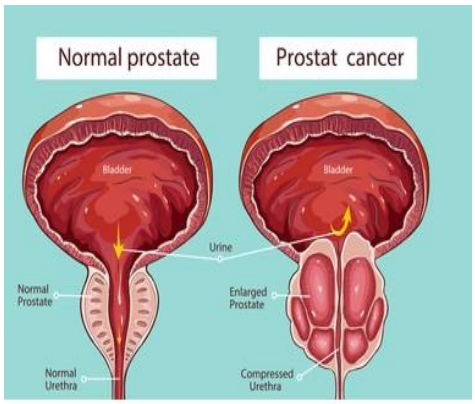
Tumor diagnostic using machine learning and image processing



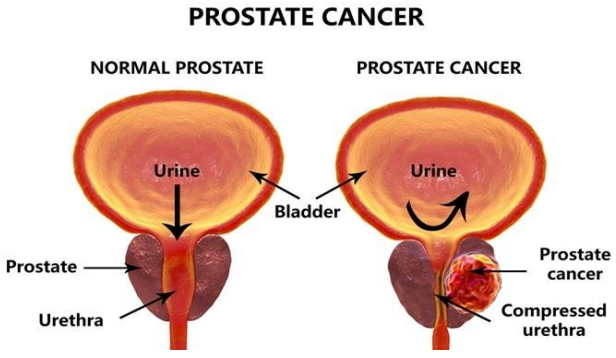


Table of contents

Motivation

The prostate is a small gland at the base of the bladder in males. It encircles the first part of the urethra. The prostate gland plays an important role during the reproductive years of life.

Prostate cancer is a disease common to elderly men, with more than 75% of cancers being diagnosed in men over the age of 65. In recent years, however, the incidence has increased in younger age groups.

In the earlier stages of the disease, prostate cancer rarely causes any specific symptoms. Later, it can present the following symptoms in the lower urinary tract:

- Difficulty passing urine
- Poor flow when passing urine
- Incomplete emptying of the bladder
- Increased frequency and urgency to pass urine
- Nocturia – excessive urination at night
- Ejaculatory problems
- Hematospermia -the presence of blood in semen-
- Erectile dysfunction

A locally advanced case of prostate cancer may obstruct the kidneys and cause renal failure. Prostate cancer usually spreads to the bones and may cause bone pain, anemia, or fractures.

For the above we have decided to do a program to help the doctor during the diagnosis to see if this tumor malignant or benign. Through the use of machine learning and image processing techniques.

1.0 Introduction

Prostate cancer is cancer that occurs in the prostate gland. A small, walnut-shaped gland found in males, it produces semen, which nourishes and transports sperm.

Prostate cancer is one of the most common types of cancer. Many types of prostate cancer grow slowly and are confined to the prostate gland and may not cause serious damage. However, while some types of prostate cancer grow slowly and may require little or no treatment, others are aggressive and can spread quickly.

Prostate cancer that is caught early, at a stage when it is still confined to the prostate gland, has the best chance of successful treatment.

In this project we try to help doctors in the process of diagnosing the tumor in order to improve the speed and accuracy of this important process. By building an intelligent model, the tumor's medical rumor can tell the doctor whether the tumor is malignant or benign.

1.1 Problem Statement

As we mentioned before, the process of diagnosis of tumor is one of the most important processes to treat cancer patients that require great accuracy.

For this purpose, we want to build an intelligent model where the model determines whether the tumor is benign or malignant based on the extraction of some characteristics of the medical image of the tumor and inserted into the model to determine the type of tumor.

1.2 Methodology

The methodology of the system involves extracting some features from the medical image of the tumor, nearly 10 properties depending on our data, and then introducing these features to intelligent learning model, then the smart learning model determine whether tumor benign or malignant.

