Yarmouk University
Hijjawi Faculty for Engineering Technology
Department of Computer Engineering

Graduation Project Report

Lung Cancer Detection Using Deep learning

Heba Abu Shareefeh 2019980186

Rudaina Alyaseen 2019980261

Mohammad Mrayyan 2019980212

Dr. Mahmoud Masadeh

Semester: First 2023/2024

Date: 10th January 2024

Students' Property Right Declaration and Anti-Plagiarism Statement

We hereby declare that the work in this graduation project at Yarmouk University is our own except for quotations and summaries which have been duly acknowledged. This work has not been accepted for any degree and is not concurrently submitted for award of other degrees. It is the sole property of Yarmouk University and it is protected under the intellectual property right laws and conventions.

We hereby declare that this report is our own work except from properly referenced quotations and contains no plagiarism.

We have read and understood the school's rules on assessment offences, which are available at Yarmouk University Handbook.

Students:

Name: Heba Abu Shareefeh	Signature: HA	Date: 6/20/2024
Name: Rudaina Alyaseen	Signature: RA	Date: 6/20/2024
Name: Mohammad Mrayyan	Signature: MM	Date:6/20/2024

Table of Contents

Students' Property Right Declaration and Anti-Plagiarism Statement....1
List of Tables....3
List of Figures....4
Abstract....6
Chapter 1: Introduction....7
Chapter 2: Background....11
Chapter 3: Design....13
Chapter 4: Implementation....9
Chapter 5: Results and Discussion....15
Chapter 6: Economical, Ethic, and Contemporary Issues....21
Chapter 7: Project Management....22
Chapter 8: Conclusion and Future Work....24
References....25
APPENDIX A: User Manual....25

List of Tables

Table 1. datasets used for lung cancer. 8
Table 2. Performance Metrics Description. 9
Table 3. Design considerations 7,8

List of Figures

Figure 1. CT scan based on intensity projection. [3] 7

Figure 2. Confusion Matrix [9] 9

Figure 3. Schematic diagram for lung cancer detection.... 11

Figure 4. cancerous image. 14

Figure 5. non-cancerous image. 1

Figure 6. use case diagram. 2

Figure 7 ANN visualization 3

Figure 8 CNN visualization 3

Figure 9: CNN architecture.... 4

Figure 10: CNN visualization 6

Figure 11: Libraries used 9

Figure 12 :Jupyter Icon……9

Figure 13 :VS code ……9

Figure 14:before one class classification ……..11

Figure 15:after one class classification ……11

Figure 16: shape distribution before realizing ……12

Figure 17:k_fold validation ……12

Figure18:model structure……13

Figure19:our CM result ……13

Figure 20:related paper result ……13

Figure 21:home,about and features sections……16

Figure 22:FAQ and contact sections……17

Figure 23:log in page……...18

Figure 24:sign up page……18

Figure 25:result page ……...19

Figure 26:Responsive Web App……19

Figure 27:pert chart ……...22

Figure 28:gantt chart……....23

Abstract

As advancements in technology continue in healthcare, detecting cancer, especially lung cancer, presents high challenges because it's complex and there's a huge amount of data from medical images and patient records to manage, making detection both challenging and time-consuming. Moreover, Sharing of medical data among healthcare professionals can greatly enhance diagnostic accuracy and treatment outcomes However, existing methods of data sharing often lack efficiency and security. Therefore, our project aims to build a model using AI algorithms to provide an effective solution for enhancing diagnostic accuracy, enabling early detection of lung cancer, and use web techniques that make our model easier to use.

Keywords— Lung cancer, Deep Learning, convolutional neural networks (CNN)

Chapter 1: Introduction

1.1 problem statement:

Globally, lung cancer is the most commonly diagnosed cancer over the past few decades [1]. Initially, cancer cells are tiny and hard to spot, but after a while, they enlarge and become more severe. For this reason, a key factor in enhancing patient outcomes is early detection.

To detect cancer, a radiologist must perform a large number of CT scans. However, because lung nodules resemble surrounding structures, such as blood vessels, it can be challenging to distinguish between blood vessels and cancer nodules in the early stages of the disease so this can be time-consuming and prone to human error. In Figure 1, three CT images are shown: The top two show lungs containing early-stage cancer; the first being malignant and the second being benign. It is difficult to differentiate between vessel and cancer nodules during this stage. The last image is of cancer in its late stage, with a large nodule size making it easier to detect, but survival rates are low [2].

Therefore, our project focuses on developing a model to early detecting lung cancer from CT images.

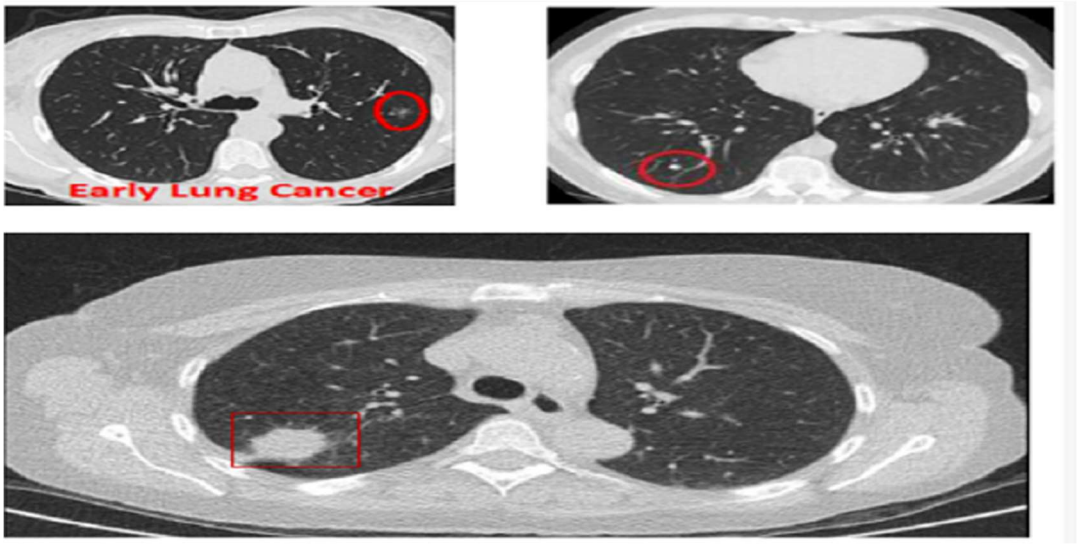


Figure 1. CT scan based on intensity projection. [3]

1.2 Background.

Doctors face difficulty in detecting lung cancer in its early stages due to its complexities and the large amount of data it requires. This can lead to the disease progressing to advanced stages.

To address these challenges, this project aims to train a model using DL algorithms on a dataset. The model will be able to predict the probability of having the disease in its early stages, enabling early treatment.

The table below show some of the datasets that can be used to train model.

Table 1. datasets used for lung cancer.

Datasets	**About data set**
Chest CT-Scan images Dataset [5]	**Data** folder consists of train , test and valid folders each folder contain 3 folders of different chest cancer types(adenocarcinoma,large cell carcinoma,squamous cell carcinoma) and 1 folder of normal CT-Scan images (normal) **train** folder contain the training images (adenocarcinoma 195 files, large cell carcinoma 115 files, normal 148 files, squamous cell carcinoma 155 files) **test** folder contain the testing images (adenocarcinoma 120 files, large cell carcinoma 51 files, normal 54 files, squamous cell carcinoma 90 files) **valid** folder contain the validating images(adenocarcinoma 23 files, large cell carcinoma 21 files, normal 13 files, squamous cell carcinoma 15 files)
CT-Scan images [6]	Images were collected from the hospital situated in Iran. Part of this CT-scan images of lungs were belonged to lung cancer patients and classified as cancerous images, and the rest of them were belong to other lung diseases, for instance patients who caught COVID-19, and classified as non-cancerous images. The total number of CT-scan images, which were used in this paper is equal to 364 that 238 of them belong to cancerous images and 126 of the rest belong to non-cancerous images. All of each these images were collected with the help of a pulmonologist in order to skip any probable error in classifying these images.
The IQ-OTH/NCCD lung cancer dataset [7]	The IQ-OTH/NCCD lung cancer dataset Lung Cancer CT Scans from Iraqi hospitals: (Normal 416 files , Benign120 files , and Malignant Cases 561 files)
CheXpert: Chest X-rays [8]	CheXpert is a dataset consisting of 224,316 chest radiographs of 65,240 patients who underwent a radiographic examination from Stanford University Medical Center between October 2002 and July 2017, in both inpatient and outpatient centers. The CheXpert dataset includes train, validation, and test sets. The validation and test sets include labels obtained by board-certified radiologists. The train set includes three sets of labels automatically extracted from associated radiology reports using various automated labelers (CheXpert, CheXbert, and VisualCheXbert).

