surrounding area of each pixel. In the second phase, we make a first preliminary assessment of the foreground areas by using Niblack's method. Since Niblack's approach often results in the introduction of additional noise, the foreground pixels are typically a subset of the Niblack output. In the third stage, we will calculate the background surface of the picture by overlaying nearby background intensities into the foreground regions that have been generated using Niblack's approach. A method that is conceptually similar has been suggested for binarizing sensor pictures [5]. In the fourth phase, we go on to the complete thresholding by mixing the computed background surface with the initial picture. This is done before we move on to the next stage. If a certain distance between the actual picture and the estimated backdrop is more than a certain threshold, then text regions will be identified. Textual information can be preserved even in very dark backdrop regions because of this threshold, which adjusts itself to the grayscale value of the surface behind it. A post-processing method is used in the very last stage in order to get rid of noisy pixels, enhance the quality of text areas, and keep stroke connection intact. The suggested approach was put to the test using a wide range of historical texts of poor quality, and the results shown that it is superior to the adaptive thresholding methods that are now considered condition [9].

## DE Skew

De skewing is the method of removing skew by rotating a picture in the opposite direction by the same amount as its skew. This produces a picture that is aligned

horizontal or vertical, with the text running across the page instead of at the aspect.

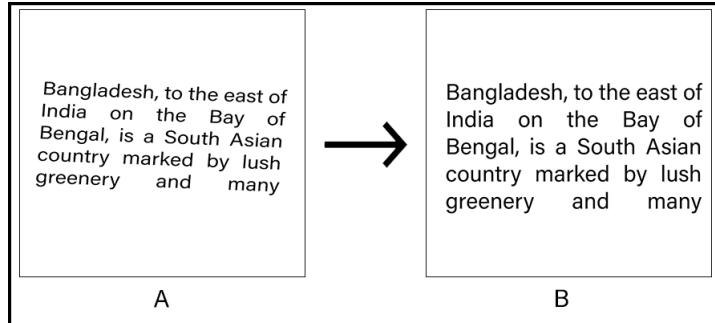


Figure 3.4: A shows before De Skew and B shows after De Skew.

When an object is misplaced, optical character recognition (OCR) becomes more complex, slower, and less precise. The OCR process can be accelerated and improved by de skewing the papers first [40].

### DE Speckle

The Despeckle filtration system eliminates visual noise without softening the edges. It aims to identify complex regions and preserve them while smoothing noise-affected regions.

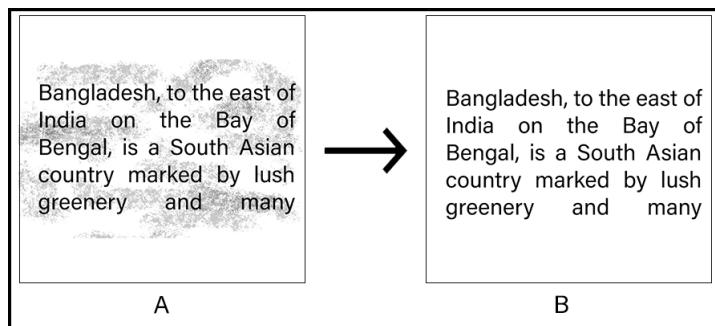


Figure 3.5: Effect of De Speckle.

The Despeckle filtering is used to smoothing out the places where noise is most visible while preserving the complexity of other parts of the image. The result is a reduction in particle and other types of distortion, without a significant impact on boundaries. Calculating the standard deviation of each pixel and its surrounding pixels allows for the determination of whether or not the region is one of high complexity or low complexity. In the event where the level of complexity is lower than the threshold, the region is smoothed down by employing a simple mean filtering [46].

## Line Removal

The Line Removal tool will get rid of any horizontal or vertical patterns that are present in the text and might possibly interfere with optical character recognition. It will function properly on black and white as well as color photographs, but it will perform more effectively on the black and white ones.

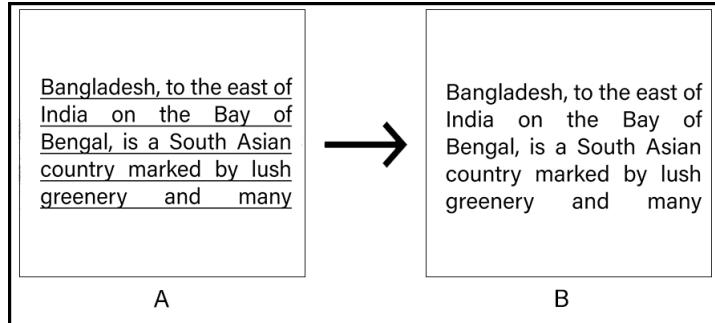


Figure 3.6: Before And after Line Removal.

When setting line removal, we have the option of specifying the direction of the line (or lines) that are to be deleted, as well as the minimum length, the maximum thickness, the maximum gap between the pixels, the region where extra pixels should be eliminated, and character reconstructions. Reconstructing a character fixes any issues with the character that came as a result of the line removal procedure. The method is carried out when the line has been removed.

## Zoning

In optical character recognition (OCR), "zoning" refers to the process of establishing zones that match to the certain characteristics of a page element. It is able to determine a zone using a non-text image, alphanumeric characters, or numerical values. OCR zoning often has the capability to zone either automatically or manually [49].

## Pattern Recognition

Pattern recognition manages to make the process of learning and detection of patterns visible, so that it may partly or totally be implemented on systems.[34] In order to get just text from the image, we processed the image in such a way as to remove all other elements. Image patterns will be saved in a different location for future reference. The picture being queried is then compared to image patterns that have been saved in the database.

If the picture is a match for the pattern that has been saved in the directory, the system will use an algorithm to identify the text, and it will produce any text it finds. The picture will be uploaded by the user. Because of the greater susceptibility of pictures to noise and other forms of environmental interference, the first step in the process is to eliminate noise by applying a filter. The application of image

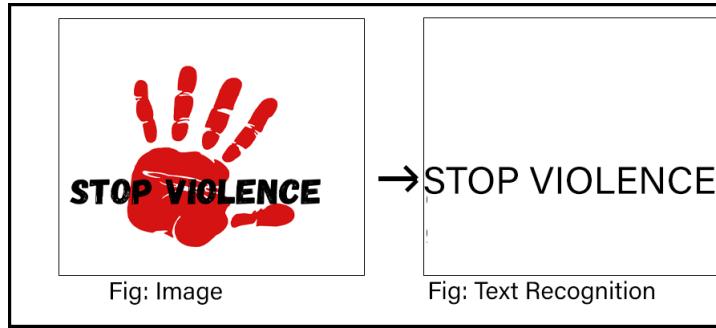


Figure 3.7: Pattern Recognition.

pre-processing stages takes place with images. Things which aren't needed are removed. The retrieved text is provided bounding boxes after being processed. The text is presented here in an image format. These pictures are turned into character representations. The system extracts characters from images via something called optical character recognition. The directory is where the character and number pictures are kept. The bounding box has been used to divide the retrieved text and picture. Each character and each number will be contained inside their own bounding boxes. Each letter, number, and character is compressed to fit the picture that is saved in the directory. A comparison is made between the extracted picture and the existing character image feature. Characters are identified once the comparison is made. Finally, the characters are shown in text format.

### Feature Extraction

A formal procedure can be used to extract the relevant shape information from a pattern, making the process of pattern classification much easier. Using feature extraction in pattern recognition and image processing could really help decrease the amount of dimensions in a set of data. For example, feature extraction is used to extract only the most relevant information from a data set and present it concisely. Representational features of input data that exceeds the algorithm's processing capacity and are suspected of being redundant will be reduced in size (also named features vector). The method of removing features from input data is known as feature extraction. If the features selected are carefully chosen, this reduced representation is expected to extract the relevant information from the input data in order to perform the desired task, rather than the full size. Tokens, for example, can be recognized for what they really are using character-descriptive rules. When two equal-height lines are connected by a single horizontal line, the result is a capital "H." Fonts and sizes are not restricted when using this method. Can distinguish between an I, L and 1 with ease using this software. Negatives? Programming is more complicated than just comparing bits and bytes [38].

### Lexical Restriction

All words are compared to a predetermined lexicon and the closest match is used to replace the originals. A dictionary is included in a lexicon. Typos like "thorn"

for "th0rn" can be corrected using this. An app's optimization using a specialized OCR may be necessary when dealing with medical or legal papers. To find formulas and jargon from a particular field, OCR software can be used [38].

### **Natural Language Optimization**

This cutting-edge technology uses a linguistic model to forecast the word order and then corrects the sentences accordingly. It's akin to the predictive text function on smartphones. This can result in writing that is easy to understand [38].

#### **3.2.4 Find Necessary Data From .txt File**

After the OCR process the image of the report is converted into a text file. Then necessary information is extracted from the text file. Information is searched by the corresponding keywords and when those words are found in the text file necessary data is extracted from the text file and stored in the database.

#### **3.2.5 Store Data to Database**

Two databases are used in this Cancer Registry System one for individual hospital information and another is used as a central database. When information is retrieved from the text file then they are saved in corresponding tables of the hospital database. In the patient table it includes personal information of each individual. For example, name, age, contact no, NID etc. Again, in the disease data table it contains all data regarding the patient's diagnosis report with its histology. Again, in the treatment history table it includes different treatment names like surgery, chemo-therapy etc. Additionally, for future use our system also saves some specific data from the reports of the patient such as HB%, RBC, Platelet, CEA etc. In this way initially our system will preserve all the necessary data to the local database of a hospital and after that a copy of this data reaches to the central database of NICRH for national record.

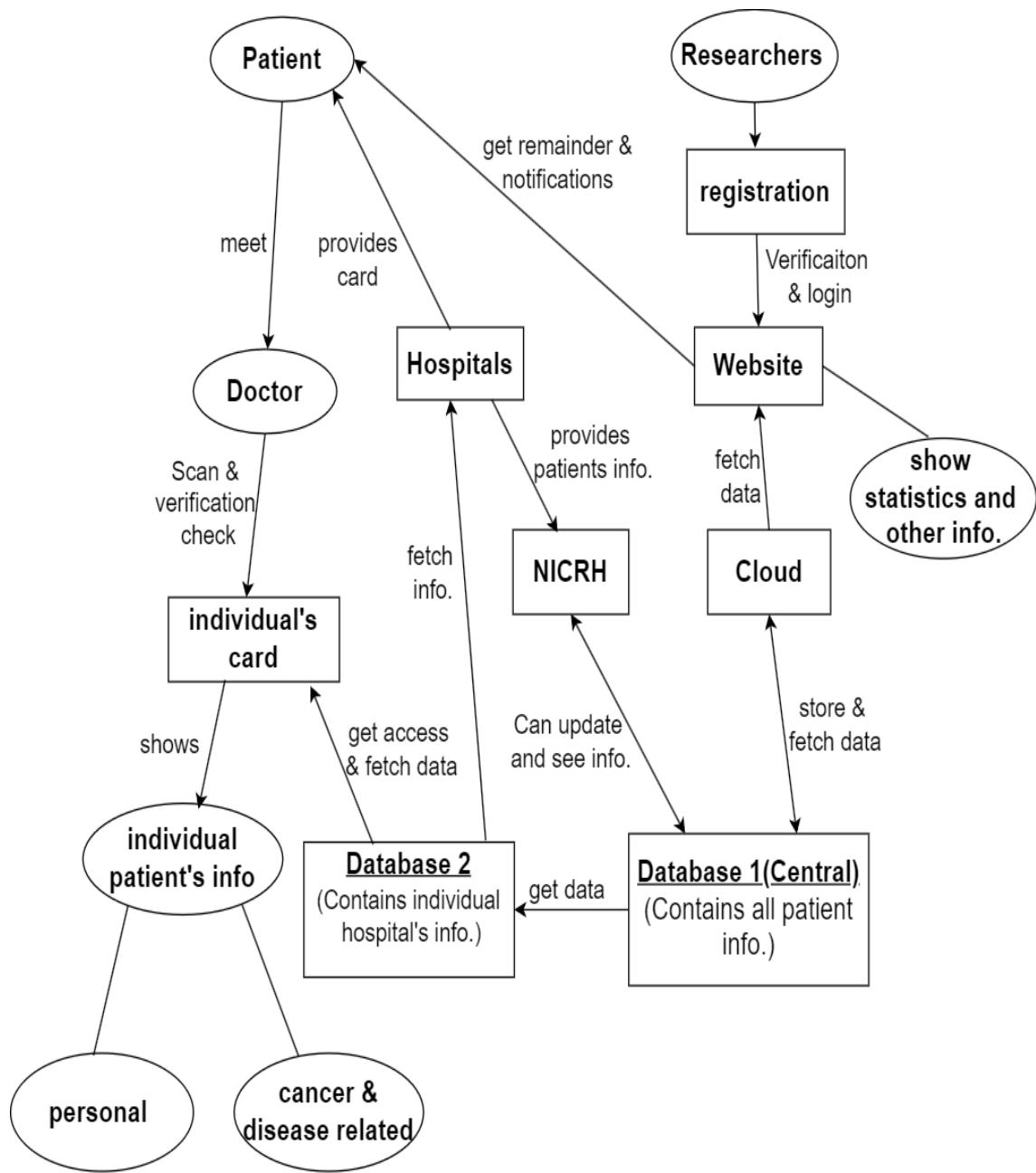


Figure 3.8: Flow chart of the proposed system

# Chapter 4

## Experimental Evaluation

### 4.1 Analysis On Survey Results

In this part we actually talk about our overall experiment and survey results. Moreover, how these values affect our overall performance. We know, for doing any research we have to go through a different type of analysis. One of these is quantitative data analysis. Actually, all types of diagrams (pie chart, bar graph, chi-squared test etc.) and different types of methods (percentage difference, standard deviation, precision, t-test etc.) from where we directly get the answer or value is called Quantitative data analysis. At the very beginning of our research, we prepared some standard questions for both patients and doctors. We prepare those questions in such a way so that the doctors and the patients can directly answer them. And based on these questions we do a survey over different patients and doctors. After getting the responses and based on the survey questions and data we actually do the quantitative data analysis. Moreover, the responses which we get from doctors and patients help us to develop our system gradually.