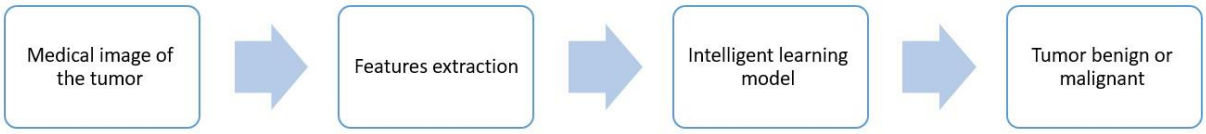


Figure 1: Process Flowchart

1.3 Use case diagram

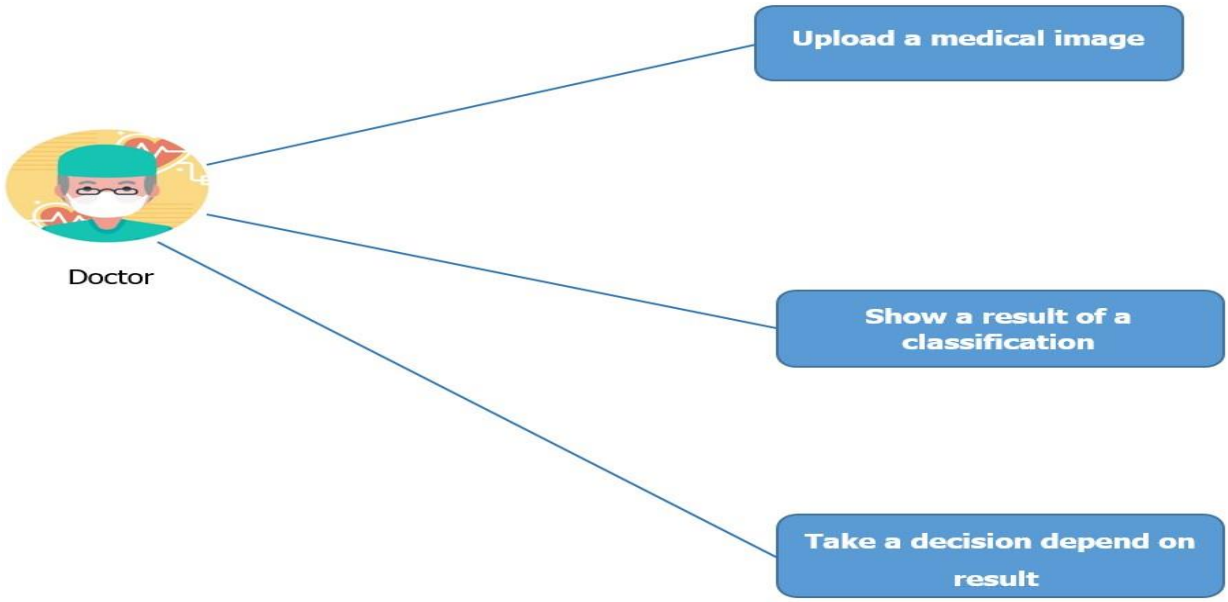


Figure 2: Use case diagram

1.4 System scenario

At first doctor must upload medical image of tumor then press on show result button. Then the system will take this image and extract features from it then enter this features value into intelligent model to classify a tumor as benign or malignant.

1.5 Class diagram

Prostate Cancer Medical Data

Attribute	Type
ID	Int
Radius	Float
Texture	Float
perimeter	Float
Area	Float
smoothness	Float
Compactness	Float
Diagnosis Result	Float
Symmetry	Float
Fractal Dimension	Float

1.6 Requirements

Functional Requirements

System is developed to diagnose Tumor using machine learning and image processing.

User Requirements

- Users are only Doctors.

Doctor Functions:

- Upload X-ray: The users containing the doctor should upload x-ray to the system / application to access the functions of the Application.
- See Evaluation: Doctors can see the evaluation that are made.

Non-Functional Requirements

1-Product Requirements:

These requirements specify or constrain the behavior of the software, include performance requirements on how fast the system must execute and how much memory it requires, reliability requirements that set out the acceptable failure rate, security requirements, and usability requirements.

1.1 Efficiency requirements:

System that we are going to develop need to be efficient by taking small space as possible and small response time.

1.1.1 Space requirement:

Our system is web application, so it works on any browser, hence we don't care about space in disk or Ram.

1.1.2Performance requirement:

Response time shouldn’t exceed 5 seconds and shouldn’t exceed 15% of the processors speed.

1.2 Usability requirements:

The system that we are going to develop should be easy to use by providing text area for user (doctor) to get the result.

1.3 Availability requirements:

The system that we are going to develop should be available 24 hours in the day and 7 days in week and should be available for all doctors in the hospital.

1.4 Interface Requirements:

The interface of the system that we are going to develop should have all the functionality of our systems and have shortcut to reach every function of the system.

2- Organization Requirements:

These requirements are broad system requirements derived from policies and procedures in the customer's and developer's organization.