1.3 Aims and objectives:

The project aims to develop an accurate disease prediction system for lung cancer using convolutional neural networks (CNNs) objectives of the project are:

- analyze medical CT scan images with high accuracy, enabling the early detection of lung cancer and quickly treatment early for patients.
- creating a centralized platform for healthcare professionals to upload images.

1.4 Evaluation of solution

For the evaluation phase, we will use confusion matrix and compute the accuracy, precision, recall, and specificity. Then we'll compare our results with related paper.

For the performance Confusion Matrix was used. **Figure 2** and **Table 2** describe them.

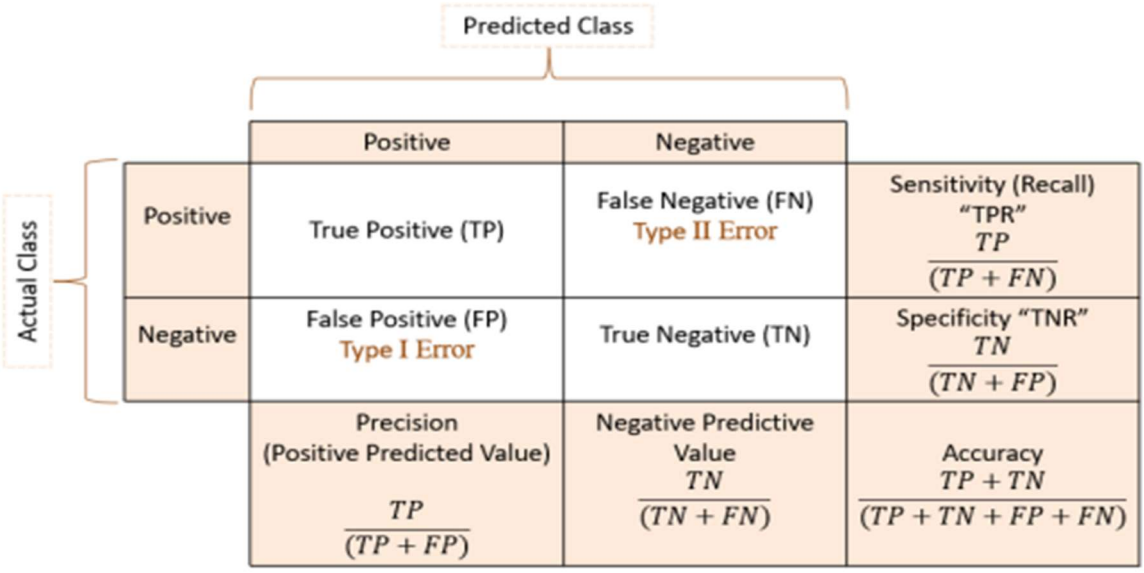


Figure 2. Confusion Matrix [9]

Table 2. Performance Metrics Description.

Performance Measure	**Description**
Sensitivity(recall)	actual positive cases that the model correctly identifies
Specificity	specificity quantifies the model's ability to correctly identify negative cases
Precision	reflects the accuracy of the model's positive predictions

1.5 List of contributions:

A list of contributions with short descriptions

The contributions of the project include:

- User-friendly front-end interface: the system features easy to use user interface, designed with modern web technologies.
- improving the accuracy of lung cancer prediction: by using efficient algorithms and calculate the accuracy.

1.6 High Level Figure of Our workflow:

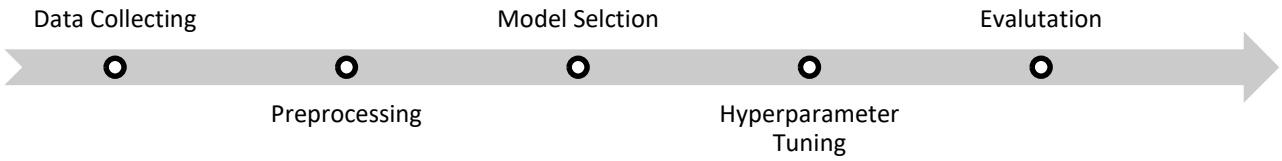


Figure 3. Schematic diagram for lung cancer detection.

Figure 3 describes the steps that we'll follow to build the model.

1.7 Summary of report structure.

We will go over every facet of the project's conception, creation, and execution in this report. The problem and a suggested solution are briefly described in the abstract. Further information about the design is covered in the introduction. The difficulty and the significance of resolving it are covered in the background section. The architecture and functionality of the system are described in the design section.

We give a thorough explanation of the technologies utilized in the system's construction as well as the methodical steps involved in its development in the implementation section. The evaluation's conclusions are presented in the results and discussion section, which also analyzes the system's shortcomings and future development opportunities. A Gantt chart and a Pert diagram show the steps taken to finish the project in the section on project management. Finally, we provide the references used throughout the report.

Chapter 2: Background

2.1 Background of the problem:

The problem is that traditional methods of diagnosing lung cancer have become ineffective because they rely on common radiographs, which may not be very accurate in identifying cancer in its early stages. Small tumors can be missed, causing a delay in their detection. Traditional methods are expensive and take a long time, so we have developed a model that predicts lung cancer based on the image, with high accuracy and a very short time.

2.2 Importance

With the development of technology, it has become important to use modern methods in detecting lung cancer, because it reduces time and effort and reduces the cost. It reduces human errors that may occur due to tiredness or lack of experience. It provides early and accurate diagnosis. Early diagnosis contributes to reducing the risk of lung cancer. The mortality rate in this disease can also reduce the workload on medical staff because they receive the initial diagnosis from the prediction model, and thus they can focus their efforts on cases that require more effort.

2.3 Target the market and their needs.

Radiologists, healthcare providers, and other medical professionals involved in lung cancer diagnostics comprise the core target market. Their main requirement is for a more accurate and efficient diagnostic tool that can operate smoothly with current workflows to identify lung cancer from CT scans in a timely and trustworthy manner.

2.4 Potential ethical and/or environmental issues.

Developing a lung cancer prediction model using deep learning preserves the privacy of the patient's data and does not use it for research and development purposes except with his permission. As for the environmental aspect, it reduces the waste of X-rays by replacing them with CT images that we upload to our website and display the results.

2.5 Summarize the different approaches currently/previously used to solve the problem.

- Traditional Methods:
  - It depends on the skill and experience of the diagnosis.
  - Relatively slow
  - Human errors may occur
  - Less effective in detection during the early stages
  - It may be expensive sometimes
  - It does not develop over time, unlike deep learning techniques, which keep pace with development and improvement

- Deep learning techniques:
  - High accuracy
  - Fast and effective
  - It works to reduce human errors
  - Detects the disease in the early stages

Chapter 3: Design

3.1 Design Overview:

3.1.1 Design Description

The project's goal is to construct a predictive model for estimating the probability of lung cancer. To enhance accessibility, we implement a user-friendly web platform. So the project comprises two primary components: the web interface and the predictive model. In the predictive model, we employ a deep learning algorithm known as Convolutional Neural Networks (CNNs).

The web interface serves is the gateway for users to interact with the predictive model. where users can input relevant data and receive predictions regarding lung cancer probability.

We have developed a simple, responsive web application that is compatible with all devices, including tablets, phones, and desktops. The application features essential pages such as Login, Sign Up, About, Home, FAQ, and a Result page. On the Result page, users can upload an image to receive predictions. We utilized HTML, CSS, and JavaScript to create the structure and interactivity of the web pages, and incorporated Bootstrap to ensure responsiveness. The backend is connected to the frontend using Flask, a micro web framework that is well-suited for integrating predictive models with web applications. This setup allows for seamless communication between the user interface and the predictive model, facilitating accurate lung cancer predictions.

3.1.2 plan to address the problem statement

Figure 4. cancerous image.

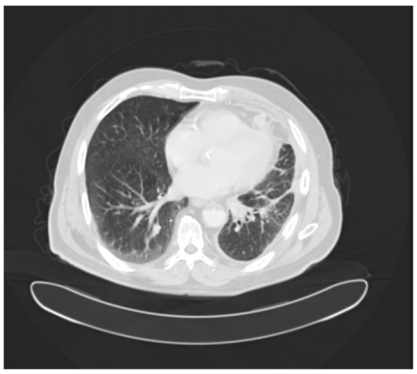


Figure 5. non-cancerous image.

1. Next, we'll preprocess the dataset, which involves tasks such as resizing images, handling imbalanced data, and partitioning the dataset into training and testing sets.
2. Constructing the model with CNNs
3. Training the model on the training set.
4. predicting the testing set and evaluating the train and test models using performance metrics such as accuracy, precision, recall, or area under the curve (AUC). This assessment aids in gauging the model's effectiveness in predicting lung cancer.
5. Develop the App Interface: designed in a way that makes it easy for users to input data, receive results
6. Integrate AI Model into the App : Integrate the trained AI model into our app's backend infrastructure. This involves implementing the necessary APIs to facilitate communication between the app's front end and the AI model.

