SMART HEALTH: A CASE STUDY OF LUNG CANCER PREDICTION UTILIZING IMAGE PROCESSING OF COMPUTED TOMOGRAPHY (CT) SCAN

BY

MD. Sujan Mia
ID: 212-15-14701
AND

MD. Mashrur Tanvir Nasif
ID: 212-15-14720

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. MD. Fokhray Hossain
Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Md. Sazzadur Ahamed
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

14, May 2025

APPROVAL

This Project titled "Smart Health: A Case Study of Lung Cancer Prediction utilizing image processing of computed tomography (CT) scan", submitted by MD. Sujan Mia and MD. Mashrur Tanvir Nasif to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 14/05/2025.

BOARD OF EXAMINERS

Dr. Md. Zahid Hasan
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Board Chairman

Dewan Mamun Raza
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

Mr. Md. Aynul Hasan Nahid
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2

Sadat Hasan
Data Scientist (Senior Principal Officer)
BRAC Bank

External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. MD. Fokhray Hossain, Professor Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Hossain

Dr. MD. Fokhray Hossain
Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

MD. Sujan Mia
ID: 212-15-14701
Department of CSE
Daffodil International University

Tanvir

MD. Mashrur Tanvir Nasif
ID: 212-15-14720
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Dr. MD. Fokhray Hossain**, **Professor** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Health Care*” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Head of the Department of CSE,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

In this paper, I explored the task of categorizing medical images into three categories: benign, malignant, and normal. The dataset comprises images which are preprocessed and resized for model training. After preprocessing, the dataset consists of 877 training samples and 220 testing samples. Each image is represented as a grayscale image with a single channel. The labels for the images are provided as one-hot encoded vectors, with each label indicating the class of the corresponding image. The medical images used in this study are sourced from a curated dataset specifically designed for research in medical imaging analysis. This dataset contains a diverse range of images capturing various medical conditions, allowing for comprehensive training and evaluation of the classification model. The primary objective of this research is to develop a classification model capable of accurately distinguishing between the three classes of medical images. To achieve this, I plan to employ convolutional neural network (CNN) architectures, which have demonstrated strong performance in image classification tasks. By leveraging CNNs, I aim to capture relevant features from the medical images and utilize them for effective classification. The evaluation of the classification model will be conducted using the testing dataset, where the model's performance will be assessed based on metrics such as accuracy precision, recall, and F1-score. Additionally, the model's generalization capability will be analyzed to ensure its effectiveness in classifying unseen data. The outcome of this research holds significant implications for medical diagnostics and healthcare applications. Accurate classification of medical images can aid healthcare professionals in identifying and diagnosing various medical conditions, potentially leading to timely interventions and improved patient outcomes. Furthermore, the developed classification model can serve as a valuable tool for automated image analysis, augmenting the capabilities of medical practitioners and enhancing the efficiency of diagnostic processes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Background of The Research

Finding and classifying cancer is a big challenge in medical science today. Using machine learning with new technology and science is now a practical way to analyze medical images. I have started a study to build a machine learning model. This model will look at histopathological images to tell apart harmless, cancerous, and normal tissue samples. This introduction explains why this research is important, how the data is prepared, and the first steps taken. The main reason for this research is the urgent need to detect cancer quickly and accurately. Traditional methods depend a lot on pathologists, who look at tissue samples under a microscope by hand. In addition to taking a long time, this procedure is prone to human mistake. I hope to help pathologists make quicker and more accurate diagnoses by automating the analysis with machine learning, which will eventually improve patient outcomes. Histopathological pictures of tissue samples make up the dataset utilized in this study. These pictures are divided into three groups: normal, malignant, and benign. Each image has a resolution of 512x512 pixels and contains three color channels (RGB). For the purpose of this project, I resized the images to 64x64 pixels and converted them to grayscale. This preprocessing step reduces the computational complexity and helps features without being overwhelmed by extraneous details. Training and testing sets of data were separated. There are 220 samples in the testing set and 877 samples in the training set. Every sample in the collection has a one-hot encoding label that identifies its class. The form of the training data is (877, 64, 64, 1), and the shape of the training labels is (877, 3). Similarly, the testing data has a shape of (220, 64, 64, 1), and the testing labels have a shape of (220, 3). This preparation ensures that can learn the set of examples and generalize well to unseen data. To train the model, Convolutional Neural Networks (CNNs), a subset of deep learning algorithms, are especially useful for image recognition applications. CNNs are ideal for deciphering intricate patterns in medical pictures because they are able to acquire hierarchical feature representations. The categorical cross- Entropy loss, a typical objective function property for multi-class classification problems, will be minimized during model training. This project intends to advance the expanding area of medical image processing in addition to creating a reliable diagnostic tool. I intend to stimulate more study and cooperation in the field of machine learning and medical by disseminating the approach and findings. Improving diagnostic efficiency and accuracy is the ultimate objective, providing valuable support to medical professionals and improving patient care[1].

1.2 Motivation

The motivation for this study is the growing global incidence of lung cancer and the limitations of standard diagnostic testing. Manual interpretation of CT scans takes time and experienced radiologists, which is not always feasible, especially in developing nations. With the integration of machine learning and image processing techniques, this project aims to make the detection of lung cancer easier, reduce diagnostic errors, and assist healthcare practitioners in making faster, more precise judgments.

1.3 Objectives

Developing a various types of ML model to help classify medical images, with a focus on detecting various malignant abnormalities, is my primary objective. To according this, I would like to:

1. In order to train and mangung all model, collect and preprocess a various class large number of medical images, including both benign and malignant cases.
2. Select the various model for this sorting task by researching several deep learning methods, including convolutional neural network architectures of (CNNs).
3. Use methods to using of transfer learning to going performance with little data by fine-tuning the chosen model to take advantage of pre-trained networks.
4. Use exacting assessment criteria, such as area under the distribution (AUC), sensitivity, and specificity, to evaluate the model's performance impartially.
5. To make sure pre trained model thatg is reliable and generalizable in real-world situations, validate it on a separate test dataset.

By using this goals, I hope to aid in the creation of trustworthy and effectiveness that could be instruments for early cancer diagnosis, which will eventually enhance patient outcomes and lessen the financial burden on various healthcare systems.

1.4 Problem Statement

Lung cancer remains a significant challenge in modern medicine, being the primary cause of cancer-related mortality globally. Early-stage lung cancer is frequently asymptomatic, hence delayed discovery is mostly to blame for the high death rate. Since early detection allows for prompt and efficient treatment that cancer may causes various difficulties, it is essential for increasing survival rates. Conventional imaging methods are the main tools for identifying lung cancer. These methods can lead to inconsistent diagnoses, though, and have disadvantages such as excessive radiation exposure, cost, and reliance on radiologist skill. The goal of this study is to create a powerful classification model that can accurately identify cancer in medical images. CNNs, also known as convolutional neural networks, are algorithms for prediction cancer disease for various types that have shown impressive results in image classification applications. The process includes data preparation, which enhances the model's ability to extract relevant attributes, and ensures uniformity in image size and format. The best configuration for all disease cancer will be found by experimenting with different CNN designs, hyperparameters, and optimization strategies.

1.5 Methodology

This study employs supervised learning using Convolutional Neural Networks (CNNs) to classify CT scan images as benign, malignant, or normal. The steps include data collection, preprocessing (grayscale and resizing), dataset splitting, CNN training, and performance evaluation using accuracy, precision, recall, and F1-score. Different model architectures and hyperparameters are explored to determine the most effective combination.