2.1 Development Requirement:

In implementing the whole system:

- We will use Python programming language.
- We will use data Mining algorithms.
- We will use Image Processing Methods.

2.2 Environmental Requirement:

The system that we are going to develop will work on any browser.

2.3 Operational Requirements:

Specify how the system will be used, our system that we are going to develop will work to enable the doctors to give their decision about Tumor.

1.7 Flow diagram

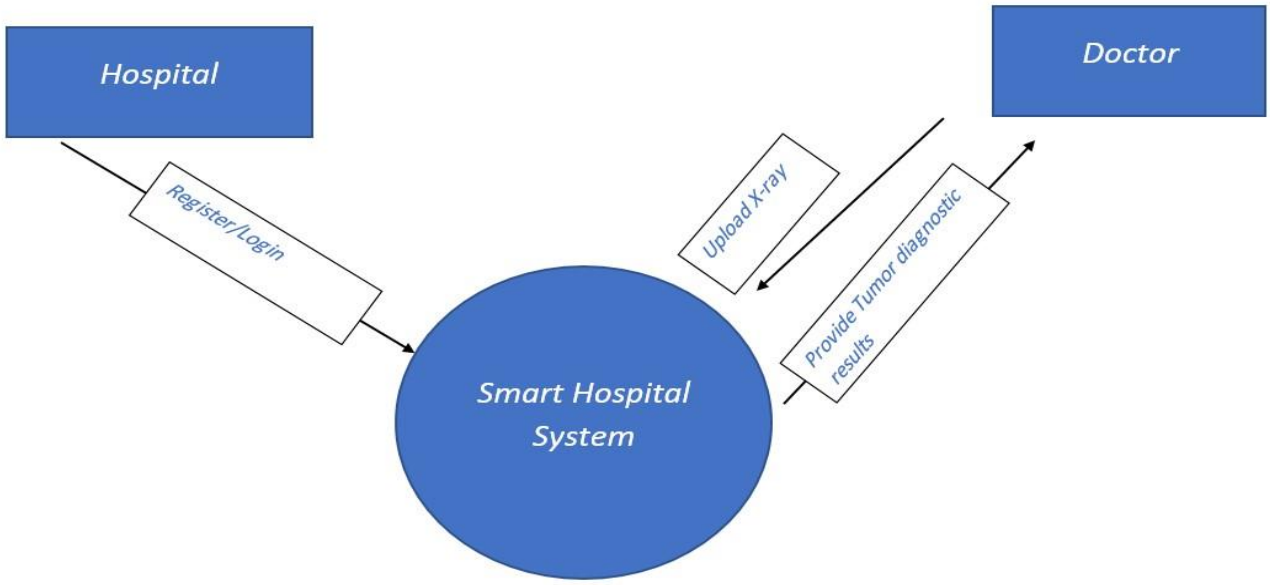


Figure 3: Flow diagram

2.0 Biological overview

Our human body consists of trillions of cells. Although each cell has the exact same DNA all over your body, every cell carries out its own function. The DNA is a long sequence of nucleotides preserved inside a cell nucleus.