3.1.3 Detailed Figure.

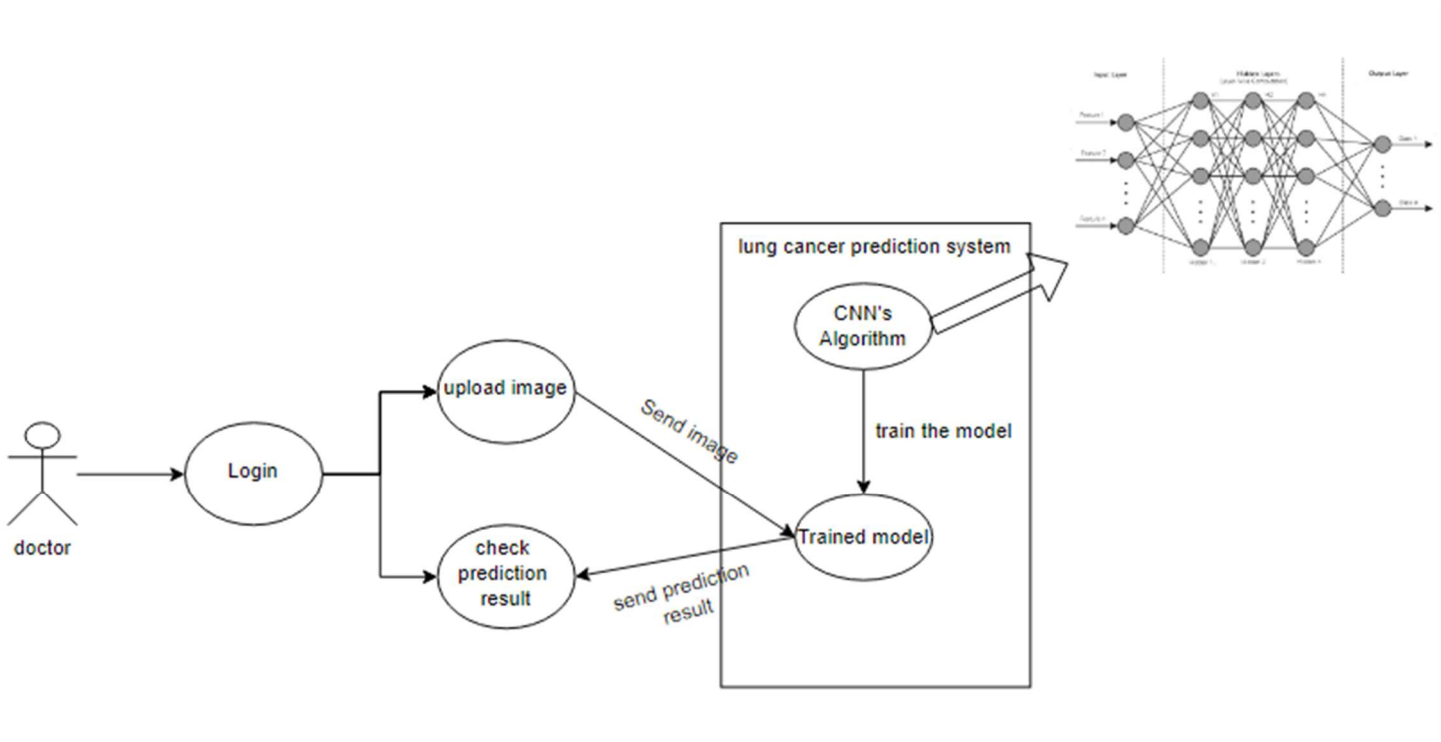


Figure 6. use case diagram.

Figure 6 shows the use-case diagram of the system. It illustrates the interactions and processes within the system

Scenarios where the end user can use the system:

- the doctor submits a CT scan image for assessment. Behind the scenes, the system's backend employs advanced machine learning algorithms to meticulously analyze the uploaded image. Once the analysis is complete, the system promptly delivers an accurate diagnosis to the doctor, empowering them with timely insights for effective patient care

3.2 Design Details:

3.2.1 Design Specifications

- Design Dimensions:
  The IQ-OTHNCCD lung cancer dataset was obtained from Kaggle [here] (https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset) , It consists of 1097 samples of lung cancer patients, covering cases categorized as Normal, Benign, and Malignant
- Quality:
  In the preprocessing stage, our objective is to implement strategies that help us avoid overfitting the data. For example, we aim to standardize image sizes across datasets and balance distributions to prevent any single class from dominating excessively.
- Safety:

Given the objective of identifying patients with cancer, mitigating False Negative (FN) errors is paramount. False Negative errors occur when the model incorrectly predicts that a patient does not have cancer when they actually do. So in our project, our primary focus is on minimizing False Negative (FN) errors.

- Environmental Factors:
  We aim to create a project that is compatible across multiple devices, and thus, we incorporate web technology into our project.

3.2.2 Design Process

In this section, we focus on the method that we have utilized, as they represent the main concept of our work.

The methods that we used is Convolutional Neural Networks:

- **Why CNNs**?

If we have 64*64 image, When using another method like ANN, each pixel is considered a separate feature, leading to a large number of features (4096 pixels in this case), which can make the system fail as the image size increases. On the other hand, CNN's Convolutional Layer performs feature selection, identifying important pixels to train the model effectively, helping in extracting distinctive patterns from the image. This makes CNN a better choice.

The images below show the significance of the convolutional layer:

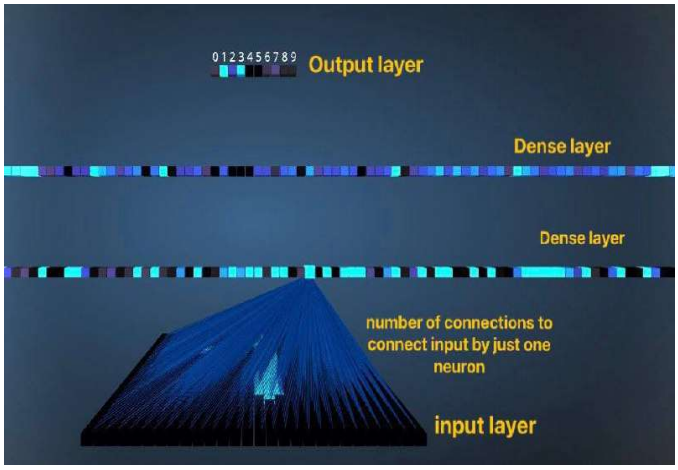


Figure 7 ANN visualization

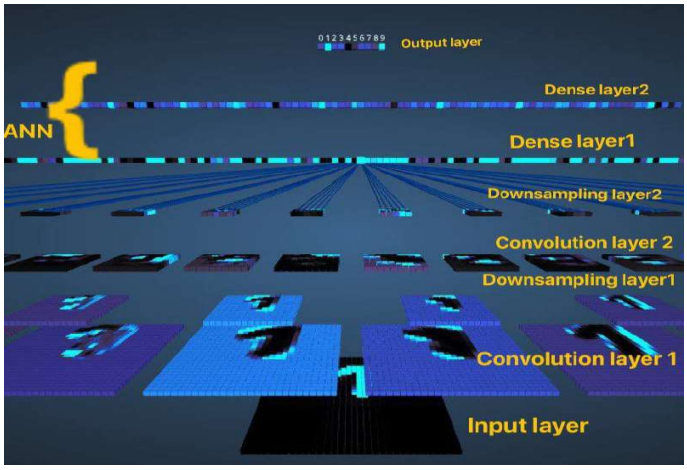


Figure 8 CNN visualization

The image 10 show the visualization of a simple image when using an Artificial Neural Network (ANN). It shows that the ANN consists of fully connected layers, where the number of connections between the input layer and each neuron is significant (In the image, the number of connections shown is only for a single neuron in the first layer and there are same connection for each neuron in same layer and additional

connections in the second layer). Therefore, if the image size is large, the ANN might fail due to the extensive number of connections, impacting its performance.

This is where the convolutional layer in CNN comes into play, as its role is to retain only the important pixels through feature selection, as show in figure 11

- What CNN?
  Convolutional Neural Network (CNN) is a type of neural network Architecture that is widely used for performing deep learning on image data.
  CNN consists of an input layer, a hidden layer(s), and an output layer. differentiates CNNs from ordinary ANNs is that the hidden layers of a CNN consist of a special series of layers called the convolutional and pooling layers.