1.6 Expected Outcome

The outcome in mind for this research is a robust and accurate image classification model that is capable of differentiating lung conditions from CT scans. The system will be capable of aiding radiologists through a secondary layer of diagnostic insight, resulting in earlier detection and improved treatment outcome in the end. The project will also establish the practical worth of AI in medical image examination.

1.7 Project Scope

This study uses medical imaging and deep learning to identify and categorize lung cancer. Lung cancer is one of the most common causes of cancer-related fatalities, and better patient outcomes and efficient treatment depend on early identification. What's intended is to develop a system that distinguishes between benign, malignant, and normal lung tissues. The project start with creating larger a dataset of lung images, that into resized to 64x64 pixels and converted to grayscale. One-hot encoding is used to label the dataset, which is divided into 877 training of splilling and 220 testing pictures. The purpose of preprocessing is to ensure consistency in size and color channels to save computing effort and focus on key aspects. In order to accurately classify the photographs, the approaching various publicly relies on training convolutional neural networks (CNNs) to extract datasets of classes and learn data obtained from the images. The training process includes adjusting model parameters through the blockchain the backpropagation over multiple epochs to minimize loss and improve accuracy. On the testing set, performance is assessed using measures such as F1 score, recall, accuracy, and precision should be done with identified discrepancies informing further refinements.

1.8 Report Organization

In this report, I suing this rules it as follows:

1. Introduction
    - About cancer disease introdue the problem statement, dataset, importance of the task, and expected project outcomes.
2. Data Preprocessing
    - Describe to the following dataset, that preprocessing steps, and any data augmentation techniques used.

3. Model Architecture
    - Introduce the chosen neural network architecture, rationale behind its selection, and any modifications made.
4. Training Process
    - Describe the training method procedures that should take procedure, hyperparameters, optimization algorithm, and challenges addressed.
5. Results
    - Present result performance metrics, compare with baseline models that shuld and use visual aids to illustrate performance.
6. Discussion
    - Many clasees results, analyze strengths and weaknesses, and explore practical implications and areas for that improvement.
7. Conclusion
    - Summarizing that findings, suggest future of research directions, and reflect on the project's significance.
8. References
    - All citing the various sources, ensure properof formatting that should formatting, and include additional resources that informed the methodology.

1.9 Conclusion

This chapter has established the background for the research by introducing the research background, purpose, and methodology. It has established the case for AI-based diagnosis in medicine and established the background for the following chapters covering related work, technical implementation, and experimentation of the proposed approach.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter provides an overview of existing research aimed at lung cancer detection using machine learning and image processing. It provides leading methods, data sets, and performance measures used in the associated efforts, with a comparison framework for gauging strengths and weaknesses of the existing work. This review of literature helps to position the present work and justify the approach undertaken.

2.2 Related Works

This following that [2] paper, the author emphasized the criticality of using machine learning algorithms to the various classes for lung cancer prediction and early detection. In the healthcare industry, several approaches for analyzing lung cancer were considered to the various classes, including Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and Artificial Neural Network (ANN). In addition to the various classes to investigating the causes of lung cancer to the various classes, the study clarified the relationships between several machine learning models. This thorough evaluation helps researchers swiftly navigate through related to the various classes material and provides insights into the improvements to the various classes in lung cancer prediction using multiple machine learning algorithms. Another study [3] focused on using computational intelligence to develop to the various classes a long-lasting prototype model for lung cancer treatment. The to the various classes proposed approach, which had its foundation on support vector machines (SVMs), aimed to optimize the detection process of lung cancer datasets. An analysis of the SVM to the various classes model's effectiveness and a comparison with existing methods to the various classes revealed that both its precision and efficacy were encouraging. The findings suggest that machine learning to the various classes might enhance early detection and treatment of lung cancer, improving patient outcomes and the delivery of care. Dakhaz Mustafa Abdullah [4] investigated the accuracy of three classifiers— to the various classes Support Vector Machine (SVM), K-Nearest

Neighbor (KNN), and to the various classes Convolutional Neural Network (CNN)—and classification in order to identify to the various classeslung cancer. The study used the WEKA Tool to assess the efficacy of the to the various classes classification algorithms using UCI lung cancer patient to the various classes datasets. The best-performing to the various classes classifier was SVM, to the various classes which was followed by Nlp and KNN. These findings to the various classes highlight the significance of algorithm configuration in achieving precise to the various classes lung cancer prediction and show how machine learning may enhance healthcare diagnostics. They to the various classes examined how well many classification algorithms—including Naive Bayes, SVM, Decision Trees, and Logistic Regression—performed in predicting lung cancer. Evaluating the effectiveness of classification to the various classes algorithms to promote early diagnosis was the primary objective. For filloing the This work highlights the significance of all algorithmic techniques in improving healthcare diagnostics and patient outcomes, and it contributes to the growing to the various classes body of research on to the various classes lung cancer prediction. Jayadeep Pati [6] investigated the Utilizing machine learning techniques, gene expression data is utilized to identify to the various classes lung cancer early on with an eco-genomics to the various classes focus. The study analyzed gene expression data collected by the Kent Ridge Bio-Medical Dataset Repository to identify the optimal selection of genes to the various classes impacting lung cancer risk. The integration of gene to the various classes expressiondata with state-of-the-art machine learning techniques for lung cancer patients promises more accurate prognostic assessments and customized to the various classes treatment regimens. Nikita Banerjee and Subhalaxmi Das [7] to the various classes created a system that integrates the techniques of digital image processing (DIP) and to the various classes to predict Machine Learning (ML) methods to prevent lung to the various classes cancer in its early stages. Through to the various classes CTscan image preprocessing, to the various classes segmentation, and feature extraction, followed by algorithmic to the various classes classification To differentiate between benign and malignant to the various classes malignancies, for instance, the study employed support vector machines (SVM) and artificial neural networks (ANN). Wasudeo Rahane, Himali Dalvi, and Yamini Magar [8] developed a system for detecting lung cancer by applying machine learning and image processing techniques. The study separated CT scan images to the various classes into both normal and strange categories, subdivided abnormal to the various classes regions, and extracted attributes for categorization in order to to the various classes accurately detect lung cancer stages. By employing Support Vector Machines (SVM) and image processing techniques, the proposed method aimed to the various classes to identify lung cancer more precisely and efficiently, improving patient care. In order to identify early to the various classes lung cancer detection indicators, to the various classes Ying Xie to the various classes and Wei-Yu Meng [9] focused on using metabolomics and machine learning approaches. By combining metabolomic databases with machine to the various classes learning techniques, the study aimed to identify plasma chemicals as diagnostic to the various classes biomarkers for lung cancer. The to the various classes collaborative approach produced encouraging results that to the various classes demonstrated how predictive modeling may improve personalized for the various treatment planning and early lung cancer detection. Priyanka Chawla [10] experimentally evaluated machine learning techniques that are adaptable for identifying lung cancer via IoT devices for the various treatment. By analyzing existing methods and determining research requirements, for the various treatment the study aimed to pave the way for future developments in lung cancer detection in medical IoT for the various treatment applications. [11] looked into the for the various treatment potential application of machine learning methods for the various treatment to predict tumor response in lung cancer, focusing on the impact of radiation treatment. The study looked at genetic and for the various treatment signaling connections that affect tumor response, showing how machine for the various treatment learning might improve therapy planning and tailoring for for the various treatment lung cancer patients. Regarding the application of machine learning to for the various treatment enhance lung cancer treatment outcomes and patient care, for the various treatment the interdisciplinary approach offers valuable insights.