#### 4.1.1 Outcome From the Interviews With Patients

Among the end-users of our system patients play a significant role in the workflow of the cancer registry system. Because the whole process of the proposed system centers around the patients. Though they are not the primary user of our system but the purpose of the cancer registry system is to give the patients personalized care and store their treatment and diagnosis data in an efficient way to give them relief from the tension of storing and carrying them and to help the doctors to view their information's structurally to analyze them for treatment purpose. So, the opinion of the patients are significantly important for identifying the requirements of the system.

For the research studies and survey purpose we have interviewed a total of 79 patients. By interviewing the participants, we have found there is no patient who does not have a smart phone in their family. Among all the participants 88.6% patients use smartphones themselves and the other family members which is the number of 70 out of 69 patients. And 7.6% of patients shared that they do not use but other family members use smartphones. From these numbers we can understand that al-

most all the participants have a smartphone in their family.

Our finding also uncovers the frequency of the patients to get diagnosed. Among the 79 participants we have found that 38 patients get diagnosed once in a month which refers to the amount of 56.7%. Then, 13.4% of the participants get diagnosed multiple times in a month. Besides, 11.9% of the patients we have interviewed go to diagnosis centers for getting diagnosed multiple times in a week. Also, 10.4% does diagnosis once in a week and 7.5% of participants have other scenarios.

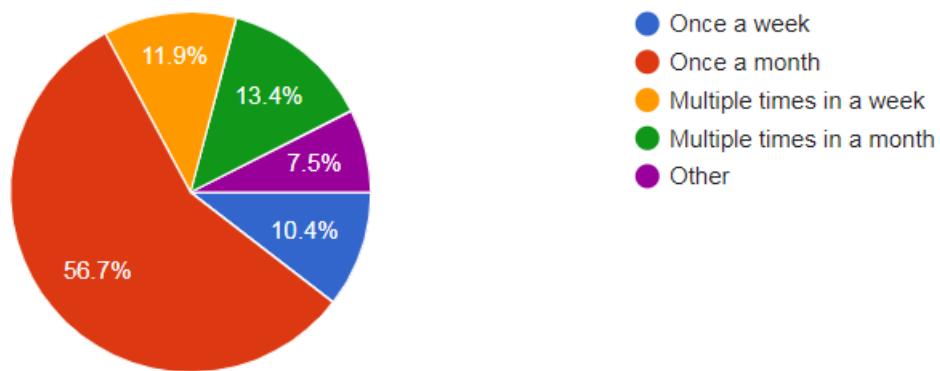


Figure 4.1: Frequency of patients for getting diagnosed.

Furthermore, we have also found the opinion on maintaining huge amounts of hard copies of diagnosis and treatment reports of the respondents of our interviews. It is found that 34.2% of people agree with the statement that they have to maintain huge amounts of hard copies of diagnosis and treatment reports which refers to the number of 27 out of 79 participants. Then, 26.6% people are neutral about the statement which is the number of 21 from all the respondents. Besides, 21.5% people revealed that they strongly agree with the statement and 16.5% people shared that they disagree with the statement.

Our survey also revealed the thoughts of the patients on the statement that more reports mean facing more difficulties to manage those from which we found that 35.4% (28 out of 79) of people agree with the statement. Then, 34.2% (27 out of 79) people strongly agree with the statement. Furthermore 22.8% people are neutral about the statement and 7.6% people disagree with the aforementioned statement which is a very small number compared to the other scenarios.

Then, we were interviewed over the question that if the patients change the doctors/hospitals, do they need to renew the medical test reports by doing the tests again and 35% people shared that they have to do tests again. Then, 28.7% of the participants said that they have to renew tests always and 26.3% said that they have to do the tests again. And the frequency of doing tests again is seldom and never is 5% for both.

It is also worth mentioning that 48.1% of people agree that a central Cancer Registry System would free them from the burden of carrying and maintaining hard copies of

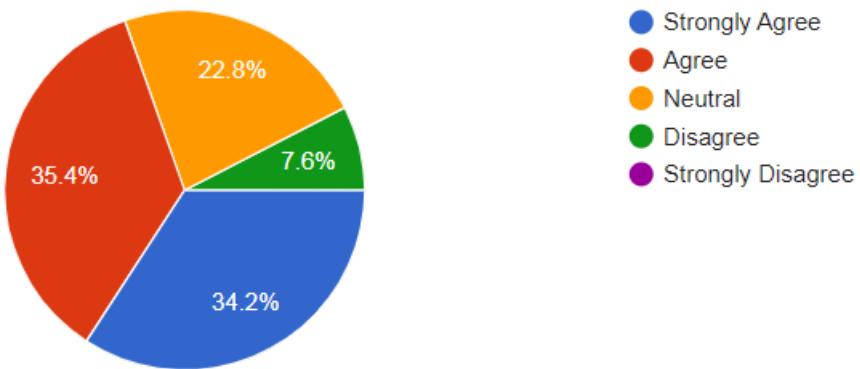


Figure 4.2: Opinion of patients on difficulty to manage reports.

test/diagnosis reports. Then 38% of people strongly agree with the statement and 13.9% of respondents are neutral about the statement. Then, 54.4% of people agree that there is a need for a central Cancer Registry System which can store diagnosis and treatment data. Besides, 31% of people strongly agree with the statement where 13.9% are neutral.

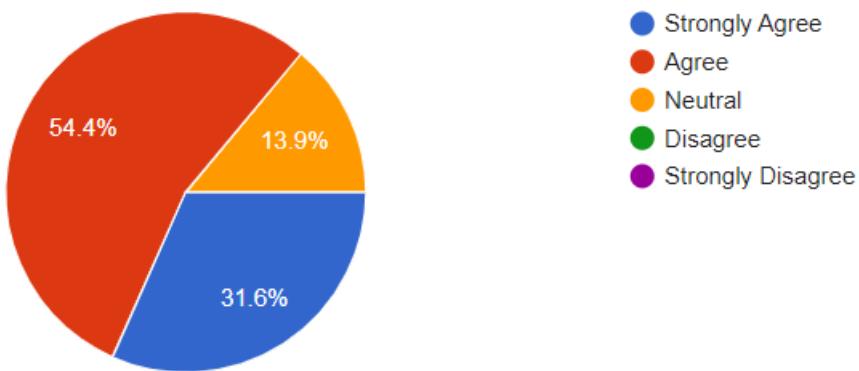


Figure 4.3: Opinion of patients for a central cancer registry system.

Then from our survey questionnaire to patients we found that 39.2% people will definitely use and carry the patient id card. 30.4% of the patients will use it and 25.3% will use it very probably where only 5.1% of people will probably not use or carry the patient id card. Then, 48.1% of the respondents agree to give their patient id card to the doctor for uploading their diagnosis and treatment report. 31.6% strongly agree and 17.7% of people are neutral about allowing a doctor to give their patient an id card which the doctor will use to upload the patient's diagnosis and treatment report to the system.

#### 4.1.2 Outcome From the Interviews With Doctors

The end-users of our system are patients and doctors where doctors are the primary users because the proposed cancer registry system will be significantly used by the doctors to view patient's diagnosis and treatment data and upload them when required. So, their opinion on the system is mostly important. A total of 25 doctors have taken part in the interview which we have done to understand their opinion about related factors for designing our system. We have found that 15 doctors out of 25 are the doctors of both public and hospital and 10 are only doctors of private hospitals.

From the respondents we have found that 66.7% of the doctors sit in less than 5 chambers and 3.7% of the doctors sit more than 4 chambers. Then, 84% of the doctors have an internet connection in their chambers where only 16% of the doctors do not have any internet connection in their chamber. Besides, 44% of the doctors shared that they have to maintain a very high number of patients where 36% of the patients have to maintain a high number of patients and 20% of the doctors said that they have to maintain a moderate number of patients.

Additionally, we have found the opinion of doctors that more reports mean more difficulties toward a patient to handle. 72% of the doctors strongly agreed with the statement and 20% of the doctors agreed where 8% of the doctors stayed neutral. Then, the interview also revealed that in 52% of the cases the patients have to renew the medical tests sometimes if they change hospitals or doctor. 28% of the doctors said often where 20% of the doctors said the patients have to renew the tests always. Then, we have asked doctors the question related to their experience on tech devices to understand their knowledge and efficiency on using them which is very much relevant for our research purpose. We have found that 45% of the doctors spend 5-6 hours of time on tech devices. And 64% of the doctors are somewhat aware about a registry or hospital management system where 20% of them are well aware. And 45.8% of the doctors have used a registry or hospital management system before. These numbers state that the doctors we have interviewed have experience on using tech devices or systems similar to our proposed cancer registry system.

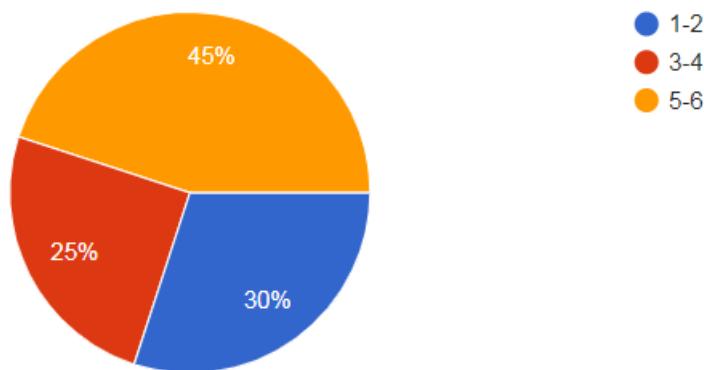


Figure 4.4: Time spent on tech devices.

It is very much important to know the opinion of doctors on the statement that a central Cancer Registry System would free patients from the burden of carrying and maintaining hard copies of test/diagnosis reports. Because, the doctors understand the actual situation of the patients and what will be better for them. 60% of the participants agreed with the statement and 32% of them strongly agreed with the statement. Therefore, we can understand the significance of a central cancer registry system which will not only free the patients from the tension of storing the reports and carrying them but also it will give the doctors a better understanding ability to the patient history.

In our proposed system patients will have a patient id card where a QR code will also be there to give easy access to the patient panel by scanning it. So, it is required to understand the experience of doctors of using a card-based registry system. 28% of the doctors are somewhat aware and 20% of the doctors said that they are well aware about a card-based registry system. Then 36% of the doctors said that they will definitely use a patient ID card with QR code to login without asking user id and password to the system.

Therefore, from the interviews with the doctors we can understand that doctors actually face difficulties when they are getting treatment for cancer. In that situation their mind stays full of grief and if they have to go through the complexities of maintaining a huge number of reports then the miserable situation gets worse. So, most of the doctors think that it is necessary to have a central cancer registry system which will reduce the tension of patients.

#### **4.1.3 Quantitative Data Analysis From Survey**

Lastly, we do Chi-Square Test and from the output values of the test we try to observe whether there is any significant relation present between different aspects of the survey questions or not. We already have the questions and its responses which we get through the survey, is used to do the test. For that, initially we have to choose some set of questions and their corresponding responses. Then we set those questions and responses in a separate spreadsheet to make the test simple. After that we check whether the values of the columns are dependent or independent to each other based on the following calculations. For,

$$\text{Expected data on a particular cell} = (\text{Row Total} \times \text{Column Total})/\text{Grand Total}$$

	Neutral	No idea at all	Some idea	Unaware	Full idea	Grand Total
Definitely	5	1	10	5	9	30
Possibly	8	3	3	8	2	24
Probably Not	1	1	1	2	0	5
Very Probably	0	0	8	8	4	20
Grand Total	14	5	22	23	15	79

Table 4.1: Observed Data for idea about card-based registry system VS opinion on using ID card for easier accessibility.