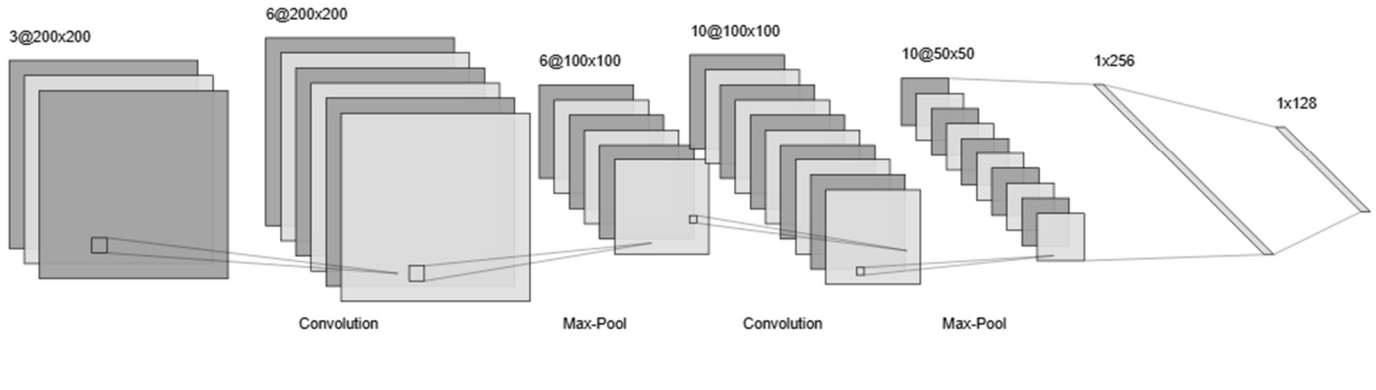


Figure 9: CNN architecture

- **Input layer:**

The input layer represents the initial stage which accepts input data in the form of multi-dimensional arrays (tensors).For example, if you have grayscale images of size 200x200 pixels, the input tensor would be of shape (200, 200, 1) because it has one channel. If you have RGB images of the same size, the input tensor would be of shape (200, 200, 3) because it has three channels

.

- **hidden layer:**

In a Convolutional Neural Network (CNN), the hidden layers consist of convolutional layers, pooling layers, and fully connected layers.

- **convolutional layers:** these layers apply convolution operations to the input data. Each convolutional layer consists of multiple filters extracting features**.**
  The number of filters in a convolutional layer determines the depth of the output feature maps. In figure 9, with an input layer of dimensions 200×200×3 and a convolutional layer consist 6 filters, the output feature's shape would be 200×200×6.

- **Pooling layer or Downsampling layer:** reduce the dimensionality of feature maps generated by convolutional layers This downsampling process retains essential information while reducing the spatial dimensions**.** In Figure 9, applying max pooling with a filter size of 2×2 will halve the dimensionality of the feature maps. So the shape become 100×100×6

  Note the shape after applying 2 Convolutional layers and 2 Pooling layers 50×50×10 about 25000 pixels and the shape of input images 200×200×3about 120000 pixels

- **Fully connected layer(Dense layer):**

  - **Flatten layer:** After the convolutional and pooling layers, the feature maps are flattened into a one-dimensional vector

    For example, if the shape of the feature maps 50×50×10 it maps to 25000×1

  - **Fully connected**: process this flattened data by connecting every neuron in one layer to every neuron in the next layer

❖ **Activation Functions in Hidden and Output Layers:**

- **ReLU**
  utilized between layers when we aim for the model to only predict positive values. It retains positive values as they are and converts negative values to zero.
  In our project we handle with image and the pixel must be positive so we will use it in hidden layer
- **Softmax**
  commonly used in the output layer of neural networks for multiclass classification and the output values represent the probability of each class
- **Sigmoid**
  Used in hidden layer as scaller and in output layer in binary classification as probability Our project binary classification so we will use sigmoid in output layer

❖ **Image below shows the visualization of CNN:**

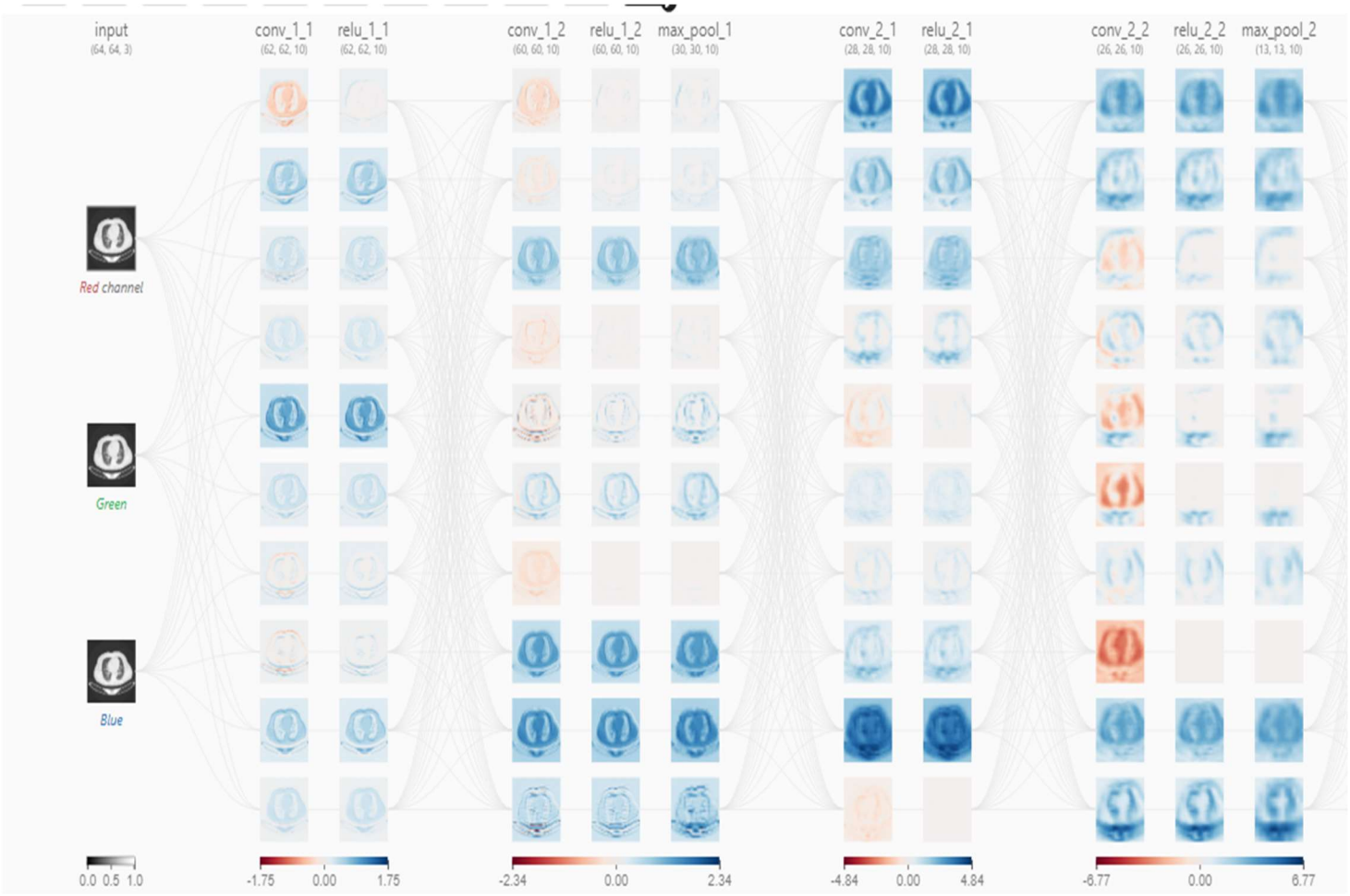


Figure 10: CNN visualization

3.2.3 Legal Aspects

Ensure that the system maintains the patient's privacy and maintains the patient's medical history and private data. The intellectual property rights of the original owner of the idea and his research must be preserved.

3.2.4 Design Constraints

Accessible, Ease of use , Responsive, Data Privacy and Secure.

3.2.5 Design Standards

Compatible with various browsers and very good performance .

3.2.6 Design Alternatives

The design can be modified to allow users to use it from the phone with ease.

3.2.7 Safety Consideration

- such as data backup and recovery, periodic test, secure design, and regular update.

3.2.8 Design considerations table.

Table 3. Design considerations.

Design consideration	Project application	Relevant location in report
Performance	Responsive, fast receiving images and displaying data.	Section 3.2.4 Design Constraints Section 3.2.5 Design Standards
Serviceability	User Support ,Data privacy	Section 3.1.4 Section 3.2.2
Economic	N/A	N/A
Environmental	The system compatible with the specific platform it will be running on	Section 3.2.1 Design Specification
Environmental Sustainability	Using efficient algorithms and data structures can help reduce the computational	Section 3.2.1 Design Specification

	resources required by the application	
Manufacturability	N/A	N/A
Ethical	System provides data protection protocols	Section 3.2.4
Health and safety	Ensuring that sensitive user data is protected from unauthorized access	Section 3.2.7 Safety Consideration
Social	User-friendly interface for users	Section 3.2.4 Design Constraints
Political	Compliance with healthcare data protection laws	Section 3.2.3 Legal Aspects

Chapter 4: Implementation

4.1 Implementation Details of Our Solution: Methods and Tools:

For trained model We used various tools to implement our solution of predicting lung cancer. We used Python programming language and dealt with multiple libraries, such as Scikit-Learn, Tensorflow, OpenCV ,Numpy, and Matplotlib.