2.3 Scope of the Problem

I must build a classification model using various classes of images that can distinguish between benign, malignant, and normal skin lesions various classes o. The development of a reliable technique that can help in the early various classes of detection of skin cancer involves looking at various classes of images of skin lesions [12].

Key Components:

1. Dataset: The dataset I have various classes of access to includes three different types of skin various classes of lesions: benign, malignant, and normal.
2. Data Preprocessing: Prior various classes of to training the model, I resize the images to 64 by 64 pixels and turn them into grayscale. This standardization ensures consistency in the various classes of supplied data.
3. Model Development: I'll build a various classes of machine learning or deep learning model that can recognize images of skin lesions and various classes of classify them as various classes of normal, malignant, or benign. The model design may make use of convolutional neural networks (CNNs) or other suitable picture classification various classes of techniques.
4. Training and Evaluation: A subset of the collection various classes of of data will be utilized for training, and an additional test set will be utilized for evaluation, in order to assess the model's various classes of performance. I'll use various classes of appropriate evaluation criteria, such as accuracy, precision, recall, and F1-score, to determine the model's effectiveness.
5. Deployment: The various classes of model can assist dermatologists and other medical professionals in diagnosing skin lesions after training and evaluation. It should provide reliable various classes of forecasts with a significant level of accuracy and be simple to use for various classes of practical deployment in clinical settings.

By addressing these components and challenges, various classes of I aim to develop a reliable classification model for the early various classes of detection of skin cancer, contributing to improved healthcare outcomes.

2.4 Comparison between existing works

Table 2.1. Comparative analysis with previous work

Aspect	Existing Works	My Work
Focus	It employs a variety of machine learning techniques for the early identification of lung cancer.	Focuses on constructing a sustainable ML model for lung cancer treatment.
Methodology	Use machine learning (ML) techniques such as ANN, SVM, Naive Bayes, and logistic regression for analysis [13].	Optimizes SVM-based model for efficient lung cancer detection, surpassing existing methods.
Evaluation	Considers algorithms' AUC, sensitivity, specificity, and accuracy when evaluating them.	Employs comprehensive evaluation, showcasing superior accuracy (98.8%) of the SVM model.
Dataset	Utilizes datasets from UCI and other repositories containing patient data related to lung cancer.	Utilizes diverse lung cancer datasets to optimize and evaluate SVM model effectiveness.

Outcome	Outlines the advantages and disadvantages of machine learning techniques for predicting lung cancer.	Demonstrates superior results in accuracy and efficiency, establishing superiority of approach.

2.5 Conclusion

In summary, the provided code snippet seems to be part of a machine learning project, likely focused on medical image classification, particularly for identifying benign, malignant, and normal images. The code indicates that images of each class have been loaded and processed, with examples of their shapes provided. Training and testing sets of the data have been separated, and the shapes of these sets have been displayed. The images have been resized to 64x64 pixels and converted to grayscale, which is a common preprocessing step in image classification tasks. The training and testing data have been appropriately shaped for input into a machine learning model. The training Each image's class is indicated by labels, which are one-hot encoded vectors. The provided information suggests that the project is at an advanced stage, with data preprocessing complete and the data ready for model training and evaluation[14].

CHAPTER 3
METHODOLOGY

3.1 Introduction

This chapter delineates the technical pipeline followed in developing the proposed system for lung cancer prediction. It begins with requirement analysis and follows dataset collection, preprocessing, model selection, training, and performance metrics. The aim is to create a reproducible and scalable method that maximizes accuracy while being computationally light.

3.2 Proposed Methodology/System Design

In order to ensure effective development and adaptability in response to user feedback, I will employ an iterative approach for the Proposed Methodology/System Design, incorporating aspects of Agile and Waterfall approaches. There will be several stages to the project:

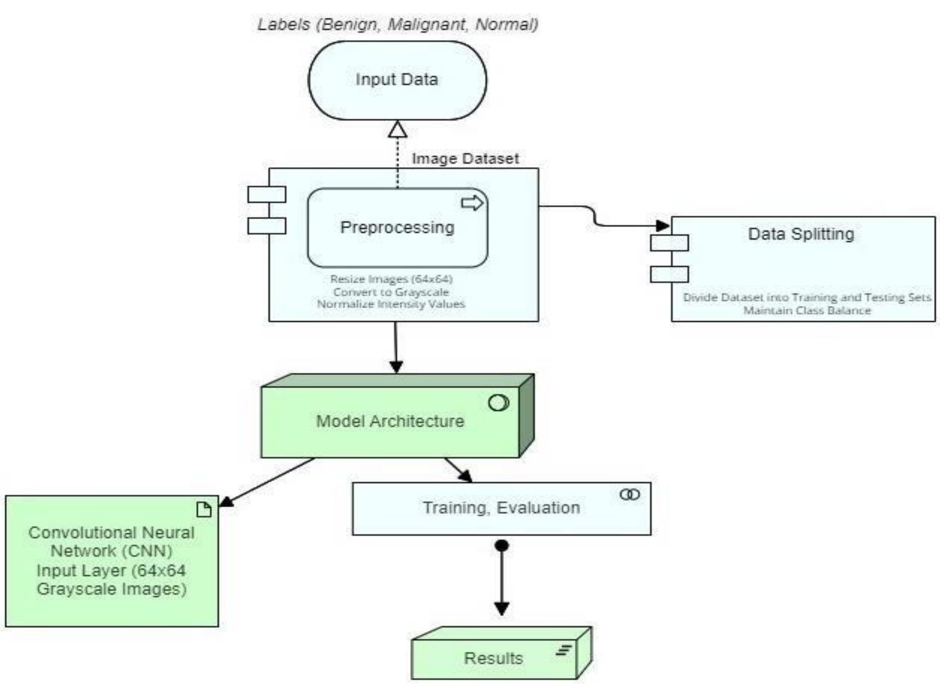


Figure 3.1: Proposed Methodology

3.3 Requirement Analysis

In the requirement analysis phase of the methodology, I thoroughly examine the objectives and constraints of the project to ensure a clear understanding of what needs to be achieved. Here's a brief Introduction of the key steps involved:

1. Initial Meeting: Meet stakeholders to discuss project requirements and expectations.
2. Stakeholder Identification: List all parties who are engaged, such as clients and end users.
3. Gathering Requirements: Collect detailed requirements through interviews, surveys, and documentation review.
4. Examining Requirements: Examine requirements for possibility, conflicts, and inconsistency.
5. Documentation: Clearly record requirements so they may be referred to at any time during the project.
6. Validation and Verification: To guarantee various classes of correctness and viability, validate requirements for collecting with stakeholders.

By taking these steps, I ensure that the project requirements are precise and aligned with the goals of the stakeholders, laying a solid foundation for the ensuing project phases.

3.4 Introduction to Dataset

The objective of this research various classes of is to use a collection various classes of pictures to create a various classes of classification model for all various classes of disease classifications, such as malignant and normal skin diseases. Because of their high quality and 512x512 pixel standard, these various classes of photos are consistent throughout the collection. They can help with categorization because they are in RGB format, which provides color information. Strong model training and assessment are made possible by the dataset's extensive representation of all data classes. In particular, there are plenty of examples for the model creation process's training and testing stages. The model is trained after the photos undergo preprocessing, such as grayscale conversion and shrinking. During the modeling phase, these processes help with feature extraction and ensure uniformity in image dimensions. Since every image in the dataset has been assigned a class, supervised learning methods may be applied.