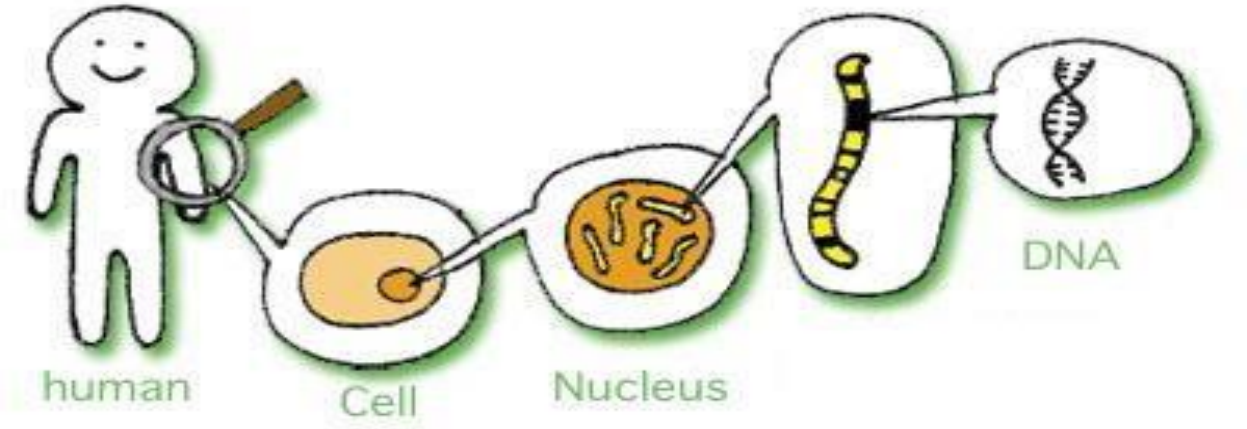


Figure 4: Photo adapted from: The Fatal Lover, Mata Hari (2016) watch online 1080

In our 3 billion nucleotides of the DNA, there are about 20,000 genes of different lengths. Suppose that the average length of a gene is 1000 nucleotides, then all genes will represent a length of 20,000,000 nucleotides; that is less than 1% of the total length of the DNA (3 billion).

At every single moment, inside our bodies, cells divide and replicate. Cell replication is also what we call growing up! As they substitute dead cells or just replicate to carry out different functions. When they do, they form new identical daughter cells. This replication starts with obtaining two identical copies of the DNA that each copy will be contained in the nucleus of a separate cell.

What is the problem?

During replication, some nucleotides might get mutated in the DNA copying process. In general, mutations are ok as the copy process occurs in microseconds, and it's expected that mutations happen with this very high-speed copying mechanism. If is occur a defect in the process of replication of cells without stopping leads to cancer.

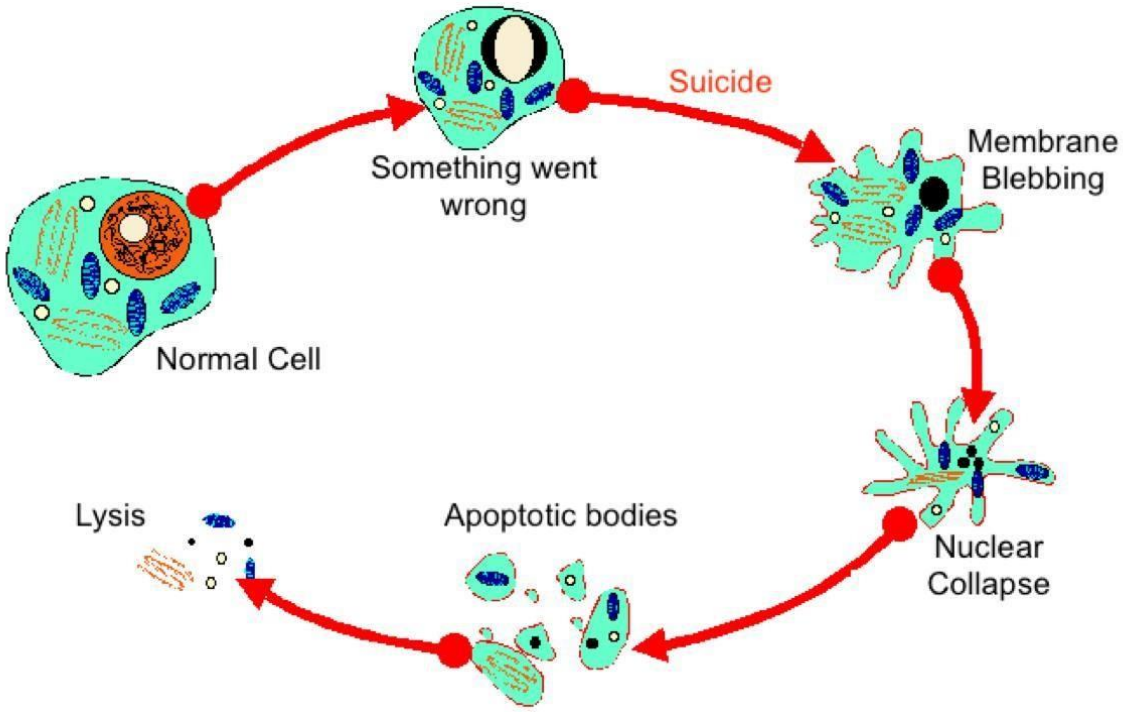


Figure 5: Photo adapted from Premed HQ program: https://www.premedhq.com/programmed-cell-death

We have mentioned the word cancer, prostate cancer, benign and malignant tumor repeatedly and repeatedly in the first part.