Scikit-learn was used for data preprocessing, model evaluation, and performance metrics calculation. The `train_test_split` function from `sklearn.model_selection` was employed to split datasets for training and testing, Categorical target label was encoded using the `LabelEncoder` from `sklearn.preprocessing` and performance evaluation metrics such as confusion matrices and precision score, recall, and F1 were calculated from `sklearn.metrics`.

OpenCV was used to handle with images .Numpy was used to help us to work with the numbers and arrays.TensorFlow and Keras were utilized to construct and train the model. Additionally, we utilized visualization library Matplotlib for data visualization.

Figure 11: Libraries used

VS code and jupyter notebook were used to write all codes as it provide excellent code completion, which is particularly useful when dealing with large and complex AI projects. It also has a powerful debugging tool allowing us to quickly review our code and find errors.

Figure 12 Jupyter Icon

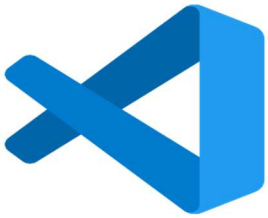


Figure 13 VS

For the web, we used HTML, CSS, and JavaScript. We also used frameworks such as Bootstrap and Flask.

- HTML: to provide structure of a web page, HTML tags allow developers to specify the type of content being displayed and how it should be presented.
- CSS: is used to defines styles for our web pages.
- JS: to make web pages interactive, we use JS to add animations that makes our website more attractive.
- Bootstrap: Framework makes front-end development faster and easier, making it more attractive, and help us to build completely responsive website.
- FLASK: is a small and lightweight Python web framework, We used it to connect the user interface to the prediction model.

4.2 Solution Infrastructure

The data set was downloaded from Kaggle , It consists of 1097 samples of lung cancer patients, covering cases categorized as Normal, Benign, and Malignant but when we examined the class distribution, we observed imbalance. To address this, we used several method like One Class Classification, Undersamlpling , Data augmentation.

- **Over-sampling:**
  duplicate random records from the minority class.
- **Under-sampling:**
  the simplest technique involves removing random records from the majority class
- **One-Class Classification:**
  One-class classification involves training a model on only one class and then using it to identify data points that do not belong to that class
- **Data augmentation:**
  Using transformation techniques like rotation, reflection and flipping to generate new images.

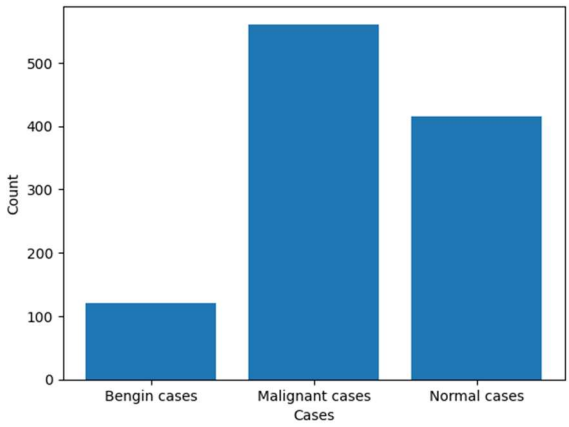


Figure 14 Before One-Class Classification

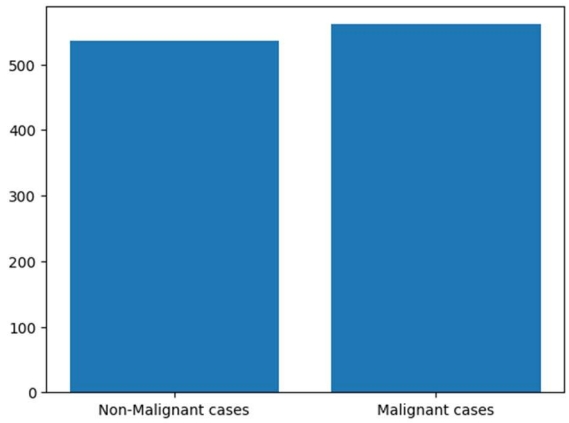


Figure 15 After One-Class Classification

The figure 14 show that the malignant cases significantly outnumber the others, which can lead to biased models favoring malignant cases and reduced overall performance. And figure 15 shows the distribution after using one class-classification .The training data consist of 70% of the total data, with the remaining 30% being used as test sets.

Now if we looked at the shape of the images, we noticed that there are some images with different shapes, so we need to resize all images to the same shape.

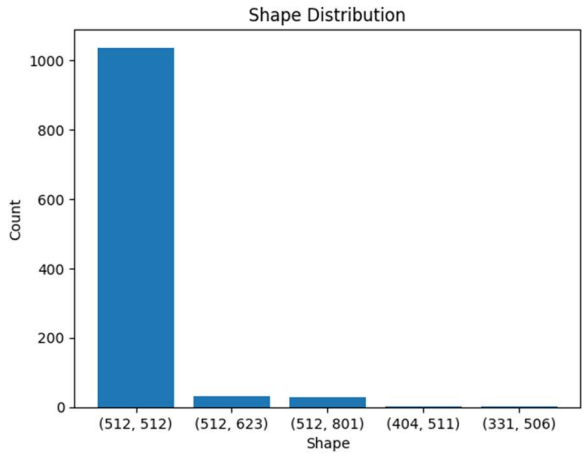


Figure 16 Shape Distribution Before Resizing

Hyper-parameter Tuning: We used grid search technique to find the optimal parameters for the model using GridSearchCV from scikit-learn library. CV stands for K-fold Cross Validation which split the training sets to k folds randomly and each time use one from these folds as testing set and the rest of them as training to find the best accuracy.

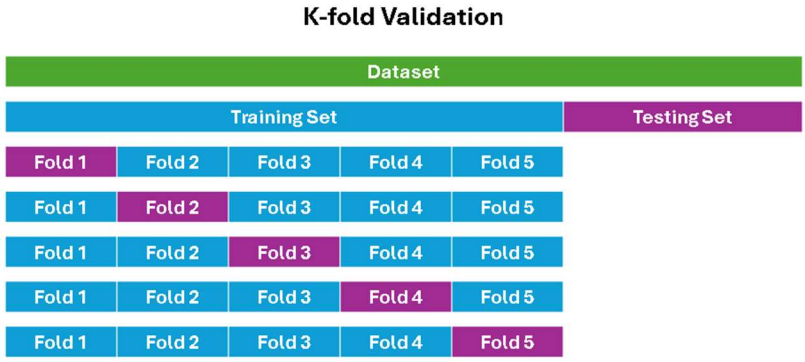


Figure 17 K-Fold Validation

For model creating we used CNN because we're working with images and the figure below shows the details of our model.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 126, 126, 64)	640
max2d_1 (MaxPooling2D)	(None, 63, 63, 64)	0
conv2d_2 (Conv2D)	(None, 61, 61, 32)	18,464
max2d_2 (MaxPooling2D)	(None, 30, 30, 32)	0
conv2d_3 (Conv2D)	(None, 28, 28, 8)	2,312
max2d_3 (MaxPooling2D)	(None, 14, 14, 8)	0
flatten (Flatten)	(None, 1568)	0
out (Dense)	(None, 1)	1,569

Total params: 22,985 (89.79 KB)
Trainable params: 22,985 (89.79 KB)
Non-trainable params: 0 (0.00 B)

Figure 18 Model Structure

Finaly, we compared our result with related paper [9]

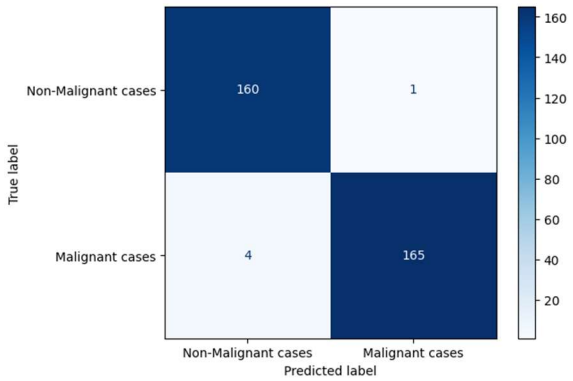


Figure 19 Our CM Result

Table 2- The confusion matrix

Confusion matrix		Predicted class	
		Non-malignant (positive)	Malignant (negative)
Actual class	Non-malignant	67 (TP)	3 (FN)
	Malignant	2 (FP)	38 (TN)

Figure 20 Related Paper Result [9]

Our model has accuracy about 99% and they have 93.548%
And these some reasons that may affect the results:

- They used AlexNet Architecture whish is may not be the best choice for this problem.
- They used some filters that may not be the best choice for this problem. (e.g. the filters may be good for some images and not good for the rest of them) and In our work we don't use any filters because we used CNN which is need lower pre-processing (CNN works to find the best filters).
- May they used less data for training than we used.