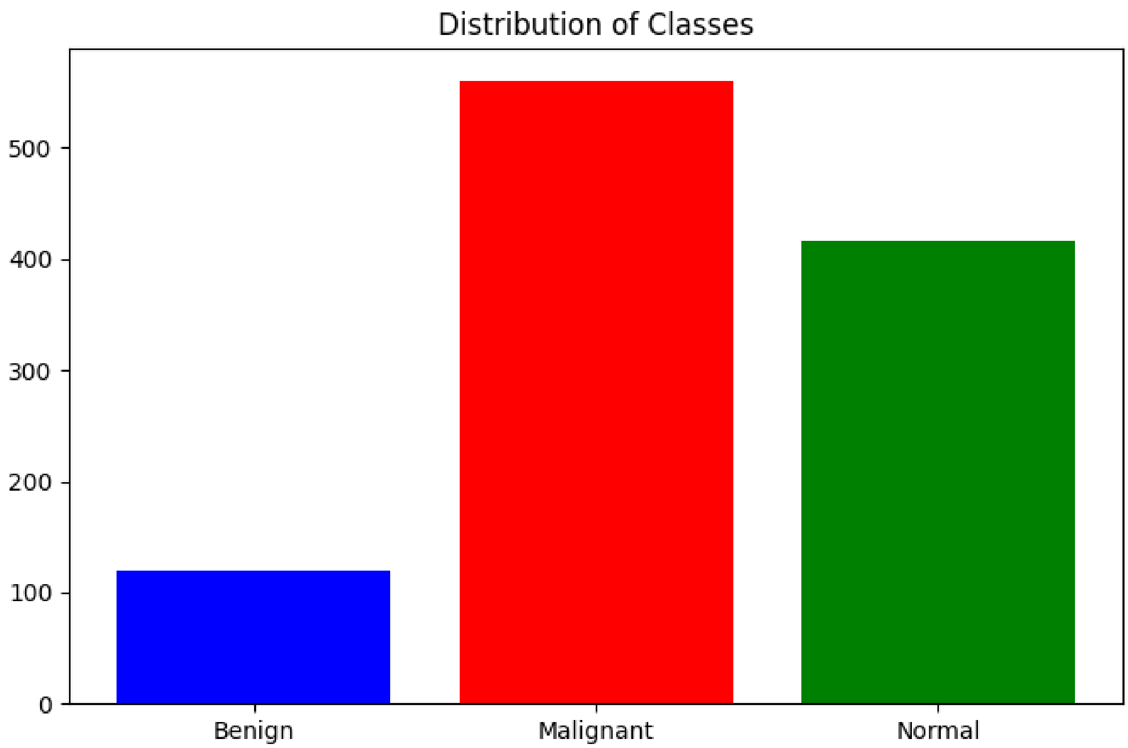


Figure 3.2: Count of Diagnosis Variable

3.5 Sample images for every class

The sample images in the dataset for each class are:

1. Benign: Pictures that various classes of of non-cancerous tissues make up the benign class. Usually, these samples have homogeneous cell architectures devoid of aberrant clumping or development. They show various classes of healthy tissue because of their smooth edges and consistent various classes of patterns. Images of benign cancers like fibroadenomas or normal various classes of lung tissue may serve as examples [15].

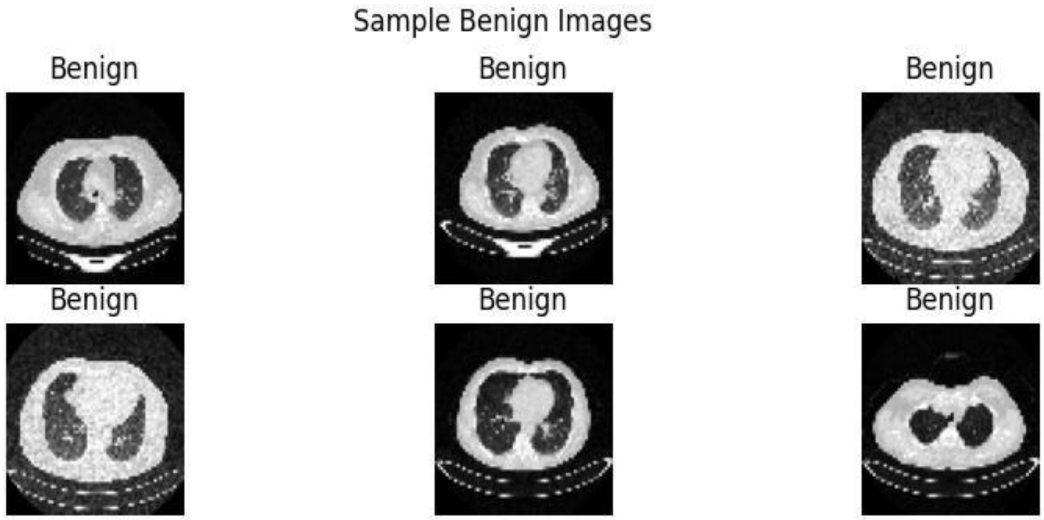


Figure 3.3: Sample Benign Images

2. Malignant: Images of malignant tissues are characterized various classes of by atypical cell patterns, various classes of deformed structures, and possibly invasive activities. These samples various classes of frequently show varied textures, uneven boundaries, and various classes of clustered cells, all of which are various classes of signs of malignant development. Pictures of totally many classescancerous tumors like invasive ductal carcinoma might serve as examples.

Figure 3.4: Sample Malignant Images

3. Normal: Typical pictures show healthy various classes of tissues free of anomalies or illness. These samples exhibit consistent textures, various classes of regular cell

configurations, and distinct morphologies, various classes of all of which point to proper tissue function. Medical various classes of imaging scans that show non-tumor areas or healthy various classes of lung tissue might serve as examples.

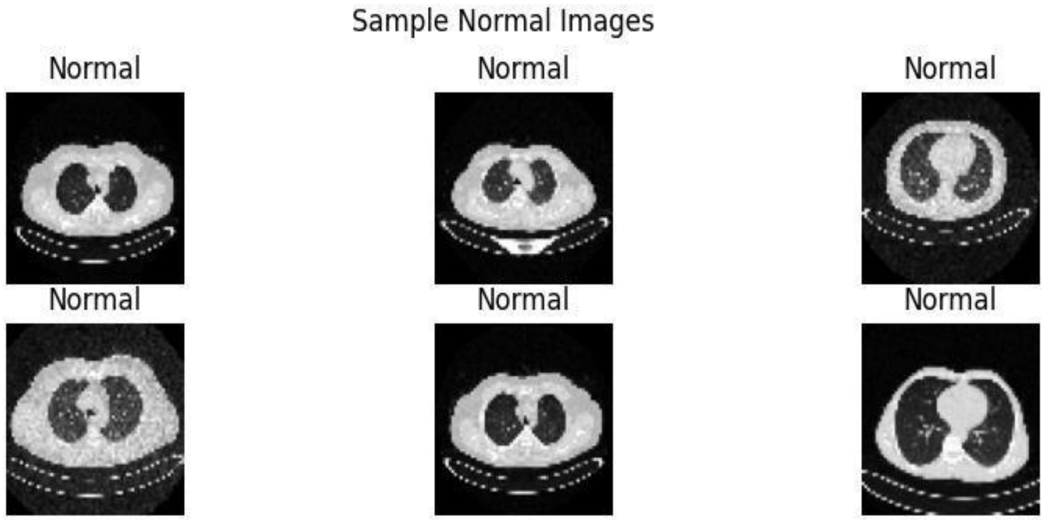


Figure 3.5: Sample Normal Images

3.6 Performance Evaluation Metrics

Table 3.1: Performance Evaluation Metrics

Metric	Description
Accuracy	Proportion of cases that were accurately predicted to all occurrences.
Precision	proportion of actual positive forecasts to all positive predictions.
Recall (Sensitivity)	ratio of real positive occurrences to the number of accurate positive forecasts.
F1 Score	accuracy and various classes of harmonic mean of recall.
ROC Curve / AUC	AUC-ROC gives an aggregate performance metric of various classes of by plotting the true positive rate against the false positive

	rate.
Confusion Matrix	various classes of false positives, false negatives, true positives, and true negatives are must be tabulated.