Don't worry we explain everything now.

Let's start what is cancer?

2.1 What is cancer?

Cancer is the name given to a collection of related diseases. In all types of cancer, some of the body's cells begin to divide without stopping and spread into surrounding tissues.

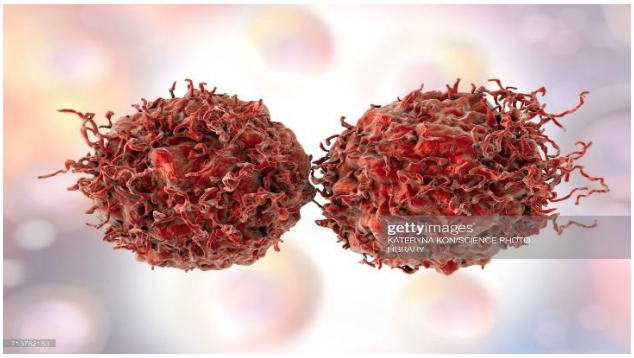


Figure 6: Prostate cancer cells, computer illustration.

Cancer can start almost anywhere in the human body, which is made up of trillions of cells. Normally, human cells grow and divide to form new cells as the body needs them. When cells grow old or become damaged, they die, and new cells take their place.

2.2 What is Prostate cancer?

Prostate cancer is cancer that occurs in the prostate gland. A small, walnut-shaped gland found in males, produces semen, which nourishes and transports sperm. Prostate cancer is one of the most common types of cancer. Many types of prostate cancer grow slowly are confined to the prostate gland and may not cause serious damage. However, while some types of prostate cancer grow slowly and may require little or no treatment, others are aggressive and can spread quickly.

2.3 Where prostate cancer starts?

Prostate cancer starts in the prostate gland. It starts when cells in the prostate grow out of control. Cancer cells can spread to other parts of the body.

Cancer cells in the prostate can sometimes travel to the bones or other organs and grow there.

2.4 Differences between Cancer Cells and Normal Cells

Cancer cells differ from normal cells in many ways that allow them to grow out of control and become invasive. One important difference is that cancer cells are less specialized than normal cells. That is, whereas normal cells mature into very distinct cell types with specific functions, cancer cells do not. This is one reason that, unlike normal cells, cancer cells continue to divide without stopping.

For this benign word is a metaphor for normal cells while malignant is a metaphor for cancer cells.

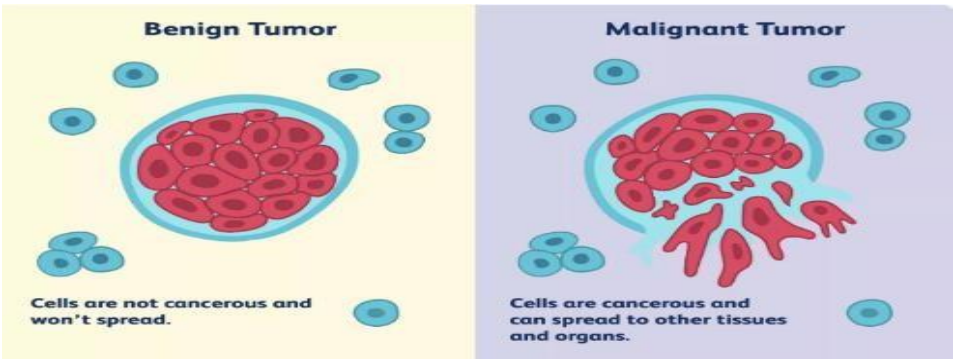


Figure 7: A picture showing difference between benign tumor and malignant tumor.

3.0 AI in cancer research

AI and Machine Learning is paving the way for the future and one of their largest and most useful applications is in the medical field. Research on curing and treating cancer has been a prominent focus for years. With the recent advancements of AI and Machine Learning, it will soon be possible to use them to make more insightful observations and create better treatment plans.

3.1 What is Artificial Intelligence (AI)?

AI is typically defined as the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, and even exercising creativity. There are applications for AI in almost any field from customer service to identifying malignant tumors.

3.2 What is Machine Learning (ML)?

Most recent advances in AI have been achieved by applying machine learning to very large data sets. Machine Learning allows computers to learn by themselves. With the power of modern computers, it is easier to use this to process large data sets. These algorithms detect patterns and learn how to make predictions and recommendations by processing data and experiences, rather than being explicitly written in program.