	Neutral	No idea at all	Some idea	Unaware	Full idea	Grand Total
Definitely	5.43	1.94	7.91	8.91	5.81	30
Possibly	4.2	1.5	6.9	6.9	4.5	24
Probably Not	0.88	0.31	1.44	1.44	0.94	5
Very Probably	3.5	1.25	5.75	5.75	3.75	20
Grand Total	14	5	22	23	15	79

Table 4.2: Expected Data for idea about card-based registry system VS opinion on using ID card for easier accessibility.

For the above calculation, responses of the columns values of patients' questions named as idea about card-based registry system and opinion on using ID card for easier accessibility are used. We match the attribute in terms of the same values and count the total no of same values for each unique attribute and then note in on the table of observed data. For example, the number of total patients' who have full idea about the registry system and want to use the id card definitely for easy accessibility is 9. And it measures in the scale of 79 people.

Now in the expected data table we put the values which we get from the observed data table by using the expected value measuring formula. Actually, expected values help us to identify whether the null hypothesis is true or false. Here, null hypothesis tells us that the values of our chosen questions column have no relation to each other. After finding all the cell values of the expected data table, we have to find out the values of the Chi-square value. For getting that first we subtract the expected values(E) of each cell value to the corresponding observed cell value(O). Then we get the new column containing the values of (O-E). Again, we make the square of the (O-E) column values and in the next phase we divide each (O-E) cell values by the corresponding E value. Finally, we do summation of all the values of new column and this summation value is the actual Chi-square value for the observed table. But to complete the test and reach to a conclusion we have to measure the degree of freedom our observed data table carry. By using the formula, we get the degree of freedom for observed data table is 12. Now by using significance and degree of freedom value we easily calculate the critical value of the table. Finally, by observing the critical and Chi-square value we can say that as the Chi-square value is greater than the critical value, so there is a relation exists and they dependent on each other. Also, it also rejects the null hypothesis.

Below some other relations between other columns are shown. As this whole process

Observed(O)	Expected(E)	O-E	$(O-E)^2$	$((O-E)^2)/E$
5	5.43	-0.43	0.18	0.03
8	4.2	3.8	14.44	3.44
1	0.88	0.12	0.01	0.02
0	3.5	-3.5	12.25	3.5
1	1.94	-0.94	0.88	0.46
3	1.5	1.5	2.25	1.5
1	0.31	0.69	0.48	1.54
0	1.25	-1.25	1.56	1.25
11	8.91	2.09	4.37	0.49
3	6.9	-3.9	15.21	2.2
1	1.44	-0.44	0.19	0.13
8	5.75	2.25	5.06	0.88
5	8.91	-3.91	15.29	1.72
8	6.9	1.1	1.21	0.18
2	1.44	0.56	0.31	0.22
8	5.75	2.25	5.06	0.88
9	5.81	3.19	10.18	1.75
2	4.5	-2.5	6.25	1.39
0	0.94	-0.94	0.88	0.94
4	3.75	0.25	0.06	0.02
			Chi-square, $\chi^2$	22.53

Table 4.3: Chi-square calculation

Significance	Degree of Freedom	Critical value	Chi square value
0.05	12	21.03	22.53

Table 4.4: Critical value calculation for Idea about card-based registry system VS opinion on using ID card for easier accessibility.

is followed to observe the other columns relation so we are not showing these again.

Here, we try to figure out the relationship between opinion on the difficulty

	Agree	Neutral	Strongly Agree	Grand Total
Agree	16	2	10	28
Disagree	3	0	3	6
Neutral	8	6	4	18
Strongly Agree	10	3	14	27
Grand Total	37	11	31	79

Table 4.5: Opinion on the difficulty of managing hard copies of report VS A system storing all diagnosis data is helpful.

	Agree	Neutral	Strongly Agree	Grand Total
Agree	13.13	3.89	10.98	28
Disagree	2.8	0.84	2.36	6
Neutral	8.42	2.51	7.07	18
Strongly Agree	12.65	3.76	10.59	27
Grand Total	37	11	31	79

Table 4.6: Opinion on the difficulty of managing hard copies of report VS A system storing all diagnosis data is helpful.

Significance	Degree of Freedom	Critical value	Chi square value
0.05	6	12.59	14.74

Table 4.7: Critical value calculation for Opinion on the difficulty of managing hard copies of report VS A system storing all diagnosis data is helpful.

of managing hard copies where a system storing all diagnosis data is helpful for patient. After finishing all the calculation as we find the Chi-square value is greater than the critical value so we again say that there is a relation or dependency exist between these two columns.

	Neutral	No idea at all	Some idea	Unaware	Full idea	Grand Total
Definitely	4	0	1	0	4	9
Possibly	2	1	3	2	0	8
Probably Not	1	0	0	0	0	1
Very Probably	2	0	3	1	1	7
Grand Total	9	1	7	3	5	25

Table 4.8: Observed Data Idea about card-based registry system VS opinion on using a patient's ID card for easier accessibility.

#### Expected Data for Table 4.3

	Neutral	No idea at all	Some idea	Unaware	Full idea	Grand Total
Definitely	3.24	0.36	2.52	1.08	1.8	9
Possibly	2.88	0.32	2.24	0.96	1.6	8
Probably Not	0.36	0.04	0.28	0.12	0.2	1
Very Probably	2.52	0.28	1.96	0.84	1.4	7
Grand Total	9	1	7	3	5	25

Table 4.9: Expected Data Idea about card-based registry system VS opinion on using a patient's ID card for easier accessibility.

In this time, we feel interested to observe the relationship between the idea about card-based registry system and opinion on using a patient's ID card for easier accessibility in terms of doctor. At this particular case as we get lower Chi-square value compared to the Critical value so we can easily say that these two columns have no relation to each other.

Significance	Degree of Freedom	Critical value	Chi-square value
0.05	8	15.51	12.78

Table 4.10: Critical value calculation for Idea about card-based registry system VS opinion on using a patient's ID card for easier accessibility.

## 4.2 Analysis on System Output

In our proposed system or prototype we used OCR (Optical Character Recognition) tool. As in our database, doctors can upload the pictures, so for doctors' easier accessibility and better search-ability we implemented the OCR. Actually, in OCR different types of image processing filters and techniques are used to make the picture in an exact order. Like binarization, de skew, de speckle, line removal, zoning, pattern recognition, feature extraction, lexical restriction, natural language optimization, natural language processing is followed. By using these techniques OCR actually converts the picture into a text document and from there we can extract any particular data and be able to save that in our database. In our analysis we get an experimental data-set based on our actual given data-set which we used to do the experiment.

### 4.2.1 Quantitative Analysis on Data-sets

Here, we basically use some of these analysis techniques to evaluate and measure the outcome of our experimental datasets. Below we describe the step-by-step process of our used method. As we check the accuracy of our datasets. So, we put a variety of inputs from our collected dataset into our system and from there because of using OCR we get a new dataset. As a consequence, at first, we feel interested in measuring the accuracy which we get through the data extraction process. Two types of techniques we used for measuring and comparing the datasets. Those are checking the accuracy based on closeness (percentage differences) and another one is the accuracy based on exactness (average accuracy).

$$\text{Percentage difference} = \frac{|Exact\ value - Approximate\ value|}{Exact\ value} \times 100\%$$

Percentage difference tells us how much difference the calculated value carried compared to the exact one. In terms of our case for measuring the percentage differences we have to find out the differences between two data sets. But the thing is each of our data set contains multiple columns which actually carries the values of Hemoglobin, ESR, WBC, Neutrophils, Lymphocytes, Monocytes, Eosinophils, Platelets, RBC, HCT/PCV, MCV, MCH, MCHC respectively. So, to identify the percentage difference first I have to determine the absolute differences in percentage between each of the samples like the input hemoglobin value and the output hemoglobin value which we get after doing the OCR. After doing this for all of the samples we calculate the average difference for each of the samples again. Finally, from all the average difference sample values, we calculate the percentage difference value. In our case the percentage difference of two data sets is 3.13% which means our measured data-set's values are almost  $(100-3.14) = 96.86\%$  close to the given or actual data-set. On the

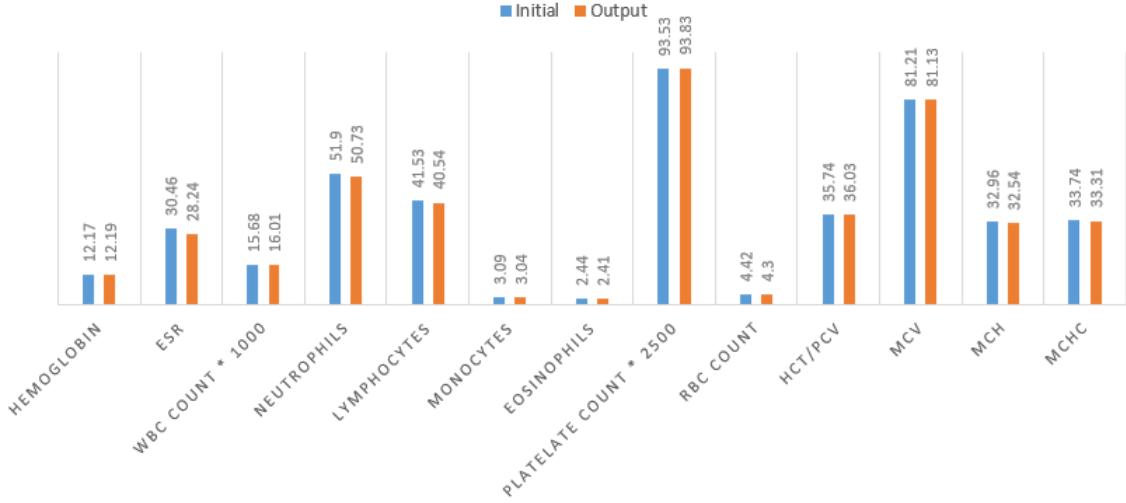


Figure 4.5: Average Difference between two data-sets.

contrary,

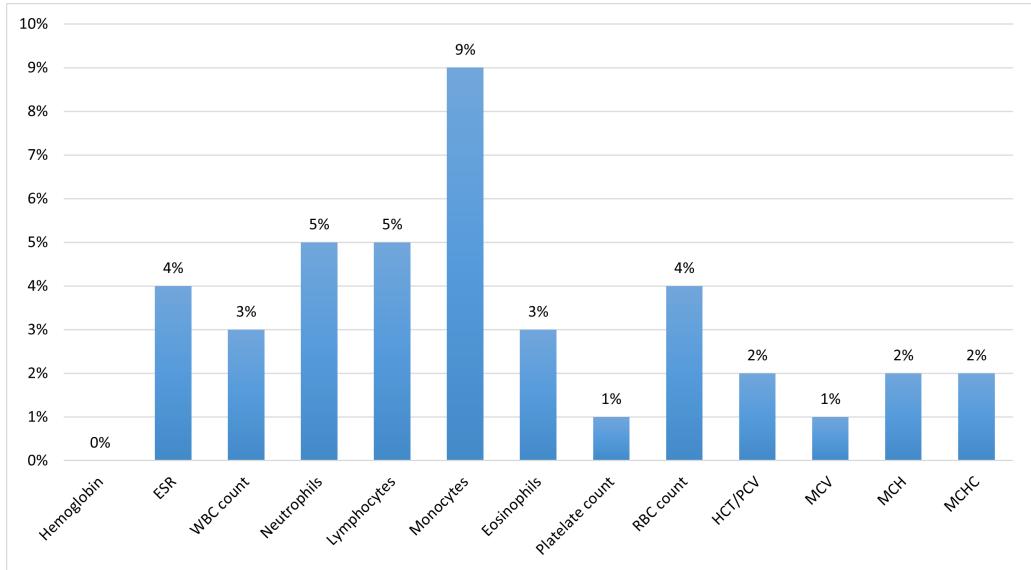


Figure 4.6: Closeness Between Two Datasets in Percentage.

$$Average\ Accuracy = \frac{No\ of\ total\ output\ matched\ samples}{No\ of\ total\ input\ samples} \times 100\%$$

On the other hand, average accuracy tells us how much actual data is exactly matched with the measured data/value. For finding the average accuracy we directly compare two data-sets, column wise one after another and find the total no of matching duplicate values for each column component like Hemoglobin and others. After that to get the accuracy value we divide the total no of duplicate values by the total no of values for each column and also take the percentage of it. Lastly, from the different average column values we calculate the final average value which is mainly the average accuracy of our proposed system. In terms of calculating average

accuracy, we find that two dataset's fully similar for 89.49% cases and (100-89.49) =10.51% values are different from each other.