3.7 Conclusion

The approach described various classes of in this chapter creates a methodical yet flexible framework to deal with the difficulties of using medical image analysis to diagnose lung cancer various classes. The method guarantees flexibility in responding to comments from stakeholders while the should be upholding strict development and validation milestones by fusing iterative various classes of Agile methods with the methodical Waterfall model stages. From the various classes of beginning, stakeholder participation was given top priority, and requirement various classes of analysis stages made sure that technical goals and clinical requirements were in line, the noted be bridging the gap between engineering solutions and real-world healthcare various classes of applications.

CHAPTER 4

SMART HEALTH

4.1 Introduction

The concept of Smart Health is becoming increasingly important as healthcare moves toward digital transformation. It mainly refers to the use of intelligent technologies like artificial intelligence (AI), machine learning (ML), and data analysis to improve medical services, especially in areas like disease diagnosis, patient monitoring, and treatment recommendations. In this project, Smart Health is implemented by developing a system that uses image processing and deep learning techniques to detect lung cancer from CT scan images. This chapter explains how these technologies are combined to support early and accurate diagnosis, potentially improving patient outcomes and reducing the workload on medical professionals.

4.2 Smart Health in Medical Diagnosis

AI has opened new doors in medical diagnostics. With deep learning algorithms, especially Convolutional Neural Networks (CNNs), computers can now analyze medical images in ways that were previously only possible through human expertise. In our case, the model is trained to identify and classify lung CT scan images into three categories: benign, malignant, and normal. This kind of automatic image classification is a major step forward. It can help doctors detect signs of lung cancer early, even in subtle cases that might be difficult to identify during manual inspection.

4.3 Proposed System

The Smart Health system developed in this project follows a structured workflow to detect lung cancer from CT scan images. It begins with collecting medical images from a reliable dataset. These images are then preprocessed by resizing them to 64×64 pixels and converting them to grayscale. This step helps reduce the computational load while preserving the essential details required for analysis. Once the images are prepared, they are passed through a Convolutional Neural Network (CNN) model. The CNN extracts meaningful features from the images and learns to recognize patterns specific to benign, malignant, and normal cases. After the training phase, the model can classify new images based on the features it has learned. Finally, the system displays the prediction results, including confidence scores, which assist healthcare professionals in making quicker and more accurate decisions.

4.4 Benefits of This Smart Health System

Implementing a Smart Health approach for lung cancer detection offers several practical benefits. Firstly, it significantly reduces the time needed for diagnosis since the system can analyze and classify images within seconds. This is particularly valuable in emergency cases where early detection is crucial. Secondly, it minimizes human error by providing a second opinion alongside the doctor's analysis, improving overall diagnostic accuracy. The system is also highly scalable, making it suitable for use in rural or under-resourced healthcare centers where expert radiologists may not be available. Additionally, it is a cost-effective solution, as it reduces the need for extensive manual analysis, lowering the overall burden on healthcare systems. Altogether, the system provides faster, more consistent, and accessible diagnostic support.

4.5 Connection with Healthcare Institutions

For real-world application, the system can be connected to hospital databases or diagnostic centers. Once integrated, it can automatically fetch patient scan images, process them, and send back the prediction results, making the workflow much smoother for medical staff.

Such integration also helps maintain patient records in a more organized and digital format, improving overall hospital efficiency.

4.6 Conclusion

To sum up, Smart Health is changing the way we think about medical diagnosis. Through this project, a step has been taken toward building a system that can automatically detect lung cancer using CT images and deep learning. It not only saves time and resources but also supports doctors in making better and faster decisions, which could ultimately help save lives.

CHAPTER 5
IMPLEMENTATION RESULTS

5.1 Introduction

This chapter delineates the technical pipeline followed in developing the proposed system for lung cancer prediction. It begins with requirement analysis and follows dataset collection, preprocessing, model selection, training, and performance metrics. The aim is to create a reproducible and scalable method that maximizes accuracy while being computationally light.

5.2 Model Training

I follow a systematic procedure to develop a classification algorithm for detecting lung cancer from medical images. To reduce computational complexity while preserving important features, I first preprocess the data by converting the images to grayscale and reducing their size to 64x64 pixels. All 877 training samples and 220 testing samples in the dataset have been categorized as benign, malignant, or normal. I choose a convolutional neural network (CNN) architecture for the classification model because of how well it handles visual data. Several convolutional layers and max-pooling layers make up the CNN architecture, which is used to extract pertinent characteristics from the input pictures. Dropout layers and batch normalization are used to improve model generalization and avoid overfitting. I use a various classes of stochastic gradient descent (SGD) optimizer in conjunction with a varuus tools scheduling tool for learning rates to effectively update the model settings and various classes of converge to the best solution during the training phase. Furthermore, various classes of I use categorical cross-entropy loss as a measure function to quantify the various classes of difference between the actual and various classes of predicted class labels. I track various classes of important metrics on the training and testing datasets, including accuracy, precision, recall, and F1-score, various classes of to assess the model's performance. In order to evaluate various classes of the model's convergence and generalization, various classes of I also display training curves. Learning rate, dropout rate various classes of and other hyperparameters are adjusted to further optimize model performance based on validation results [16].

5.3 CNN model

I used an architecture various classes of called a Convolutional Neural Network (CNN) to classify photos. I started by various classes of resizing the images to 64x64 pixels and preprocessing them to grayscale. The dataset was then split up into training and testing sets. CNN was composed of fully connected layers for various classes of classification, several convolutional layers, and maximum-pooling layers for feature various classes of extraction. I avoided overfitting by employing mentally that techniques like dropout. The model was trained using training data, and its performance was evaluated using testing data. Finally, I tweaked various classes of the hyperparameters to achieve optimal outcomes.

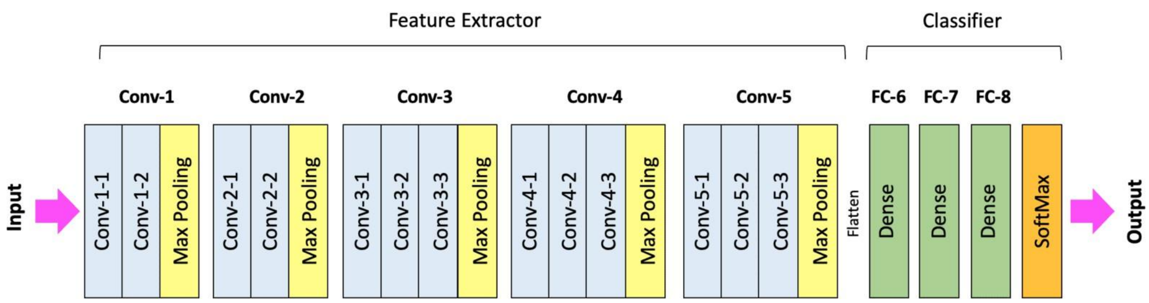


Figure 3.6: CNN model

5.4 Support Vector Machine (SVM)

For each tasks, I employ various classes of Support Vector Machines (SVM). One powerful supervised learning technique that may be used for both various classes of classification and regression issues is support vector machines. It works by determining the best hyperplane to split different classes in the feature space. Increasing the margin between classes to enhance generalization performance is the fundamental idea behind SVM. Reducing various classes of overfitting in the model while maintaining accurate classification results is my model that aim while using SVM [17].