4.3 The trade-offs

Designing deep learning model involves balancing between accuracy and complexity, complex model may have high accuracy but require more computational time but in medical model we need high accuracy, and balancing between specificity and sensitivity, in our model we focused on the sensitivity because we need to reduce the risk of missing Malignant cases (reduce FN/ increase TP)

4.4 Assumptions of our implementation.

Our solution assumes that the prediction model can generalize well to unseen data, i.e., it can predict lung cancer for new CT scan image that does not present in the training data. The quality of the input data is crucial for the accuracy of our prediction mode

Chapter 5: Results and Discussion

5.1 Result and Details

We selected a dataset containing cross-sectional images of the lungs with annotation of the cancerous and non-cancerous parts. We then resized and processed the images to obtain accurate results. Then we built the model using CNN. We trained the model on the processed image set to ensure reliable performance. Next, predict the test set and evaluate the training and test models using performance metrics such as precision, precision, recall, or area under the curve (AUC). This evaluation helps measure the effectiveness of the model in predicting lung cancer. All these steps were performed on the back and front side. As for the front end, we designed a simple user interface to upload the image and display the result. The result included the type of cancer and prediction accuracy. Finally, we connected the backend and frontend using Flask.

Here are some images showing the web interface design:

Figure **21 :**Home, about and features sections

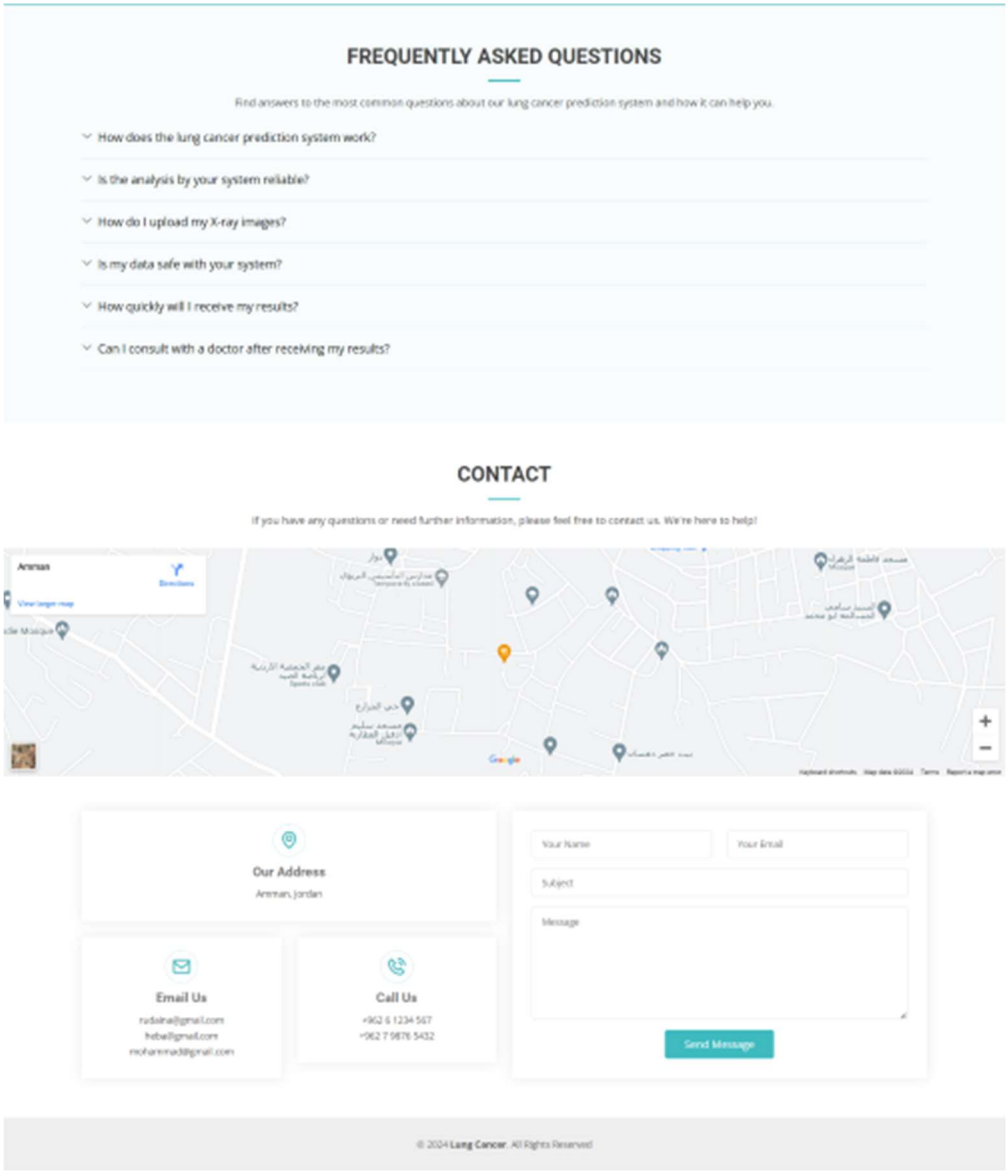


Figure **22:**FAQ and contact sections

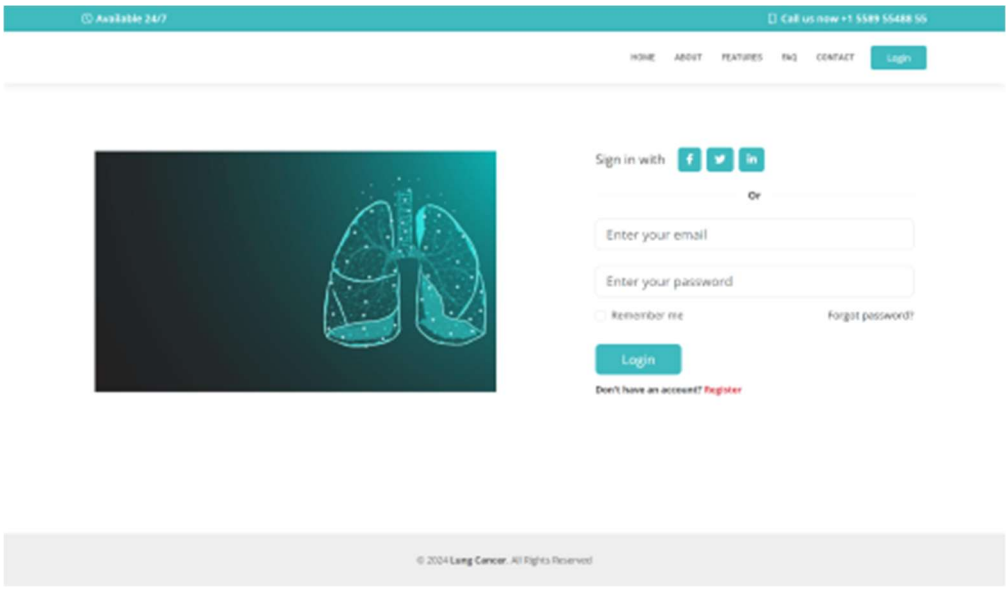


Figure **23:**Log in page

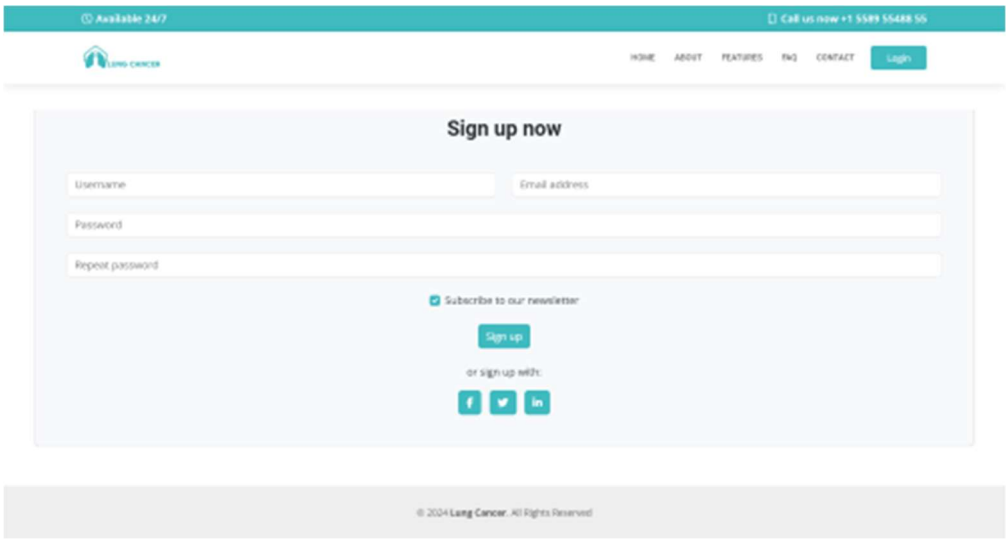


Figure **24:**Sign up page

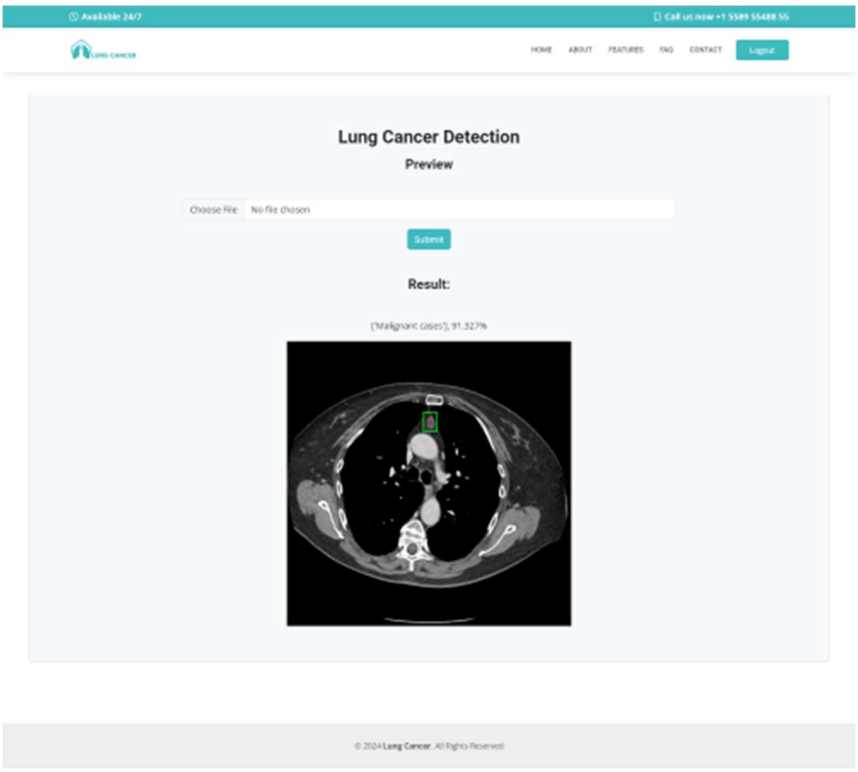
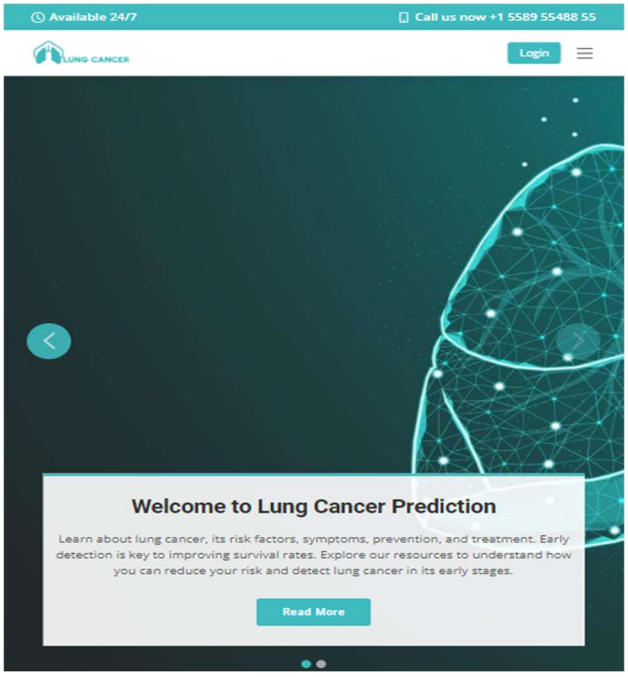


Figure **25:**Result page

The images below show that the Responsive Web application is compatible with all devices:



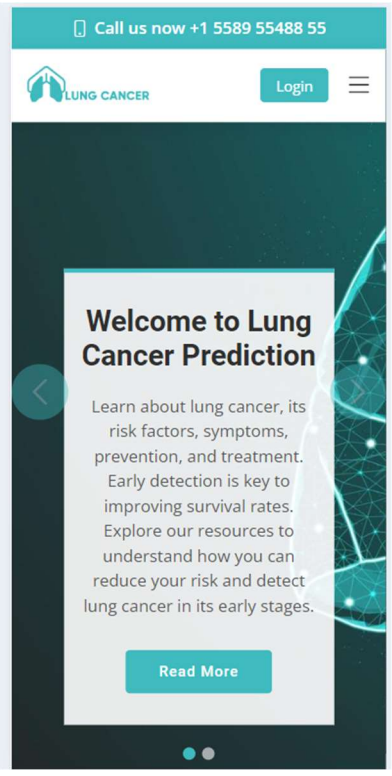


Figure **26:**Responsive Web application

5.2 Discuss the strengths and weaknesses of your solution/system.

Strengths:

- The system works non-stop 24 hours and doesn't need to sleep or even take rest like humans (they have a limit to their capabilities).
- Responsive, works on all devices, whether tablet, phone or desktop.
- It predicts the disease with very high accuracy.