In general, these two accuracy computing methods are fully different. Percentage differences indicate the closeness between two data set values whereas average error expresses how much exact values we get from our measured data set compared to the initial given data set. For example, I have two values 1 and 3. Here, 1 is the given value and 3 is the measures value. If I calculate the percentage difference then it will be  $(|3 - 1|/1) = 3\%$  difference. On the other hand, in terms of average accuracy it will be 100% different. As 1 and 3 are two different values and there are no similarities between these values.

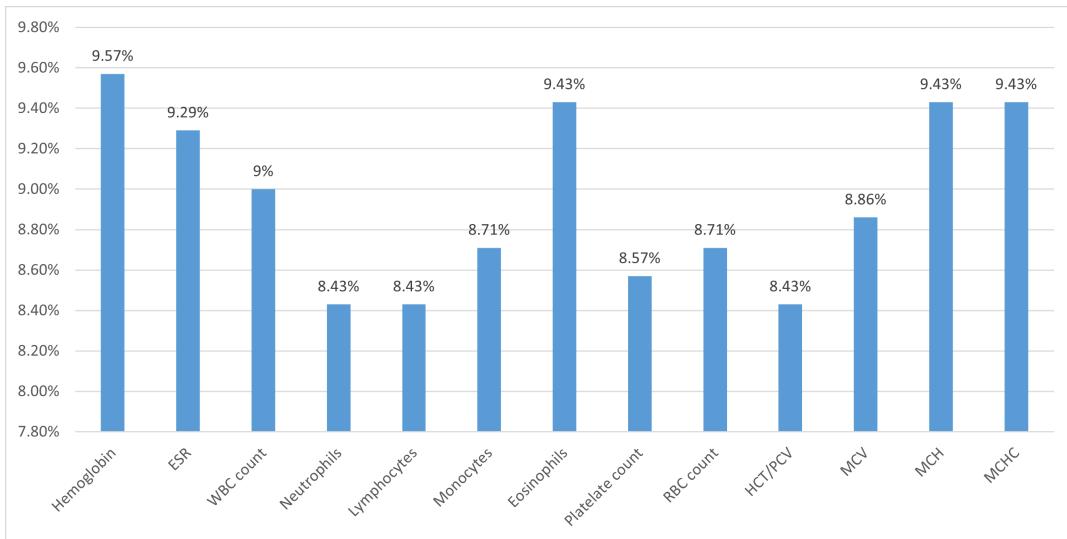


Figure 4.7: Exactness Between Two Datasets in Percentage.

$$\%RSD/Precision = \frac{\text{Standard Deviation}}{\text{Average}} \times 100\%$$

Secondly, we calculate the precision value from our measured data set. For that first we compute the average for each column. In the same way we calculate the standard deviation (STD) for each of the columns again. Standard deviation tells us how spreads our experimented value into the initial data set. After doing that we simply divide the standard deviation of all columns by the average value to calculate the relative standard deviation (RSD). And the computed values are also multiplied by 100. RSD is used in terms of comparing the mean of a data set and whether the data set is large or small. Lastly, we get the average of all relative standard deviations to get the final percentage relative standard deviation (%RSD).

Precision is the measure of how much detailed information given. Also, it indicates the degree to which exactness is applied. Again, it expresses the repeatability or consistency of values in a particular data set. Generally, %RSD is the measure of precision. In case of our dataset, we get a precision value  $39.4\% = 0.394$  or  $0.4 < 1$ , which is a good measurement. Because, precision value less than 1 is good for computation.

$$t = \frac{\bar{X}_a - \bar{X}_b}{\sqrt{\frac{S_a}{N_a} + \frac{S_b}{N_b}}}$$

Lastly, we calculate the T-test for measuring the statistical difference between the means of two groups. There are overall three types of T-tests. They are one sample T-test, two sample T-test and the last one is paired T-test. In this case, first we individually calculate the average of all columns in each of the data set. Then from there we calculated the difference of two mean values. Again, we compute the values of  $((Standarddeviation)^2)/Total$  data of the data set, use this process for two individual dataset and add the summation of these two values. After this we make the square root of this value and the mean difference is divided by this value. In this way we get the value of T-test where  $|t| = |-0.073| = 0.073$  or 0.1 which is close to 0. We also know that closer the t value to 0 means, there is no significance difference between those groups. From our estimated T-test value we get  $t=0.1$  approximately which indicates that there are no significant difference between our initial data set and the estimated/measured data set.

### 4.3 Estimated Cost of the Proposed System

After going through the research studies and the requirements found from the analysis of the surveys we have proposed and designed a prototype of a central cancer registry system. But it is equally important to estimate a cost for the proposed system so that the actual cost of implementing the system could be lessened. For these reasons we have estimated a rough cost for our system which is the initial cost of implementing the system. During the cost calculations different scenarios have been assumed to get a rough estimation.

As we have calculated the initial implementation cost of the system, it is roughly assumed that a total number of five thousand patients will use the system and for the worst case scenario. The cost includes the cloud service cost which is calculated by the assumption of using Linode cloud service, the domain cost and the cost of Textract API for which we have used AWS (Amazon Web Services). We have also assumed that all users, a total of five thousand people will use the system on a daily basis and the cost is calculated on daily and monthly usage of the proposed cancer registry system.

**Let's assume,**  $Total user = 50000$

$Daily max file uploaded by a registered user = 2\$$

$Daily user (worst case scenario) = 50000$

$Max upload size of 2 files = 1MB$

**One day storage used by user,**

$= (Daily user \times Daily Max file upload \times 2 files size)$

$= (50000 \times 2 \times 1) = 100000MB = 100000/1000 = 100GB.$

**Monthly storage used by registered users**  $= 100GB \times 30 = 3000GB$

$= (3000/1000)TB = 3 TB$

**For per page reload we need 1 MB Bandwidth**

$$= (\text{Daily user}) \times (\text{per page reload}) \times (2 \text{ filesize})$$

$$= 50000 \times (1 + 1) = 100000 \text{ MB} = 1 \text{ TB}$$

**Total monthly cost for Bandwidth, storage and 8GB memory by using Linode cloud service = 350\$**

**Card cost = 0.22\$**

**Domain cost monthly = 0.83\$**

**By using Amazon Textract API, per 1000-page cost in 1 month**

$$= 1.50\$ = 1.50/30 = 0.05 \text{ \$ per day}$$

**If each report is 1 page**

**2 reports**

$$= 2 \times 1 = 2 \text{ Pages}$$

**Daily cost for detecting document text API,**

$$= (\text{per day page cost} \times \text{daily user} \times 2 \text{ report}) / (1000 \text{ page})$$

$$= (0.05 \times 50000 \times 2) / 1000 = 5\text{\$}$$

**Monthly cost for detecting document text API =  $30 \times 5 = 150\text{\$}$**

**Total monthly cost for whole registry system**

$$= (350 + 150 + 0.22 + 0.83)\text{\$}$$

$$= 501.05\text{\$}$$

It was mentioned earlier that the estimated cost is based on the assumption of the implementation phase. It may seem that the cost is little higher but when the system is used generally then the costs will be reduced dramatically. And the estimated cost of our system could not be compared as there still remains scarcity of cancer registry systems which will not only give the patients a personalised care but also can be used as a central cancer registry. And the registries we have found in the related works, their features and the cost are disclosed. So, if we analyse the estimated cost of our system then it seems to be feasible when it will be used by a mass number of populations.

# **Chapter 5**

## **Design & Implementation**

An analysing over a large amount of data identified the requirements for the Cancer Registry System which will solve the problems that the doctors and patients face. According to those requirements we intend to propose a model of a central cancer registry system. In this chapter the design and implementation process of the proposed system is elaborated.

### **5.1 Design of Prototype**

According to the analysis the main requirement of the system is to successfully store a patient's diagnosis and treatment report. Besides, it is necessary to view the patient information in a general format so that the doctor can easily understand, observe and analyse the reports. For example, if a patient does two blood tests from two different diagnosis centres then the format or template of those tests won't be the same. Therefore, in our proposed system we have selected a general format and implemented that so that all types of blood tests could be seen in a general format which will help the doctor to easily find the keywords they are searching. The whole workflow of the system is divided into three parts which are admin, doctor and patient. The registration part of the doctor and patient needs the approval of the admin. Moreover, when a patient is registered to the central cancer registry system then an unique qr code is generated for him/her and it is also added to the patient id card. So, when a patient visits a doctor s/he will give the id card to the doctor so that s/he can go to the patient panel. Firstly, a doctor needs to log in to his panel and then he can go to the patient panel for further work from there. Though a doctor can go to the patient panel by scanning the qr code from the id card but they can also go to the patient panel by entering the specific patient's id. Once a doctor enters the patient panel the session for the doctor ends there and another session starts from there.

The patient panel shows the patient information, diagnosis and treatment data. The doctor can upload the image of the diagnosis reports which will be saved to the central database of the NICRH(National Institute of Cancer Research and Hospital). The system works as a central registry and the database is maintained by the NICRH so there is no primary authority to the different hospitals. Hence, if a patient ever changes the doctor or hospital then s/he does not need to be worried for the previous reports as they are maintained by the NICRH database. In diagnosis data the blood

test part we have implemented OCR which will convert the image of the blood test to text file and from that the necessary information is extracted in order to maintain a general format of the blood test. Lastly, if there is any need to upload to the image of a prescription then the doctor can upload it and in the therapy part the doctor can also add therapy. Thereby, through the proposed cancer registry system the treatment and diagnosis reports of a patient are efficiently stored and viewed according to the requirements which were identified by the analysis.

### 5.1.1 Doctor Panel

As doctors play the mainland major role in this system, we provide unique features to the doctor. For that in the doctor panel at first, each doctor must have to sign up. To sign up s/he must have to provide his/her name, email address, a valid password and then click on the register button to make it done. After successful sign up s/he is able to log in the system. But before that login admin approval is needed as only the admin has the power to make a doctor registered in the system. When a doctor logs in the system s/he can see one of the main features of our system which is displayed and that is the QR based login system. From figure 5.1 we can see that there are two types of login features a doctor can find. One of these is the QR code scanning system. If a patient is unable to provide his valid username and password to the doctor, then in this process the doctor has to scan the QR code which is printed on the card of the patient and then show it on the camera. And this allows the doctor to login to that patient's profile without knowing his user id or password. By the use of this QR code, doctors can directly get access to that particular patient's profile. Another thing a doctor can do and that is the manual login. For that he has to provide the particular patient's id to the given box and then press go. In this time the same things happen, doctors can access the patient profile and see the details of the patient's history. When a doctor enters a patient ID or scans the QR code then he can go to the patient portal (figure 5.2).

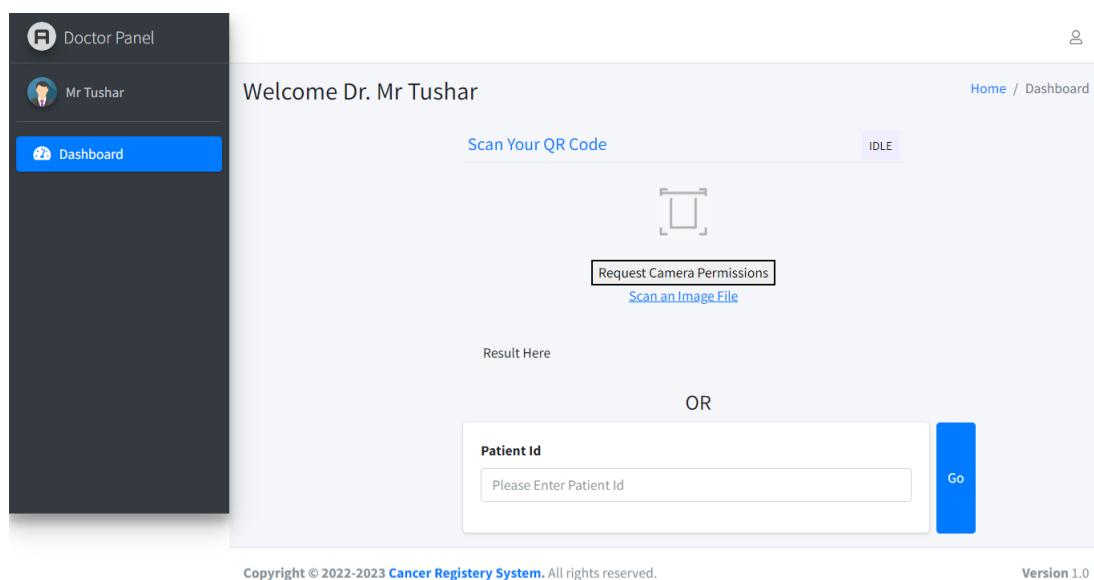


Figure 5.1: UI of Doctor Panel.