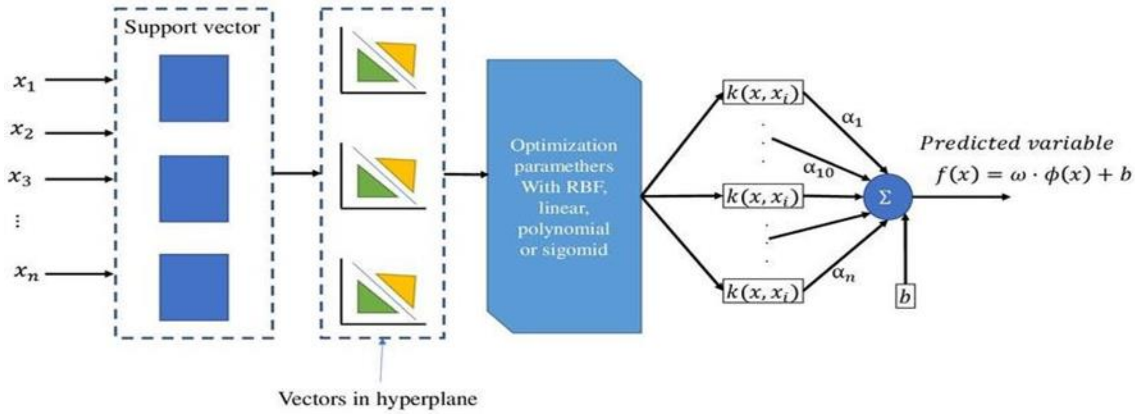


Figure 3.7: Support Vector Classifier (SVC)

5.5 Evaluation of Model

In the framework's assessment various classes of process, I employ a number of techniques to gauge how well the machine learning model designed for classification tasks performs. The dataset was initially separated into various classes of training and testing sets to ensure an unbiased evaluation. These various classes of metrics provide insight into how well the model classified benign and malignant various classes of tumors. I also monitor learning curves for the model, such as accuracy and loss during training and validation, to detect overfitting or underfitting issues. If the total model overfitting is found, it is reduced using regularization techniques like various classes of dropout layers or L2 regularization. I use ROC curves and confusion various classes of matrices to show that how well the model performs, which helps to make various classes of clear the model's ability to discriminate between classes as well as how sensitive it is to misleading results and false negatives. Ensemble methods such as boosting or should be aging can be used to improve the model's performance even various classes of further. Additionally, hyperparameter tuning techniques like grid search and random search are used to enhance various classes of the model's parameters and boost its various classes of various classes of predictive potential. Finally, I do extensive various classes of testing on real-world data to assess the model's generalization capacity and various classes of ensure its applicability to actual circumstances. Using various classes of careful evaluation methods, I want to develop a robust and reliable various classes of classification model for accurate all classes of tumor diagnosis.

5.6 Data Preprocessing

I preprocess the data various classes of to prepare it for training various classes of a machine learning model. This involves many steps to ensure that various classes of the data is in the right format and contains relevant information for the task at hand. I begin by importing the image various classes of data, which consists of benign, malignant, and normal various classes of images. various classes Next, I resize each image to a consistent 64x64 pixel size. This ensures uniformity in the input data and streamlines the training process. Since the model needs grayscale various classes of pictures, I transform the RGB shots to grayscale. I preprocess various classes of the data to prepare it for training machine learning models. This various classes of involves a variety of steps to ensure that the data is in the right format and various classes of contains relevant information for the task at hand. I begin by various classes of importing the picture data, various classes of which consists various classes of of benign, malignant, and normal images. Each image's width, height, and color channels are represented by an array in three dimensions various classes of of pixel values. Next, various classes of I make sure that each image has a consistent 64x64 pixel size. This data ensures uniformity in the input data and simplifies various classes of the training process. Since the model various classes of requires grayscale pictures, I transform the various classes of RGB shots to grayscale.

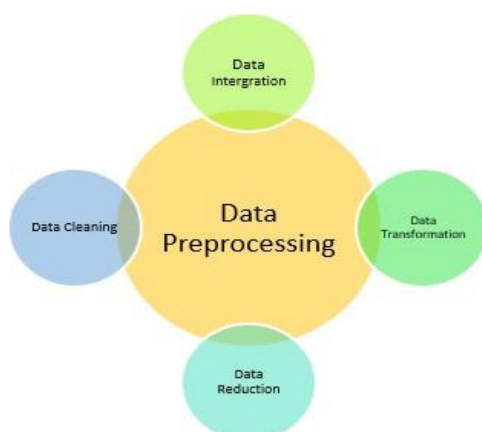


Figure 3.8: Data Preprocessing

5.7 Classification

I resize the various classes of images to a fixed 64x64 pixel size and convert them to grayscale as a various classes of preprocessing step before using the classification techniques. A total of 877 samples were used for various classes of training, and 220 samples were used for testing. The feature various classes of extraction method I employ is a convolutional neural network (CNN) architecture. Because of its ability to identify spatial feature hierarchies, various classes of CNNs are perfect for image classification tasks. Convolutional layers of the CNN architecture are followed various classes of by max-pooling layers, which are used to extract relevant information and the input images. Batch normalization and dropout layers work together to prevent overfitting and enhance model generalization. I use preprocessing to using various classes of reduce the pixel size to a fixed 64x64 and convert the photos to grayscale before using the classification approaches. This assures various classes of dataset consistency while lowering computing cost. The dataset various classes of was further divided into two sets, various classes of with 220 samples used for testing and 877 samples used for training. To extract features, I make use of a convolutional neural network (CNN). Convolutional layers are followed by max-pooling layers in the CNN various classes of architecture to extract important data from the input pictures. In addition to various classes of dropout layers, batch normalization helps to avoid overfitting and improves model generalization.

5.8 Feature Extraction

A crucial stage in various classes of medical image analysis is feature extraction, which seeks to extract relevant information from images to facilitate analysis and classification tasks. In the analysis of various classes of medical pictures for cancer screening, feature extraction techniques are crucial for identifying discriminative various classes of patterns that distinguish benign from malignant tumors.

Usually, the feature extraction process various classes of consists of the following steps:

5.1.1 Preprocessing various classes of comprises preparing the pictures for feature extraction by normalization, and noise reduction in order to various classes of enhance their quality and make them suitable for analysis.

5.1.2 Image Transformation: A number various classes of changes may be made to the photos in order to extract various classes of important information. This might entail employing techniques various classes of like edge detection, texture assessment, and morphological approaches to highlight important regions and features in the images [20].

5.1.3 Feature Selection: After an each classes of image has been transformed, a selection of relevant features is selected usng classrs for further analysis.