Weakness:

- It is possible that the quality of CT images may sometimes affect the accuracy of the model.

Chapter 6: Economical, Ethic, and Contemporary Issues

6.1 Preliminary Cost Estimation and Justification

The development of the project is undertaken with no financial costs as it uses open-source technologies and free development tools. The main tools to use in the development process include VS Code with the Jupyter extension to trained and built a model and HTML, CSS, JavaScript, bootstrap for front-end ,flask to integrate between front end and java script, python for back end.

6.2 Relevant Codes of Ethics and Moral Frameworks

We have adhered to the guidelines set forth by the IEEE Global Initiative for Ethical Considerations in AI and Autonomous Systems in our research project. Launched by the Institute of Electrical and Electronics Engineers (IEEE) in 2016, this initiative aims to develop comprehensive guidelines and recommendations for the ethical design and deployment of AI and autonomous systems. By following these guidelines, we strive to ensure that our research contributes to responsible and ethical practices in AI, aligns with human values, and promotes the public good

6.3 Ethical Dilemmas and Justification of Proposed Solution

- Dependency on technology can lead to doctors relying too heavily on the CNN-based system, potentially neglecting their own expertise and observations. Emphasizing the role of the CNN as a tool to support decision-making rather than replace human expertise helps mitigate this risk. It's important to maintain a balance where technology aids medical professionals rather than supplants their judgment entirely.

- The use of patient data raises concerns about data privacy and consent. Without proper safeguards, there's a risk of unauthorized access or misuse of sensitive medical information. to address this dilemma, access to the data is to authorized personnel only

6.4 Relevance to Jordan and Region (Social, Cultural, and Political)

Healthcare holds significant importance in addressing the challenges of lung cancer detection and treatment. By developing a centralized platform for medical data training , we aim to address the unique needs and obstacles encountered by healthcare professionals in Jordan to detect lung cancer in the early stage. This project aligns with the government's vision of advancing healthcare technology and supports the nation's goals of improving healthcare services and outcomes for patients affected by lung cancer.