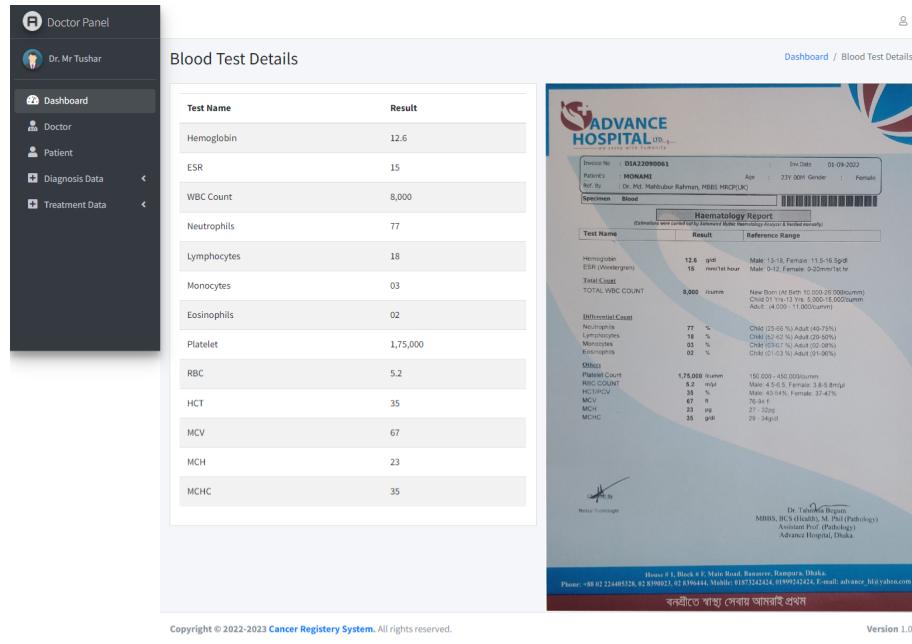
The screenshot shows a medical software interface. On the left is a dark sidebar with a navigation menu. The main area has a header 'Blood Test Report' with a 'Search' bar and a 'Search' button. Below the header is a table titled 'Data Analysis' with columns for Date, Hemoglobin, ESR, WBC Count, Neutrophils, Lymphocytes, Monocytes, Eosinophils, Platelet, RBC, HCT, MCV, MCH, and MCHC. Three rows of data are shown. Below the table is a section titled 'Reports' with a table showing 'Patient' information (John wick and Sujan) with dates (2022-08-01 07:24:53 and 2022-08-01 06:35:23) and 'Action' buttons (View and Delete). At the bottom of the sidebar, there are links for CT Scan, MRI, PET Scan, Ultrasonography, X-ray, Biopsy, Histopathology, and Treatment Data.

Date	Hemoglobin	ESR	WBC Count	Neutrophils	Lymphocytes	Monocytes	Eosinophils	Platelet	RBC	HCT	MCV	MCH	MCHC
2022-09-16 06:26:17	13.3 gm/dl	15	8000	70	25	03	02	150000	4.5	39	86	29	33
2022-09-16 06:21:49	10.8 gm/dl	66	11000	71	24	03	02	210000	3.8	32	84	27	35
2022-09-16 06:09:01	12 gm/dl	25	9500	70	24	03	03	220000	3.9	29	85	27	36

Figure 5.2: UI of Diagnosis Panel.

Moreover, a doctor has the access to make any update in the patient information if needed. From figure 5.2 on the left side the page there are Patient Profile, Diagnosis Data and Treatment Data. In the patient profile the doctor can see the patient's background and edit the information if required. Then if the doctor clicks on the Diagnosis Data then it elaborates into nine parts which are Blood Test, CT Scan, MRI, PET Scan, Ultrasonography, X-Ray, Biopsy and Histopathology same as figure 5.2 . When a doctor clicks on the Blood Test then a page with the blood test information against the date it was updated appears figure 5.2 .Then in this page the doctor is able to add the blood test if required by clicking on the button Add Blood Test . When a doctor clicks on the Add Blood Test button then a new page figure 5.2 appears and in this page the doctor has to select the image of the blood test from the device and press the blue button Add Blood Test for uploading the blood test. When an image of a blood test is uploaded then some specific information from the blood test is extracted from the image. For this process we have used the OCR process to convert the image into a text file and form that text file only necessary information is extracted and used in the table of the blood test (figure 5.2 & figure 5.3). We can see the extracted information if we click on the View button from fig4 where all the information from the blood test is stacked together. Here blood tests with any format are shown in a general format. Not only that but also a doctor can view the uploaded picture and delete the particular picture. If he wants to download the picture then he can easily download different formats of that picture like CSV, Excel, Pdf and can print it. Another special feature of the doctor panel is the searchability into the table. As we have implemented OCR in these blood test report parts, all the values from the picture will be extracted and saved to the database date wise and shown in the analysis blood test part. As we have mentioned before blood tests for different diagnosis centres differ with each other so the OCR method is used to extract the data from the image of the blood test and show them in a general format so that the doctor can easily identify the keywords with specific information he is searching for. When a doctor searches for any particular information about any patient then date wise updated information will be shown here.

In the same way doctors can add specific pictures to the CT Scan, MRI, PET Scan, Ultrasonography, X-Ray, Biopsy and Histopathology with the view functionality.



Copyright © 2022-2023 Cancer Registry System. All rights reserved.

Version 1.0

Figure 5.3: UI of Blood Test Report Page.

Except the blood test the other parts of the Diagnosis Data have similar functionality. For example, if a doctor wants to add a CT Scan report then firstly he has to click CT Scan and a page containing CT Scan information will appear (figure 5.4) and from there he has to click on the Add CT Scan button. Then a new page will appear like fig8 where an empty box is dedicated if a doctor wants to comment on the report and below the empty box there is a button for selecting the image of the CT Scan from the device. After clicking on the blue button Add CT Scan the image of the CT Scan will successfully upload to the system and the doctor can view the image of the uploaded CT Scan by clicking the View button. Then a new page will appear from where doctors can see CT Scan details. The rest of the parts of the diagnosis data have similar patterns and functionalities where the doctor can see all reports sorted by the date updated along with individual detailed information when clicked on the view button.

Moreover, doctors can provide advice to the report panel. There is another segment present and that is the treatment data segment. In the treatment data segment there are two parts one is Medicine Data and another one is Therapy. If a doctor wants to upload any image of prescription and add some details then at first he has to click on Medicine Data. Here in the medicine data part the doctor can upload the prescriptions of the necessary medicines. The process of uploading the image of prescriptions and viewing them is similar to the process of CT Scan. Then if any therapy is required for any patient then a doctor can add therapy by clicking on the Add Therapy button. After that a new page will appear where doctors will have to fill up the information of Therapy Type, Starting Date, End Date and number of doses. Like all the other cases the doctor can view, edit and delete the pictures. To go to another patient's profile doctor can use the back button and by using the logout button doctor can safely log out from his account.

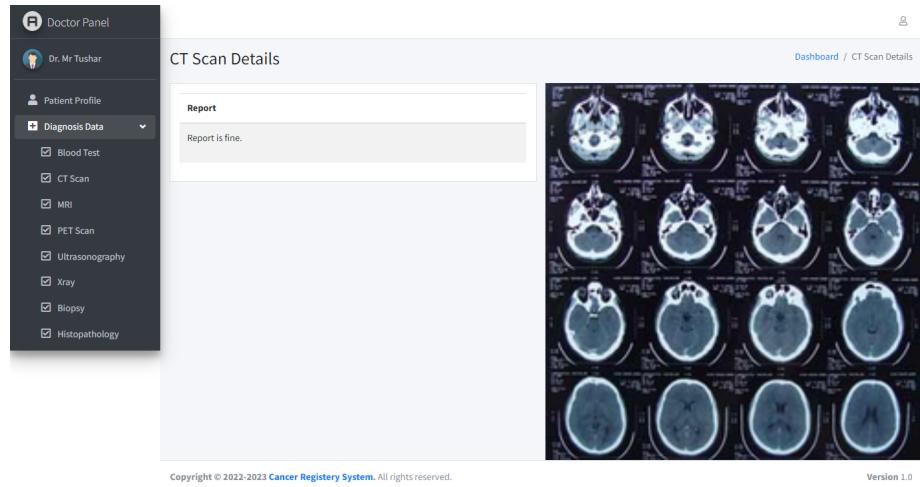


Figure 5.4: UI of CT Scane Page.

### 5.1.2 Patient Panel

For the successful login to the patient panel, a patient has to register himself by providing necessary things like name, email and password. After registering a patient a login to the system. Where in the patient profile page he can provide all the information regarding himself to update his profile. Then in the segment of diagnosis data the patient has also the access to add the blood test according to himself. Like a doctor, a patient also can view the uploaded picture and delete the particular picture. If he wants to download the picture then he can easily download different formats of that picture like CSV, Excel, Pdf and can print it. Because of the implementation of OCR, extracted value from the picture will be saved date wise to the database in the analysis blood test part. Doctor can search for any particular thing according to his will. When a doctor searches for any particular information about any patient then date wise updated information will be shown from here.



Figure 5.5: Patient's ID Card.

Patients can also add specific pictures to the other parts of the diagnosis test like in the CT scan, MRI, PET scan and other parts with the view functionality. Moreover, in the treatment data segment he can upload his prescribed prescription if he wants. In the therapy parts the patient has to fill up the therapy receiving data, its type

Patient Profile

**Patient Profile**

Name: Md. Jaynal Abadin Email: abc@abc.com

National Id: 1234567 Religion: Islam Occupation: Businessman

Address: Bashabo Occupation Address: Kakrail

Phone Number: 0188546762 Age: 45

Disease: Cancer Other Disease: Heart Disease

Current Treatment Institute: Lab Aid Previous Treatment Institute: Square

Password: (Leave blank if you want to keep previous password) Confirm Password: (Leave blank if you want to keep previous password)

**Update Patient**

Copyright © 2022-2023 [Cancer Registry System](#). All rights reserved. Version 1.0

Figure 5.6: UI of Patient Profile Page.

Add Blood Test

**Add Blood Test**

Image(Only JPG/JPEG/PNG are allowed formats)  
No file chosen

**Add Blood Test**

Copyright © 2022-2023 [Cancer Registry System](#). All rights reserved. Version 1.0

Figure 5.7: UI of Adding Blood Test Page.

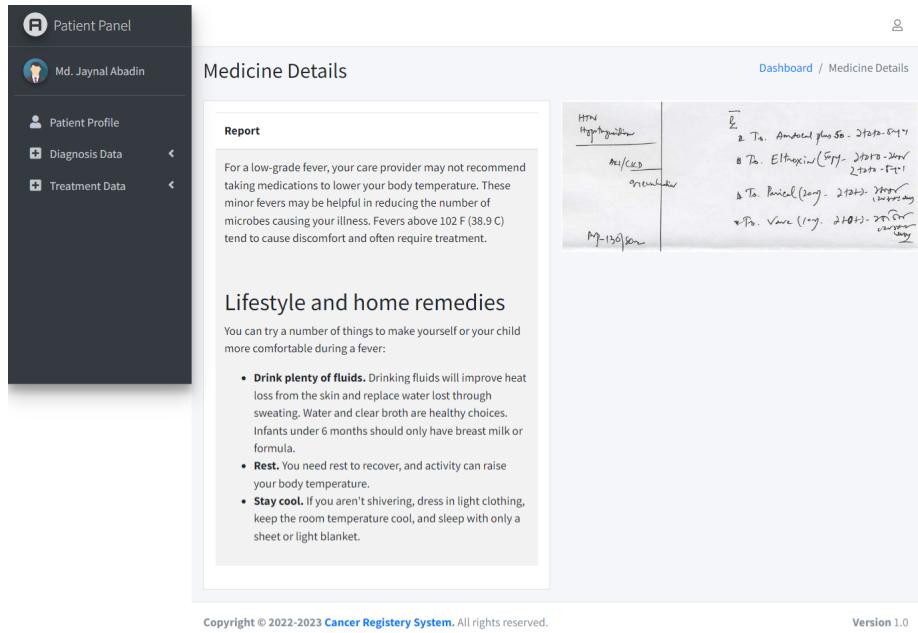


Figure 5.8: UI of Medicine Page.

and number of doses with the starting and ending date. Additionally, he can upload therapy reports and pictures for further uses. In this panel like a doctor, a patient also can view, edit and delete the pictures which he uploaded. Lastly, by pressing the logout button a patient can logout from his account.

### 5.1.3 Admin Panel

In the whole system, like doctor admin play a significant role. Actually, admin resides at the top of the hierarchy in terms of the editable process of anything in the system. After logging in as an admin, in the admins dashboard s/he is able to see all the doctors, patients along with the necessary documents and reports saved in the database. When any changes are needed regarding the patient page then both admin and doctor have access. Though the doctors and patients can register to the system by themselves but for further processing they need admin's approval. Doctors only can change or edit the patient profile, diagnosis and treatment information. But when it comes to change, edit or delete the username, password or email then only the admin has the access to do it. Admins can add doctors and patients and edit their information. Moreover, only the admin can see all the patients QR code and search on it. Again, the admin is the only person who distributes the QR code through a card to the patients. And this is one of the major features of our system. In this case every QR code holds an unique user id so that the user authentication can be preserved. Moreover, admin can edit and delete all the patients and doctors information from the system if needed.