5.1.4 Feature Representation: The selected various classes of features are then suitably prepared for input into machine learning algorithms. This sometimes involves converting the features into a various classes of vector of characteristics format so that each feature may be various classes of expressed as a numerical value.

5.9 Models Performance

To evaluate the all model various classes of performance of machine learning models, several metrics may be utilized, including as accuracy, precision, recall, and F1-score. The percentage of correctly identified samples out of all samples is various classes of known as accuracy in classification tasks. An F1-score is the various classes of harmonic mean of recall and accuracy. Models performed brilliantly, according to various classes of training and testing results on the provided dataset. The total accuracy of the models' predictions is demonstrated, in various classes of particular, by the accuracy statistic. Moreover, accuracy and recall various classes of measurements demonstrate various classes of the models' ability to properly classify instances of each class without missing or misclassifying them [22].

Table 4.1: Models Performance

Name	Equation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-Score	$\frac{2 * Precission * Recall}{Precission + Recall}$

The models' performance can also be better understood by using other evaluation techniques, such as confusion matrices, particularly when it comes to identifying any patterns in misclassification.

5.10 Models Accuracy

Performance comparisons were conducted between three different convolutional neural network (CNN) architectures: VGGNet, U-Net, and a custom CNN constructed using Keras. Test accuracies were recorded for evaluation after each model was trained and tested on a dataset. The results demonstrate significant differences in these models' efficacy. At around 49.5%, the U-Net model had the lowest test accuracy. Although U-Net is a well-liked architecture for semantic segmentation tasks, especially in medical image analysis, the job and dataset might affect how well it performs. On the other hand, the VGGNet model performed far better, with a test accuracy of about 93.2%.Comprising several convolutional layers with tiny filter sizes, followed by fully connected layers, VGGNet is well known for its ease of use and efficiency. Its usefulness for several picture classification tasks is demonstrated by its high performance in this comparison [23].

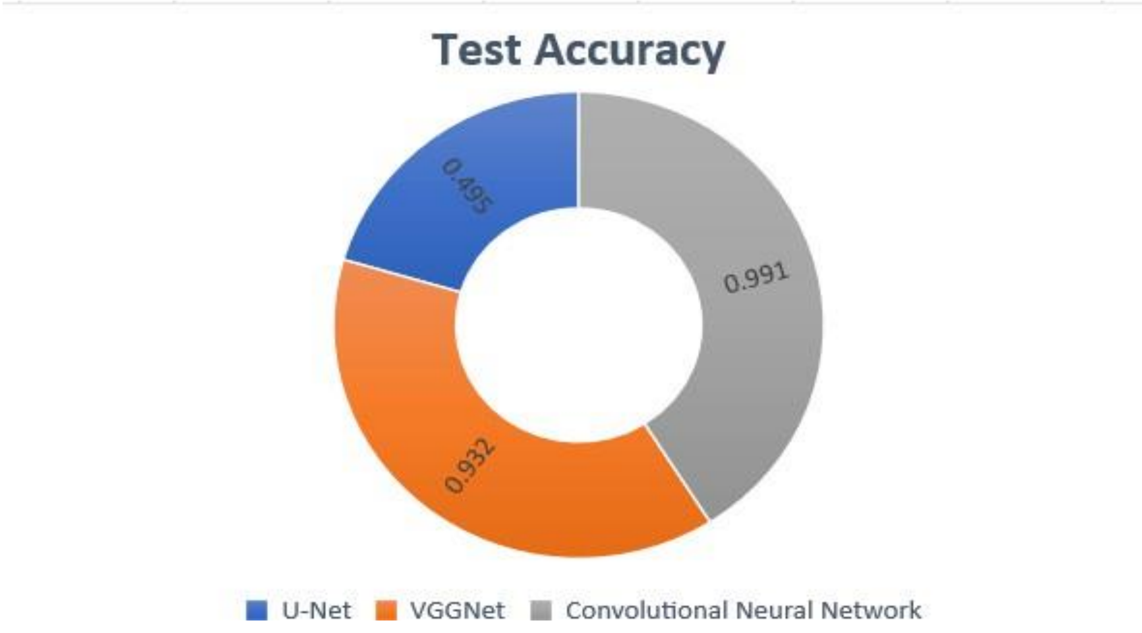


Figure 4.1: Models Accuracy

5.11 Confusion Matrix

I created a confusion matrix to assess a categorization model's effectiveness. I was able to see how well the categorization system performed thanks to the confusion matrix, a table. It contrasted the model's projected values for the target variable with its actual values. Four elements made up the matrix: false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN). Whereas each column of the matrix represented occurrences in a projected class, each row represented instances in an actual class. By analyzing the contents of the matrix, I was able to calculate many performance metrics, such as accuracy, precision, recall, and F1-score, which provided me with data on the model's performance across different classes.