There are 3 major types of machine learning:

(1) Supervised: The data is labeled. The model must identify the labels and group accordingly. In other words, the model is provided with the input and is told the desired output. This process is done numerous times until the desired output is achieved.

(2) Unsupervised: The data is not labeled. The model has to identify different features and classify based on the distinct characteristics. Here, the input is given, but there is no desired output. The computer makes logical classifications or groupings.

(3) Reinforcement learning: The agent/computer changes its outcomes based on the reward. The reward is based on the feedback from the environment. This process is continuous as the agent/computer gathers more data.

3.3 Deep Learning

Deep learning is part of a broader family of machine learning methods based on a wider range of data resources that requires less data preprocessing by humans.

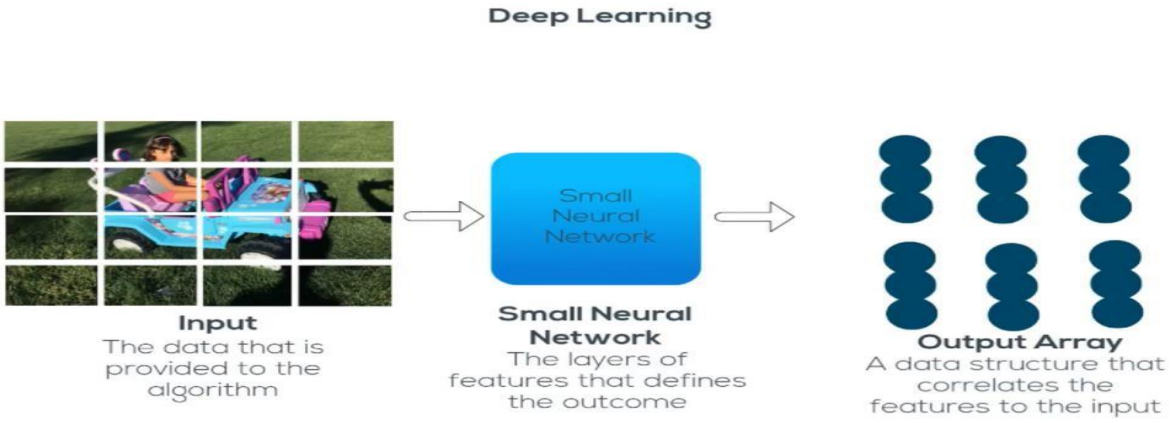


Figure 8: Deep Learning

5.0 Data description

(1) Number of total records: **100**.

(2) Number of total fields (columns): **10**.

(3) Number of Missing Value records: **0**.

(4) Number of Non-Missing value records: **100**.

Column Name	Description
ID	ID number
Radius	Distances from center to points on the perimeter.
Texture	Standard deviation of gray-scale values.
Perimeter	Size of the core tumor.
Area	Area of the core tumor.
Smoothness	Standard error local variation in radius lengths.
Compactness	Standard error perimeter^2 / area - 1.0
Diagnosis Result	The diagnosis of breast tissues • M = malignant • B = benign
Symmetry	No description available
Fractal Dimension	"Coastline approximation" - 1

5.1 Tools identification

(1) Openrefine tool

Open-source and free google tool to handle data quality issues.

(2) Python programming language

Open-source programming language for coding our models.

(3) Jupyter notebook platform

Platform to run python code on it.

(4) Spyder

Platform to run python code on it with more functionality.

Figure 9: Utilized tools

6.0 References:

https://www.ikca.org.uk/news/what-is-prostate-cancer/

https://www.cancer.gov/about-cancer/understanding/what-is-cancer

https://blog.xrds.acm.org/2016/10/cancer-biology-computer-science/

https://www.verywellhealth.com/what-does-malignant-and-benign-mean-514240

https://blog.goodaudience.com/machine-learning-in-cancer-research-75cf4d7ee29a

7.0 Conclusion statement

In conclusion we would like to say that there are a lot of things that we have not identified which ones we will use in the project as hyperparameters values, optimization algorithm, and the model that we will use it because all these things are depend on try and enhance during implementation phase but, we will roughly choose all these things at the end of project. Finally, we would like to thank **Dr. Mustafa Abu Bakr** and **Eng. Marwa Ahmed** to assist us in this project.