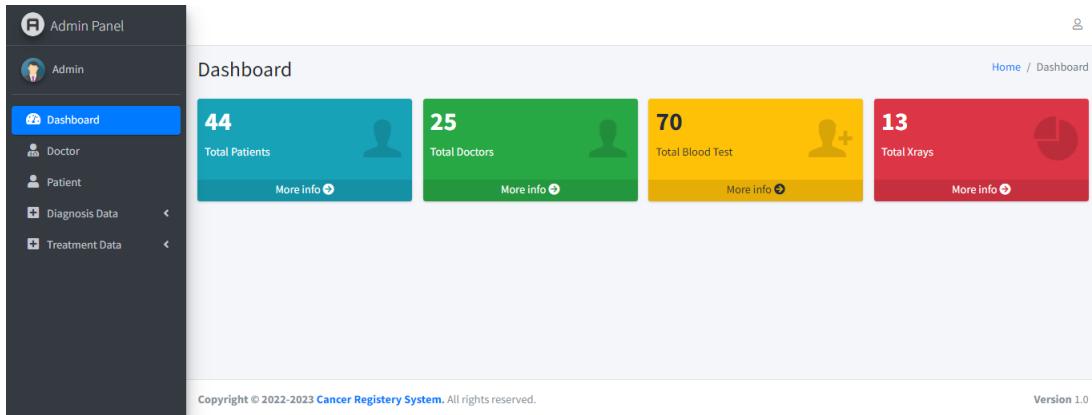


Figure 5.9: UI of Admin Dashboard.

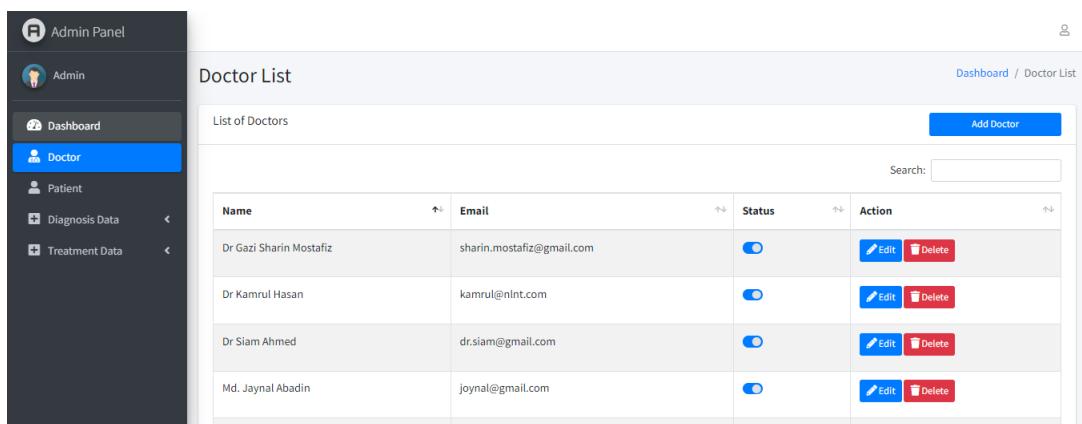


Figure 5.10: UI of Doctor List Page.

Like we have mentioned the diagnosis data and treatment data in the doctor panel there are also these segments in the admin panel. But their functionalities are slightly different from the doctor and patient panel. Here when an admin wants to upload a report then he has to select a patient by entering his id number in the search bar Fig. After that the report is saved to the database against the entered id of a patient. And in the information page of the segments of diagnosis data and treatment data the reports are showing for dedicated patients. Admin can view those reports, edit them or delete them if necessary. Thus, the admin has the full authority over every functionalities of the proposed system and resides at the top of the hierarchy.

QR Code	Id	Name	Email	Status	Action
	3	Gazi Sharin Mostafiz	sharin22222.mostafiz@gmail.com	<input checked="" type="checkbox"/>	<a href="#">Edit</a> <a href="#">Delete</a>
	5	Sujan	sujan@gmail.com	<input checked="" type="checkbox"/>	<a href="#">Edit</a> <a href="#">Delete</a>
	7	John wick	john_wick@123.cii	<input checked="" type="checkbox"/>	<a href="#">Edit</a> <a href="#">Delete</a>
	10	Mehedi	mehedi@gmail.com	<input checked="" type="checkbox"/>	<a href="#">Edit</a> <a href="#">Delete</a>

Figure 5.11: UI of Patient List Page.

## 5.2 Implementation of the Prototype

The architecture of the prototype is developed with laravel by creating a development environment of PHP Laravel with a database. We have followed the MVC pattern to develop our system according to the requirements so that the users can experience a user friendly and interactive system. The prototype's formation is done with the XAMPP server by creating a database using PhpMyadmin. All the information of this system is stored in the database. For example, when a patient or doctor is registered his information is saved to the database and when they intend to login to the system then their input is compared with the stored info. Every uploaded information is structurally stored in the database.

Furthermore, when an image of a blood test is uploaded then that image is converted to text file and from that text file we extract necessary data and store them to the database. The image to text conversion is done by the OCR process. For developing the prototype functional for OCR we have used the Textract API of AWS( Amazon Web Services). When a user uploads a report to our registry system, the Textract API of AWS service assists in extracting the printed content of the text file. And the text generated from the given image is returned in an array format.

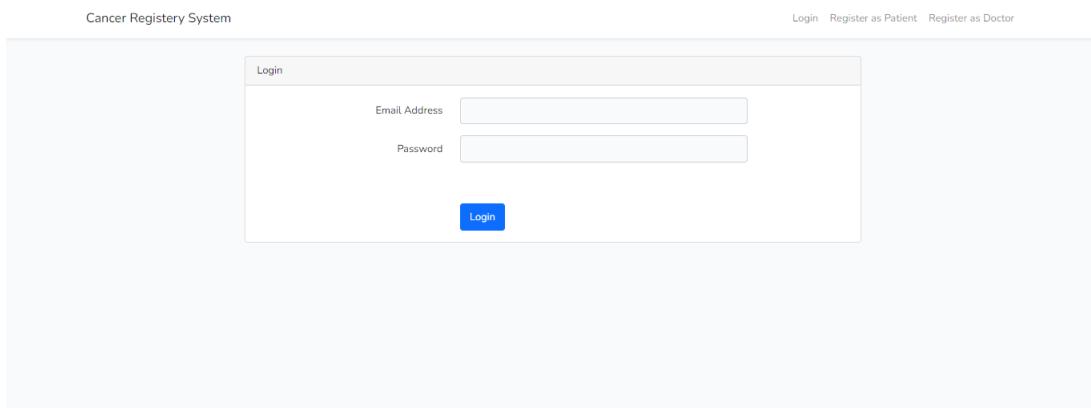


Figure 5.12: Homepage Of Webpage.

If we expand the array by using the dd() function of Laravel which dumps the

contents of a variable to the browser while preventing the execution of other scripts, then we will receive output information such as attributes or the original section. Then we extract the necessary data from that array and store it in the database which is then shown in a general format in the system. This cancer registry's principal responsibility is to keep record of cancer patients and their treatment progress. To create a cost-effective and user-friendly cancer registry system, we separated our registry system into three distinct panels, each with two database. The first page of the prototype we create is the login page, where one login page for our system user (doctor, patient, admin). If any of user not registered with our system, there individual Register system for doctor and patient. You will be redirected to the login and login page after clicking on the desired type. You must provide the following information during registration: username, E-mail, number, gender, address and password. After, user can then logging the system using their username and password.

# **Chapter 6**

## **Discussion**

The aforementioned sections contain the clear idea of the current scenario of the patients and doctors where both of them suffer because no proper system exists that will help to maintain and store the treatment and diagnosis reports which is necessary for an ongoing treatment. Furthermore, we tried to present the challenges that patients face by the research studies and surveys. The influential findings from the research studies and surveys identified the current miserable conditions of the patients where an extra tension and hassle are added on top of the headache they are going on. Moreover, the crucial findings are discussed over the aforementioned sections and a crucial system is proposed to solve the problems and fulfil the requirements of the patients and doctors. Hence, we are on the verge of bestowing the discussion on the findings and implications of the proposed cancer registry system and the impacts of them.

### **6.1 Understanding the Context of the Complications of Patients**

Bangladesh is an under-developed country where still a large number of people live under the poverty line. Scarcity over resources is a common phenomena in this country. Hence, the headache of treatment of diseases is considered as an extra burden here. When scarcity strikes and there is not enough money to run a family then the thought of treatment is considered as luxury. In that situation people think how fast they will be cured and the extra expenses will be over. But when a patient is diagnosed with cancer then the situation of the patient can be considered with the metaphor that the whole sky has fallen on his head. From our survey we have found that most of the people who travel round and round from hospital to home and home to hospital are the middle class people. Because if anyone from the elite class of our country is diagnosed with cancer then they travel abroad for advanced treatment. Therefore, the population of our research are mostly the people from the middle class. And the additional expenses of treatment is a great burden for most of them. In this scenario if any of their diagnosis reports gets lost then most of the cases they have do the diagnosis again which leads to extra expenses. On the other hand if ever a patient changes his hospital or doctor then in most of the cases the previous reports are not accepted by the doctors for various reasons. In the majority of the cases they have to get the diagnosis again. Thus, these extra

expenses become a headache for them. These extra expenses increase the cost of their treatment and sometimes the cost becomes unbearable and the patient stops treatment.

Furthermore, whenever a cancer patient visits a doctor he has to carry a lot of reports with him. This is an extra hassle for a patient. Besides, from a huge collection of reports it is difficult to find the required report or document. And most of the time it costs the valuable time of a doctor to find the required report from the mountain of documents. Most often when a new patient comes to a doctor then the doctor advises diagnosis tests even if those tests already exist. Because in most of the cases patients can not show the required report at the right time to the doctor. And the hard copies are also tough to store because they are easily demolished. It takes a good amount of care to store the hard copies of the reports and takes a lot of space. From our survey and research studies we have found that older reports are commonly lost for various reasons. Moreover, cancer treatment is a lengthy process which takes a lot of time to get cured. So, the patients need to continue treatment for a long time. But as time passes by patients become less serious about their treatment. Because, most of the patients tend to become careless once their disease is getting better or they are a little bit cured. And from our observations we have found that when a cancer patient is after taking chemo or another type therapy some of the patients become better than before. And this is the phase where they become careless and do not take proper care of the reports along with their health and they suffer. Whatever the reason is their report is lost they have to take additional pain on their head with extra expenses. Besides, it is not easy for doctors to find out the necessary documents or reports from a patient's big collection as a result they give new tests. Thus, the patients have to suffer the most because of the mismanagement of the diagnosis and treatment reports which is the reason for the emerging aforementioned problems.

## 6.2 Designing According to the Requirements

The primary user of our proposed cancer registry system is the doctors so simplicity and usability are the main key factors. From our survey we have found that a huge number of doctors are used to manoeuvring smart phones. But we have to keep in mind that they are not tech-skilled people so the system requires a simple UI so that they can easily understand and use our proposed cancer registry system. Moreover, the doctors are not the only users of this system. Patients can also upload their reports to the system. For example, someone is being treated under a particular doctor but the patient is away from the doctor's reach. Now if it is necessary to have a diagnosis and show that report to the doctor then the patient can do the diagnosis to his nearby diagnosis centre and then upload his report to the system so that the doctor can check it and give the patient advice. But if we compare the technological knowledge of patients to doctors then the patients have less knowledge. That is why our system is designed in a way that both patients and doctors can easily interact with the system.

The doctors see patients and do the work related to them when they are in their

chamber. From our survey it has been found that around 84% of doctors have an internet connection in their chambers. Hence, the majority of the doctors have internet connection in their chambers which is a positive side for deploying our system as it requires internet connection. And when any information is uploaded to the system then it is stored in the database. The architecture of the database is designed in a way that some level of security can be maintained in the system. In the previous sections we have shown that the admin is at the top of the hierarchy who can see anyone's information. After that doctor is the second in that hierarchy who can see and edit patient's information along with his own. After that patient is at the edge of the hierarchy who can only see and edit his information. In the architecture of the database security is maintained in a way that a doctor can not see or edit any other doctor's or admin's information. Same goes with the patient where a patient can not see or edit any other patient's or doctor's or admin's information. These restrictions not only ensure security of the system but also keep the system simpler where everyone registered has their own model of functions to operate. Therefore, the proposed cancer registry system is designed by maintaining the requirements of the patients and doctors where they can successfully operate and be part of advanced technological change to solve their never ending problems and end their miseries.