Chapter 7: Project Management

7.1 Timeline of Project Schedule

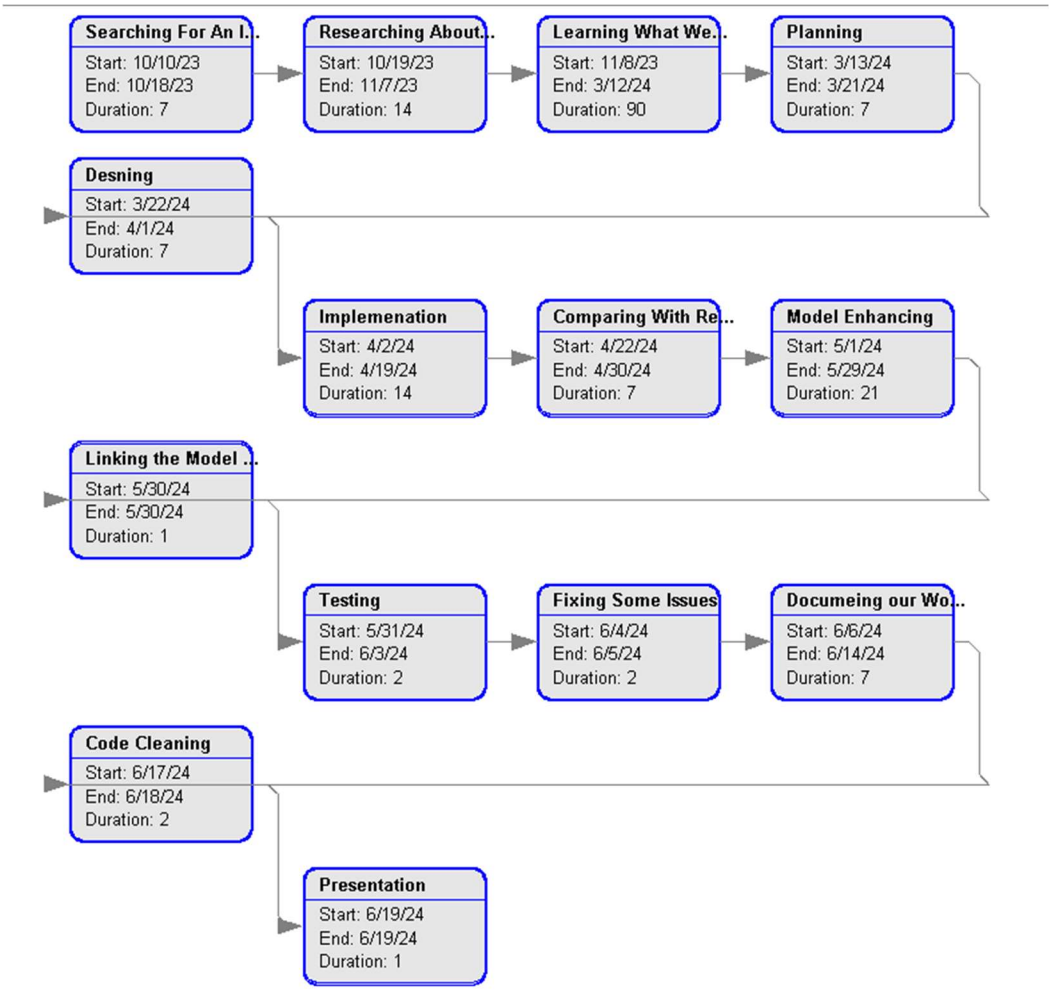


Figure 27: PERT Chart

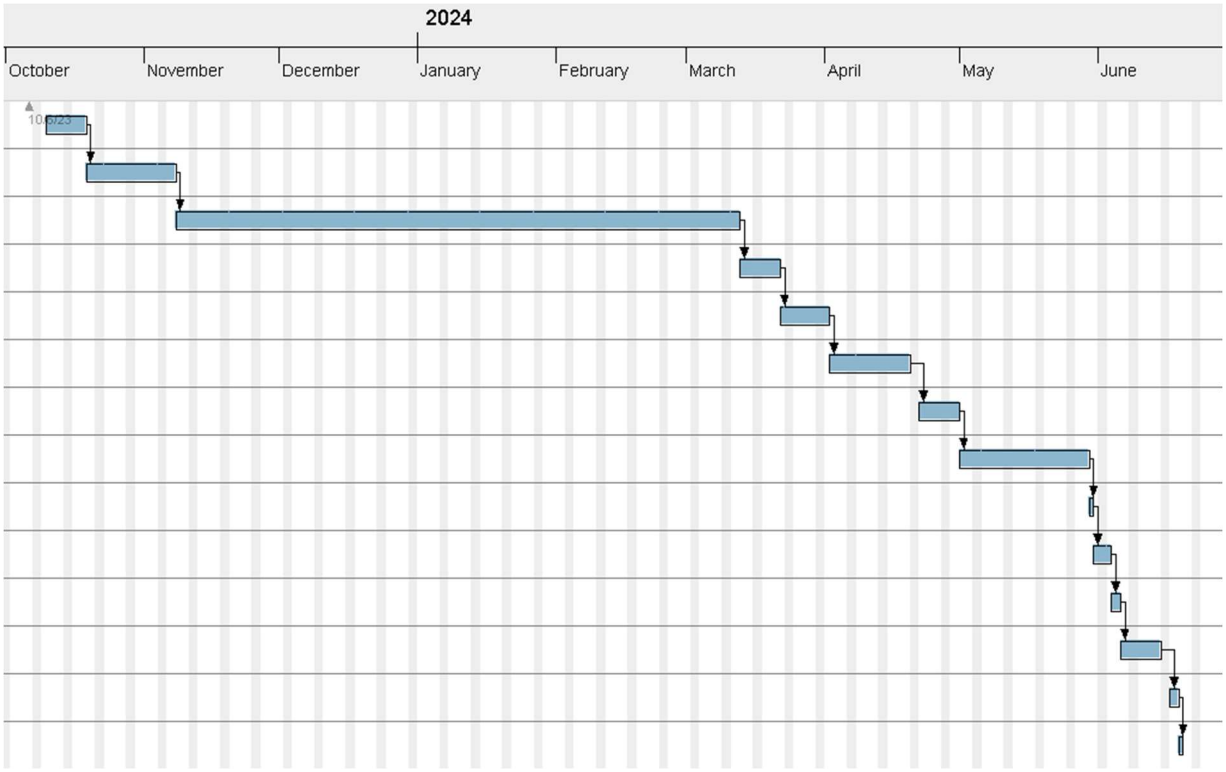


Figure 28 Gantt Chart

7.2 Resource and Cost Management

Using existing resources, such as open-source software, developing the project did not cost money. The project costs time and effort

7.3 Quality Management

To ensure quality management throughout the project, we implemented the following measures:

- We made sure everyone on the team knows the requirements and understand their role in meeting them.
- Implement automated testing and validation processes to ensure the project meets quality standards.
- Define and track key performance metrics for the project, such as AUC, Precision, and Recall

7.4 Risk Management

Plans development and compliance will mitigate any risk of delivery delay. The biggest risk that might be faced is the lack of time

Chapter 8: Conclusion and Future Work

8.1 Main contributions of the work.

This project has successfully developed a lung cancer prediction System marking a significant advancement in healthcare resource management. The main contributions of this work are as follows:

- **Centralized Platform Creation:** We achieved the objective of simplifying the organization of medical data by creating a centralized platform for healthcare professionals to upload images and obtain probabilities of lung cancer.

- **Early Detection:** analyze medical CT scan images with high accuracy, enabling the early detection of lung cancer and increases treatment options for patients.

- **Reduced Workload for Healthcare Professionals:** This frees up time for clinicians to focus on patient care

8.2 Further future work

We could add a database that records patient histories, test results, types of treatment received by patients, and any changes in their health status over time can help improve models and provide accurate and useful information to enhance lung cancer detection

References

[1] M. B. Schabath and M. L. Cote, "Cancer progress and priorities: Lung cancer," Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, vol. 28, no. 10, pp. 1563-1579, 2019.

[2] I. Shafi, S. Din, A. Khan, I. D. L. T. Díez, R. d. J. P. Casanova, K. T. Pifarre and I. Ashraf, "An effective method for lung cancer diagnosis from CT scan using deep learning-based support vector network," Cancers, vol. 14, no. 21, p. 5457, 2022.

[3] "Researchgate.net," [Online]. Available: https://www.researchgate.net/figure/CT-scans-based-on-maximum-intensity-projection_fig1_365221751.

[4] E. Pratt, "Healthline," 03 01 2023. [Online]. Available: https://www.healthline.com/health/lung-cancer/screening-and-early-detection.

[5] M. Hany, "Kaggle," Chest CT-Scan images Dataset, [Online]. Available: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images.

[6] N. Maleki, "Medeley Data," CT-Scan images, [Online]. Available: https://data.mendeley.com/datasets/p2r42nm2ty/1http://dx.doi.org/10.17632/P2R42NM2TY.1.

[7] H. F. Al-Yasriy and M. S. Al-Huseiny, "Kaggle," The IQ-OTHNCCD lung cancer dataset, [Online]. Available: https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset?select=The+IQ-OTHNCCD+lung+cancer+dataset.

[8] S. AIMI, "Azurewebsites.net," Stanford AIMI shared datasets, [Online]. Available: https://stanfordaimi.azurewebsites.net/datasets/8cbd9ed4-2eb9-4565-affc-111cf4f7ebe2.

[9] "Blogspot.com," Data Science and Machine Learning, [Online]. Available: https://manisha-sirsat.blogspot.com/2019/04/confusion-matrix.html.

[10] [Online]. Available: https://www.thirdrocktechkno.com/blog/how-to-outsource-software-development-projects-the-right-way/?fbclid=IwAR2kSQg8g54LSVUR7tsyq51-UFKe5cHCVAQPsyrr2A2CdVWSuunVg7h5_8g.

[11] [Online]. Available: https://www.researchgate.net/figure/The-steps-of-Image-Processing-Technique-for-Lung-Disease-Diagnosis-LDD-using_fig1_344479889.