### 6.3 Comparison With Existing Works

For establishing the proposed cancer registry system as a feasible prototype the development and advocacy of the system has pointed out nineteen cancer registry systems for storing cancer data which are already mentioned in related works. Among them only five have some similar functionalities to our proposed cancer registry system. Though few of them have the common functionality of being a central cancer registry, most of them do not have a secure accessibility to patient panel like our system. And the most contrasting feature is that the existing systems do not have any general format for showing the diagnosis where reports of different formats and templates can be viewed in a general format for easy understand ability. In fact, most of the existing cancer registry systems store demographic data for statistical analysis. Dr. Hossain (Resident, Radiation oncology, National Institute of Cancer Research and Hospital) gave positive feedback on our working and asserted that,

*“Your proposed system is novel because you have implemented an OCR for showing the different reports from different diagnosis centers in a general format and QR code login system which makes it unique from other existing registry systems .”*

The following table compares our proposed cancer registry system to other existing systems or literature which is appropriate for our research:

Existing System / Paper Name	General Features	Common Factors With Proposed Cancer Registry System	Contrasting Factors With Proposed Cancer Registry System
OWise - Breast Cancer and Support (2021) [19]	<p>An app to store the diary entries of patients where they receive individualised care by uploading all of their treatment information, including voicemails, voice recordings and pictures in one convenient location and doctors have access to patient information.</p>	<p>1. Ensures personalised care to the patients. 2. In both systems patient can upload their information and doctors can view them for treatment purpose.</p>	<p>1.The uploading criteria of patient data is like diary entries where there is no structural way to categorise different types of reports or data and no general format to show reports of different templates and formats. 2. In this system only patients can upload their information where doctors do not have this feature.</p>
Central Cancer Registry (CCR) (2019) [20]	<p>In this system the oncologists can utilise stored data in the cancer research where the registered doctors can establish and amend patient profiles, write prescriptions for them, view their treatment histories and stage the cancer using the TNM staging system, which has two types: cType and pType.</p>	<p>1.Only registered doctors can add and edit patient information. 2.Doctors can view information in a structural way.</p>	<p>1.This model does not have any secured access system by which doctors can go to the patient panel. 2.In this system only doctors can add patients and their information . 3.The primary use of the patient data is statistical analysis and staging cancer.</p>

Cancer, Cancer Control and Bangladesh (2012)[21]	Focuses on the demographic information of cancer patients which are used for statistical analysis.	Stores the demographic information of cancer patients.	1. This system does not store patient's diagnosis and treatment data. 2. The only basis of this system is to store demographic data of patients where doctors and patients do not have any access.
Global Cancer Registry Software (2021) [17]	This registry is renowned for its capabilities for patient care, medical research, clinical investigations and open database access.	1. In both systems patients get medical care. 2. Patient's diagnosis and treatment reports are stored for medical care and clinical studies. 3. Doctors can view patient data to understand the current situation of a patient.	1. Does not have any secured access system for login. 2. Only patients can upload their data but doctors can not which is not feasible for our country as a huge number of cancer patients are illiterate.
Kentucky Cancer Registry (2019) [51]	It is a user-friendly and dynamic data input system, it may save incomplete abstractions for later completion and perform thorough validations and edit checks for single fields, between fields, and between records during data entry.	1. Both are interactive and handy data entry systems. 2. Have edit option for future update and keeps records during data entry with respective dates.	1. This system is primarily used for statistical analysis of cancer patients so there is no feature present for patient care. 2. Only doctors and admin of the system can upload patient data.

Table 6.1: Comparison of the proposed system with existing works[18], [39], [42], [50], [52].

## **6.4 Limitation**

In every working system there exist some limitations. And these limitations appear due to different reasons. As this is our first practical implementation of such a huge system. So, in our system we also have some limitations.

Firstly, we use OCR for data extraction only on the patients' blood report. However, there are also some other reports present in our system like MRI, X-Ray, CT scan, Biopsy etc. which also carry values and data. But in less amount compared to the blood report. We extract all data from the blood report whereas in case of others we directly save the picture to the database. So, this creates another limitation of our system. Lastly, OCR is one of the important features of our system and it plays a crucial role in the overall data extraction process. But as OCR worked through the multi-layer process so sometimes it creates problems. When we upload multiple pictures to the system at that time OCR faces problems with the low-quality pictures. As a consequence, we get less accurate value from that picture compared to the values of high-resolution pictures. And in this way getting less accurate value turns to another major limitation of our system.

# Chapter 7

## Conclusion And Future Work

The most crucial task for any cancer control program (CCP) is the availability of reliable cancer incidence data from which major risk factors in a population can be deduced. As a result, ‘Cancer Registries (CRs)’ constitute the basis of any meaningful CCP in every country. It is impossible to eliminate cancer patients in the short term. However, a well-developed and cost-effective cancer registry system enables us to accomplish this goal by allowing us to quickly identify and treat cancer patients. Numerous studies, research projects and other cancer-related initiatives have been ongoing for several decades. Bangladesh is a densely populated country where a large number of people are affected and die of lethal cancers each year. A well-designed registry system enables us to monitor the growth and providing accurate data on the incidence, prevalence, and mortality of cancer patients. While the high specific costs related to conducting and carrying out registration activities may account for some of the cost differences by size, other factors, such as the quality of data initially submitted to registries by reporting sources like hospitals and pathology laboratories, may be involved. Costs may be reduced overall if data collection is more automated or efficient. Therefore, the focus of our paper is to structure such a cancer registry system which will reduce the suffering of both doctor and patient. Considering the economic condition of Bangladeshi people we have tried our best to make the system cost effective as much as possible. Moreover, we try to make sure a user-friendly interface for both doctor and patient. Additionally, we surveyed both doctors and patients when developing the system and tried to include the features that they valued the most. To conclude with a positive note, launching a generalized cancer registry system for both doctor and patient is one of the crying needs of Bangladesh right now— and our cancer registry is nothing but a leap towards that cherished goal.

Due to the massive advancement of the IT sector, different useful and important systems are built. But the thing is there is no such system present, which appears at a first glance without any limitations and also does not require any further changes. Developer tests the system on a regular basis and makes necessary changes day by day to fix and overcome the existing limitations. In the same way, in our proposed prototype system we also have some limitations which we will fix in future. We designed our proposed prototype in such a way so that all the extracted values can be arranged date wise in a systematic manner and saved in the dataset. As we work with the Cancer Registry System so the data is very important. As we are using OCR from AWS, so in terms of data processing sometimes it predicts

wrong values because of blur or unclear picture. As a consequence, the AWS does not predict the report value correctly and saves almost close values in the array indexes or sets the value to the wrong index position. However, these index values are inappropriate and for this reason at the present situation we get almost 96.86% accuracy from our system. Though this accuracy value is quite good. But as we are working on the data collected from the medical field, it has to be more accurate. Otherwise, any accident can happen at any time. We will work on its extraction process, so that we can achieve almost 100% accuracy. Again, as this is the initial phase of our system deployment. So, at present we run and check its accuracy with other features in our local machine. But we have developed our system in terms of Bangladeshi context. We know that Bangladesh is a developing country so there are a lot of villages where the usability of the internet is not so regular. So, if we make this system accessible offline then the people of that area also get the benefits of having such a system. As a result, in the near future we will try to launch a software-based registry system which is also accessed by people offline. We will also try to make a portable application by which any one can able to use this system without even installing them in their machines. Finally, there are different medical reports present in our proposed Cancer Registry System like MRI, X-Ray, CT scan, Biopsy etc. and we only implement OCR only on the patients' blood report for data extraction. However, there are always some values and data left in all of the reports. But the amount of other report's data is less, compared to the data carried by the blood report. Each of the minor medical data is important to analyze the patient's overall health condition, so these data are crucial for the better analysis. As a result, we have to extract all the data from all of the reports and save these extracted data into the database. In this way to achieve the goal we extract all data from the blood report along with others. So, this creates another limitation of our system which we will implement in future.

# Bibliography

- [1] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [2] S. Huq, “Common cancers in bangladesh: Their trends through last three decades,” *Bangladesh Med J*, vol. 17, no. 3, p. 5563, 1988.
- [3] J. Kanai, S. V. Rice, T. A. Nartker, and G. Nagy, “Automated evaluation of ocr zoning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 86–90, 1995.
- [4] P. Akhter, M. Uddin, and S. Sharma, “Patterns of malignant neoplasm: A three year study,” *Bangladesh Med J*, vol. 27, pp. 29–32, 1998.
- [5] M. Seeger and C. Dance, “Binarising camera images for ocr,” in *Proceedings of Sixth International Conference on Document Analysis and Recognition*, IEEE, 2001, pp. 54–58.
- [6] L. C. Harlan, J. Abrams, J. L. Warren, L. Clegg, J. Stevens, and R. Ballard-Barbash, “Adjuvant therapy for breast cancer: Practice patterns of community physicians,” *Journal of clinical oncology*, vol. 20, no. 7, pp. 1809–1817, 2002.
- [7] A. L. Potosky, L. C. Harlan, R. S. Kaplan, K. A. Johnson, and C. F. Lynch, “Age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer,” *Journal of clinical oncology*, vol. 20, no. 5, pp. 1192–1202, 2002.
- [8] J. Ferlay, “Cancer incidence, mortality and prevalence worldwide,” *GLOBO-CAN2002*, 2004.
- [9] B. Gatos, I. Pratikakis, and S. J. Perantonis, “An adaptive binarization technique for low quality historical documents,” in *International Workshop on Document Analysis Systems*, Springer, 2004, pp. 102–113.
- [10] S. M. A. Hussain and M. Zaman, *National Cancer Control Strategy and Plan of Action 2009-15*. Dec. 2008.
- [11] M. A. Moore, H.-R. Shin, M.-P. Curado, and T. Sobue, “Establishment of an asian cancer registry network-problems and perspectives,” *Asian Pac J Cancer Prev*, vol. 9, no. 4, pp. 815–832, 2008.
- [12] R. Sankaranarayanan, N. Bhatla, P. E. Gravitt, *et al.*, “Human papillomavirus infection and cervical cancer prevention in india, bangladesh, sri lanka and nepal,” *Vaccine*, vol. 26, pp. M43–M52, 2008.
- [13] M. A. Rahiman and M. Rajasree, “A detailed study and analysis of ocr research in south indian scripts,” in *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, IEEE, 2009, pp. 31–38.

- [14] M. Zaman and M. Baki, “Cancer registry report-national institute of cancer research and hospital 2005-2007,” *Dhaka: WHO*, 2009.
- [15] A. Alwan, D. R. MacLean, L. M. Riley, *et al.*, “Monitoring and surveillance of chronic non-communicable diseases: Progress and capacity in high-burden countries,” *The Lancet*, vol. 376, no. 9755, pp. 1861–1868, 2010.
- [16] M. L. Gjerstorff, “The danish cancer registry,” *Scandinavian journal of public health*, vol. 39, no. 7\_suppl, pp. 42–45, 2011.
- [17] M. P. Robillard and R. DeLine, “A field study of api learning obstacles,” *Empirical Software Engineering*, vol. 16, no. 6, pp. 703–732, 2011.
- [18] M. A. Kalam and T. Ahmed, “Cancer, cancer control and bangladesh,” *Bangladesh Journal of Plastic Surgery*, vol. 3, no. 1, pp. 1–2, 2012.
- [19] S. Roopa and M. Rani, “Questionnaire designing for a survey,” *Journal of Indian Orthodontic Society*, vol. 46, no. 4\_suppl1, pp. 273–277, 2012.
- [20] R. Verma and J. Ali, “A-survey of feature extraction and classification techniques in ocr systems,” *International Journal of Computer Applications & Information Technology*, vol. 1, no. 3, pp. 1–3, 2012.
- [21] S. A. Hussain and R. Sullivan, “Cancer control in bangladesh,” *Japanese journal of clinical oncology*, vol. 43, no. 12, pp. 1159–1169, 2013.
- [22] E. Ferrara, P. De Meo, G. Fiumara, and R. Baumgartner, “Web data extraction, applications and techniques: A survey,” *Knowledge-based systems*, vol. 70, pp. 301–323, 2014.
- [23] V. Le and S. Gulwani, “Flashextract: A framework for data extraction by examples,” in *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation*, 2014, pp. 542–553.
- [24] M. A. Moore, S. Sangrajrang, and F. Bray, “Asian cancer registry forum 2014-regional cooperation for cancer registration: Priorities and challenges,” *Asian Pacific Journal of Cancer Prevention*, vol. 15, no. 5, pp. 1891–1894, 2014.
- [25] A. N. Richardson, “Quality of cancer registry data,” 2014.
- [26] S. Vijayarani and A. Sakila, “Performance comparison of ocr tools,” *International Journal of UbiComp (IJU)*, vol. 6, no. 3, pp. 19–30, 2015.
- [27] A.-M. Forsea, “Cancer registries in europe—going forward is the only option,” *Ecancermedicalscience*, vol. 10, 2016.
- [28] M. Hyman and J. Sierra, “Guidelines for writing good survey questions,” *NMSU Business Outlook*, vol. 14, 2016.
- [29] S. Subramanian, F. K. Tangka, M. C. Beebe, D. Trebino, H. K. Weir, and F. Babcock, “The cost of cancer registry operations: Impact of volume on cost per case for core and enhanced registry activities,” *Evaluation and program planning*, vol. 55, pp. 1–8, 2016.
- [30] Z. Mohammadzadeh, M. Ghazisaeedi, A. Nahvijou, S. R. N. Kalhori, S. Davoodi, and K. Zendehdel, “Systematic review of hospital based cancer registries (hbcrs): Necessary tool to improve quality of care in cancer patients,” *Asian Pacific journal of cancer prevention: APJCP*, vol. 18, no. 8, p. 2027, 2017.

- [31] M. Jacobs, J. Johnson, and E. D. Mynatt, “Mypath: Investigating breast cancer patients’ use of personalized health information,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–21, 2018.
- [32] J. Noor and M. G. Z. Husna, “Astha: Handy and manageable cancer management system for bangladesh,” Apr. 2018.
- [33] C.-J. Chiang, Y.-W. Wang, and W.-C. Lee, “Taiwan’s nationwide cancer registry system of 40 years: Past, present, and future,” *J Formos Med Assoc*, vol. 118, no. 5, pp. 856–858, 2019.
- [34] J. Memon, M. Sami, R. A. Khan, and M. Uddin, “Handwritten optical character recognition (ocr): A comprehensive systematic literature review (slr),” *IEEE Access*, vol. 8, pp. 142 642–142 668, 2020.
- [35] S. Pervez, A. A. Jabbar, G. Haider, *et al.*, “Karachi cancer registry (kcr): Age-standardized incidence rate by age-group and gender in a mega city of pakistan,” *Asian Pacific journal of cancer prevention: APJCP*, vol. 21, no. 11, p. 3251, 2020.
- [36] D. Zhang, C. Yin, J. Zeng, X. Yuan, and P. Zhang, “Combining structured and unstructured data for predictive models: A deep learning approach,” *BMC medical informatics and decision making*, vol. 20, no. 1, pp. 1–11, 2020.
- [37] K. Thayer, S. E. Chasins, and A. J. Ko, “A theory of robust api knowledge,” *ACM Transactions on Computing Education (TOCE)*, vol. 21, no. 1, pp. 1–32, 2021.
- [38] *Build your own Optical character recognition (OCR) System using Google’s Tesseract and OpenCV*, howpublished = [https://cdn.analyticsvidhya.com/wp-content/uploads/2020/05/ocr\\_1.png](https://cdn.analyticsvidhya.com/wp-content/uploads/2020/05/ocr_1.png), note = Accessed: 2022-06-16.
- [39] *Central Cancer Registry (CCR)*, howpublished = [https://play.google.com/store/apps/details?id=com.cmedhealth.flutter\\_app\\_ccr\\_new&hl=en&gl=us&fbclid=iwar1pf2gvohc5nthpkuf0xp51qgvil6hwex6l\\_mk3rjvy-js7c2dbheqxpks](https://play.google.com/store/apps/details?id=com.cmedhealth.flutter_app_ccr_new&hl=en&gl=us&fbclid=iwar1pf2gvohc5nthpkuf0xp51qgvil6hwex6l_mk3rjvy-js7c2dbheqxpks), note = Accessed: 2022-08-21.
- [40] *Deskew/Autodeskew: What is this Image Processing Feature? ORPALIS :: Experts in Document Imaging & Document Management*, howpublished = <https://www.orpalis.com/blog/deskew-autodeskew-definition/>, note = Accessed: 2022-07-17.
- [41] *Despeckle*, howpublished = <https://www.iccp-portal.org/cancer-registries>, note = Accessed: 2022-07-14.
- [42] *Global Cancer Registry*, howpublished = <https://www.advancemarketanalytics.com/reports/87853-global-cancer-registry-software-market/>, note = Accessed: 2022-02-17.
- [43] *How do apis work?* <https://tray.io/blog/how-do-apis-work>, Accessed: 2022-09-01.
- [44] *How Image-to-Text Works (aka Optical Character Recognition)*, howpublished = <https://www.makeuseof.com/tag/ocr-image-to-text/>, note = Accessed: 2022-08-20.

- [45] *How to Train Tesseract OCR in Python?*, howpublished = <https://www.projectpro.io/article/how-to-train-tesseract-ocr-python/> 561, note = Accessed: 2022-08-22.
- [46] *Image Effects Component.* , howpublished = <https://www.websupergoo.com/helpie/default.htm?page=source%2f2-effects%2fdespeckle.htm>, note = Accessed: 2022-07-10.
- [47] *International Cancer Control Partnership (ICCP)* , howpublished = <https://www.iccp-portal.org/cancer-registries>, note = Accessed: 2022-07-16.
- [48] *Introduction to Pattern Recognition*, howpublished = [http://homepage.tudelft.nl/a9p19/papers/pr\\_intro.pdf](http://homepage.tudelft.nl/a9p19/papers/pr_intro.pdf), note = Accessed: 2022-07-20.
- [49] *Introduction to the Image Processing Functions*, howpublished = <https://www.leadtools.com/help/sdk/v22/image-processing-functions/introduction.html>, note = Accessed: 2022-06-14.
- [50] *Kentucky Cancer Registry*, howpublished = <https://www.kcr.uky.edu/software/cpdmsnet.php>, note = Accessed: 2021-07-04.
- [51] *Module 1: Cancer Registry*, howpublished = <https://www.tn.gov/health-health-program-areas/tcr/cancer-reporting-facility-training/module1.html>, note = Accessed: 2021-04-14.
- [52] *OWise - Breast Cancer Support*), howpublished = <https://play.google.com/store/apps/details?id=nl.onesixty.owise>, note = Accessed: 2022-07-20.
- [53] *What Is A Cancer Registry*, howpublished = <https://www.ncra-usa.org/about/become-a-cancer-registrar/what-is-a-cancer-registry>, note = Accessed: 2022-05-04.

# **Survey Questionnaire For Patients**

## **Form Id:**

1. Your name (optional) \_\_\_\_\_
2. Your age \_\_\_\_\_
3. Your gender :
  - (a) Female
  - (b) Male
  - (c) Other
4. Contact number \_\_\_\_\_
5. Your email address (if any) \_\_\_\_\_
6. Your Profession \_\_\_\_\_
7. District you live in \_\_\_\_\_
8. How many members do you have in your family (including you)? \_\_\_\_\_
9. Who is the earning member(s) of your family? \_\_\_\_\_
10. Does anyone in your family use a smart device like smart phone, laptop or others?
  - (a) Myself
  - (b) Myself and other family members
  - (c) Other family members
  - (d) No one of my family
11. Which type of disease are you suffering from? \_\_\_\_\_
12. How often do you visited the hospital in a week?
  - (a) 1-2 days
  - (b) 2-4 days
  - (c) 4-6 days
  - (d) Never visited
13. How often do you get diagnosed?
  - (a) Once a week
  - (b) Once a month
  - (c) Multiple times in a week
  - (d) Multiple times in a month

(e) Other

14. Do you have to maintain huge amount of hard copies of diagnosis and treatment reports?
- (a) Strongly Agree
  - (b) Agree
  - (c) Neutral
  - (d) Disagree
  - (e) Strongly Disagree
15. "More reports means facing more difficulties to manage those", What do you think about this statement?
- (a) Strongly Agree
  - (b) Agree
  - (c) Neutral
  - (d) Disagree
  - (e) Strongly Disagree
16. According to you, in which way a patient lost his/her reports?
- (a) Changing doctor
  - (b) Changing hospital
  - (c) Other:
17. What do you think that if a patient changes the doctors/hospitals, then he/she needs to renew the medical test reports by doing the tests again?
- (a) Always
  - (b) Often
  - (c) Sometimes
  - (d) Seldom
  - (e) Never
18. "There is a need of a central Cancer Registry System which can store diagnosis and treatment data" - what do you think?
- (a) Strongly Agree
  - (b) Agree
  - (c) Neutral
  - (d) Disagree
  - (e) Strongly Disagree
19. "A central Cancer Registry System would free you from the burden of carrying and maintaining hard copies of test/diagnosis reports" - to what extend do you agree with the statement?

- (a) Strongly Agree
- (b) Agree
- (c) Neutral
- (d) Disagree
- (e) Strongly Disagree

20. Do you familiar with any card based registry system?

- (a) Well aware
- (b) Somewhat aware
- (c) Neutral
- (d) Unaware
- (e) Not aware at all

21. "A card with QR code helps you to login without remembering user id and password to the system" - If a similar type of card is given to you, do you feel happy to use/carry it?

- (a) Definitely
- (b) Very Probably
- (c) Possibly
- (d) Probably Not
- (e) Definitely Not

22. "A doctor will use your patient id card for uploading your diagnosis treatment reports if required" - to what extent do you agree?

- (a) Strongly agree
- (b) Agree
- (c) Neutral
- (d) Disagree
- (e) Strongly Disagree

23. Express your opinion on using a Cancer Registry System. \_\_\_\_\_

# **Survey Questionnaire for Doctors**

## **Form Id:**

1. Your name \_\_\_\_\_
2. Your Age \_\_\_\_\_
3. Your gender
  - (a) Female
  - (b) Male
  - (c) Other \_\_\_\_\_
4. Your specialization \_\_\_\_\_
5. You are a doctor of -
  - (a) Public Hospital
  - (b) Private Hospital
  - (c) Other \_\_\_\_\_
6. How many chambers you are sitting in - \_\_\_\_\_
7. Do you have internet access in your chamber(s)?
  - (a) Yes
  - (b) No
  - (c) Maybe
8. How many patients do you have to maintain in a week? \_\_\_\_\_
  - (a) Very High
  - (b) High
  - (c) Moderate
  - (d) Low
  - (e) Very Low
9. "More reports means more difficulties toward a patient to handle", what do you think about this statement?
  - (a) Strongly Agree
  - (b) Agree
  - (c) Neutral
  - (d) Disagree
  - (e) Strongly Disagree
10. According to you, in which way a patient lost his/her reports?

- (a) Changing doctor
- (b) Changing hospital
- (c) Other: \_\_\_\_\_

11. What do you think that if a patient changes the doctors/hospitals, then he/she needs to renew the medical test reports by doing the tests again?

- (a) Always
- (b) Often
- (c) Sometimes
- (d) Seldom
- (e) Never

12. What technological devices do you use in your office hours?

- (a) Smart phone
- (b) Analog Phone
- (c) Laptop
- (d) Desktop
- (e) Other: \_\_\_\_\_

13. Approximately how much time (in hours) do you spend on your tech device?

- (a) 1-2
- (b) 3-4
- (c) 5-6
- (d) Other: \_\_\_\_\_

14. How much do you know about registry/hospital management system?

- (a) Well aware
- (b) Somewhat aware
- (c) Neutral
- (d) Unaware
- (e) Not aware at all

15. Have you ever used any registry/hospital management system?

- (a) Yes
- (b) No
- (c) Maybe

16. If yes then what type of data/ information stored in the system?

- (a) Demographic info.
- (b) Treatment info.

- (c) Diagnosis info.  
(d) Other \_\_\_\_\_
- 

17. If no then do you ever felt any need to have a registry system which would help you to manage and observe patients data?
- (a) Yes  
(b) No  
(c) Maybe
18. "There is a need of a central Cancer Registry System which can store diagnosis and treatment data" - what do you think?
- (a) Strongly agree  
(b) Agree  
(c) Neutral  
(d) Disagree  
(e) Strongly Disagree
19. "A central Cancer Registry System would free patients from the burden of carrying and maintaining hard copies of test/diagnosis reports" - to what extent do you agree with the statement?
- (a) Strongly agree  
(b) Agree  
(c) Neutral  
(d) Disagree  
(e) Strongly Disagree
20. Do you know about card based registry system?
- (a) Well aware  
(b) Somewhat aware  
(c) Neutral  
(d) Unaware  
(e) Not aware at all
21. "A card with QR code helps you to login without asking user id and password to the system " - If a similar type of card is given to you, do you feel happy to use it?
- (a) Definitely  
(b) Very Probably  
(c) Possibly  
(d) Probably Not

(e) Definitely Not

22. You can use a patient's id card for uploading his/her diagnosis treatment reports if required - do you think this process will be helpful for patients?

- (a) Strongly Agree
- (b) Agree
- (c) Neutral
- (d) Disagree
- (e) Strongly Disagree

23. Express your opinion on using a Cancer Registry System. \_\_\_\_\_