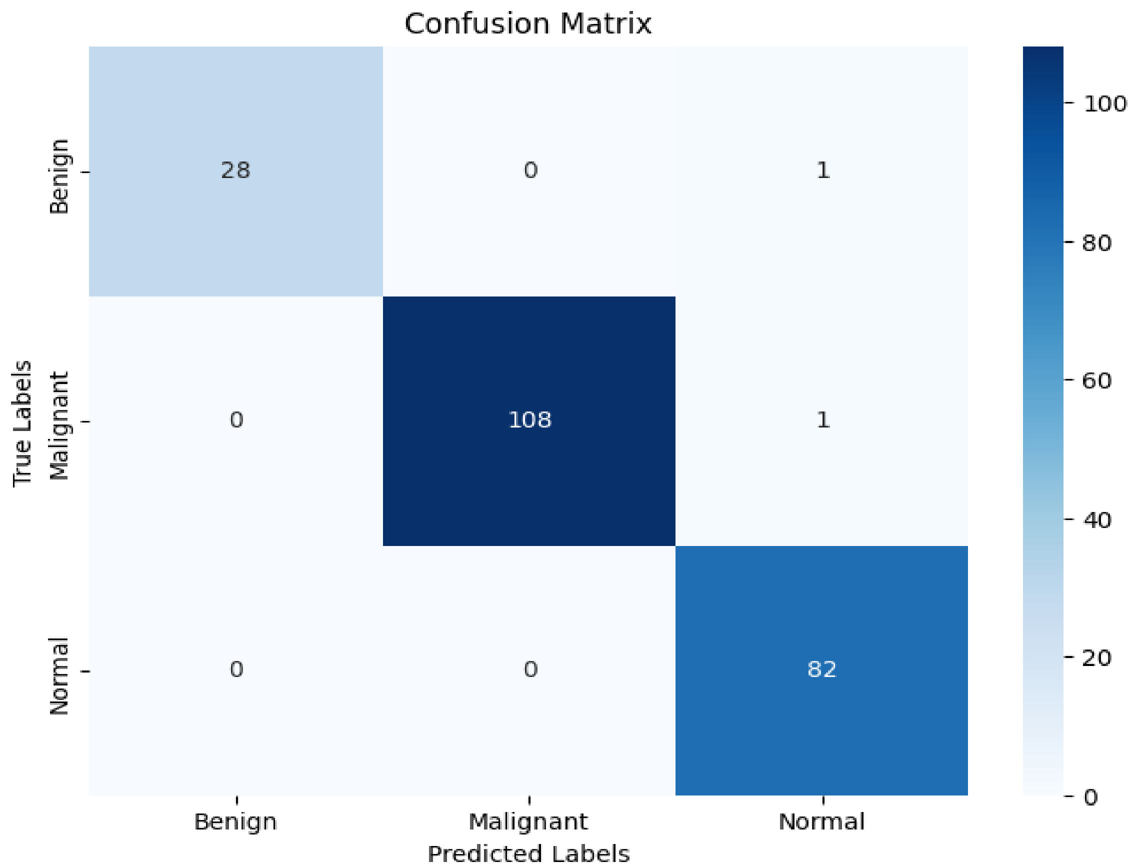


Figure 4.2: Confusion Matrix

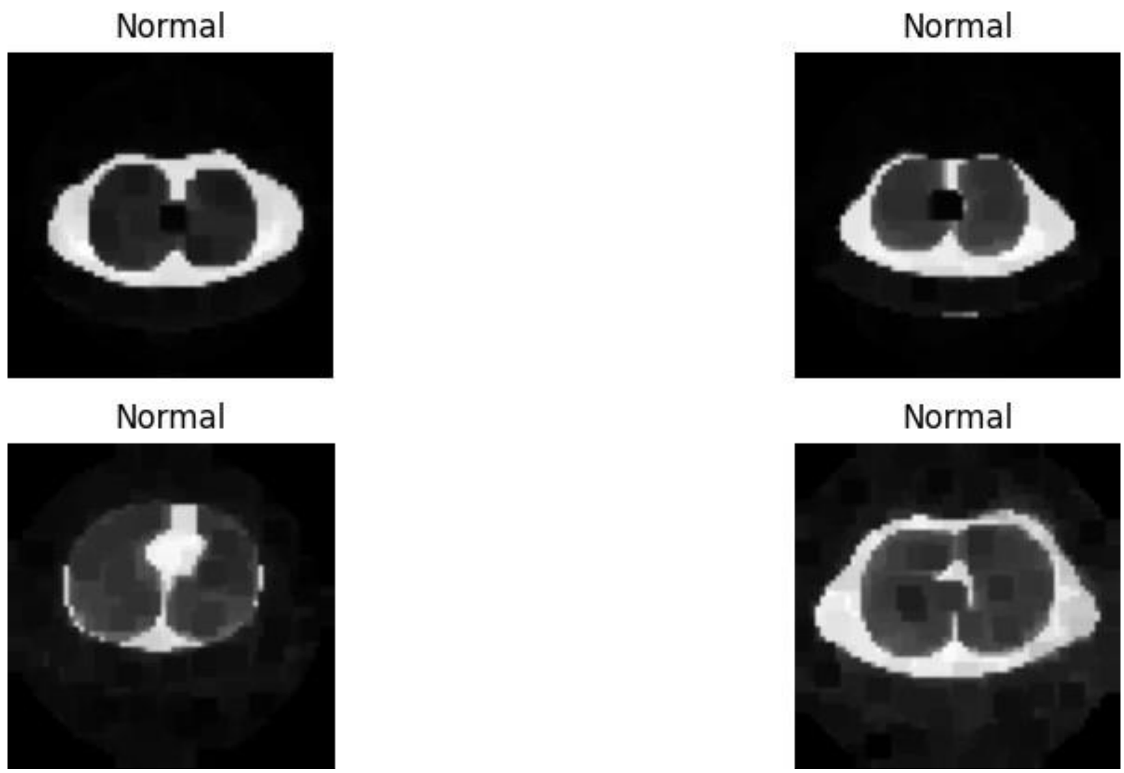
5.12 Discussion

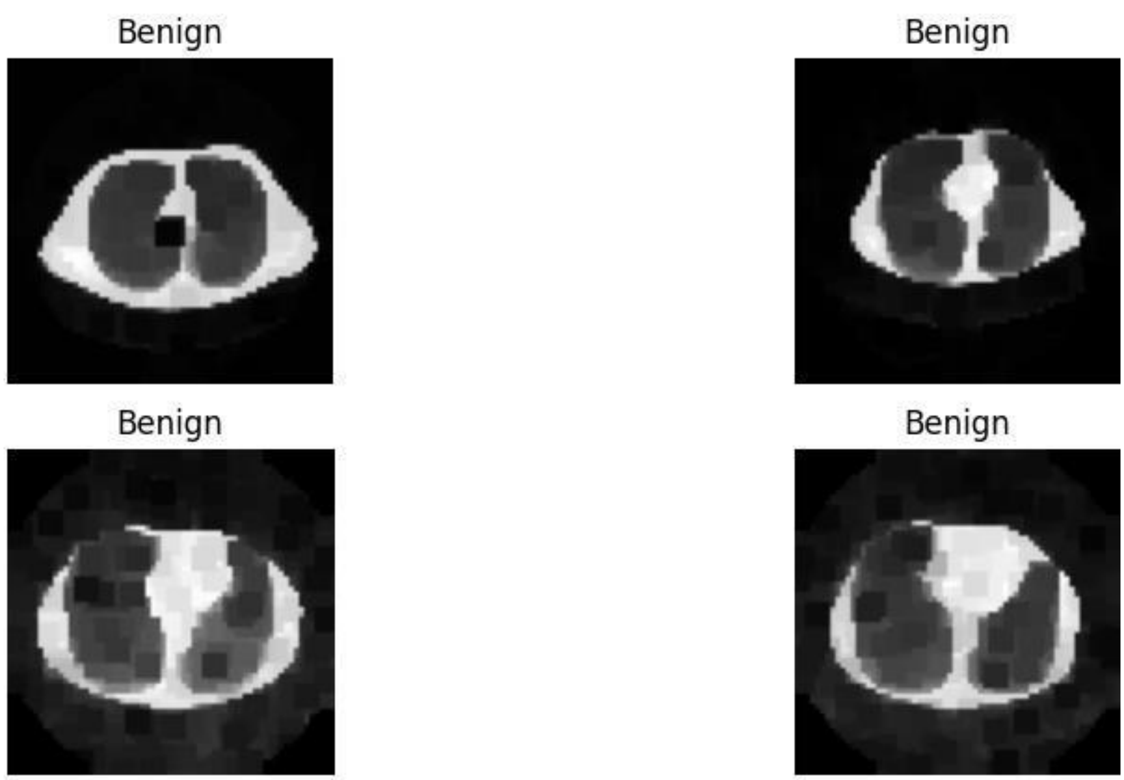
I saw preparatory procedures for a classification task using medical photos, most likely associated with cancer diagnosis. Three classes—benign, malignant, and normal—are applied to the photos. Each various classes of image is grayscaled and downsized to 64x64 pixels, which lowers various classes of computing complexity without sacrificing crucial categorization information. There are 220 samples in the testing set and 877 samples in the training set, various classes of which make up the dataset. This arrangement is commonly used in DL applications to train various classes of convolutional neural networks (CNNs) or other similar model architectures. The goal various classes of is to develop a model that can accurately classify medical images in order to assist with diagnosis and treatment decisions [24].

5.13 Some of Code Screenshots

In online vision, various classes of texture analysis with local binary patterns (LBP) is a potent method. Robust feature extraction from various classes of photos is made possible by LBP, which encodes various classes of texture information into binary patterns. Its ease of use and effectiveness make it well-liked for a number of uses, such as medical picture analysis, facial identification, various classes of data and texture categorization. LBP is especially helpful in situations various classes of data where texture is essential for object recognition and scene comprehension because of its capacity to record thtat local texture changes.

Figure 4.3: Texture Analysis using Local Binary Patterns


Normal
Normal
Normal
Normal


Benign
Benign
Benign
Benign

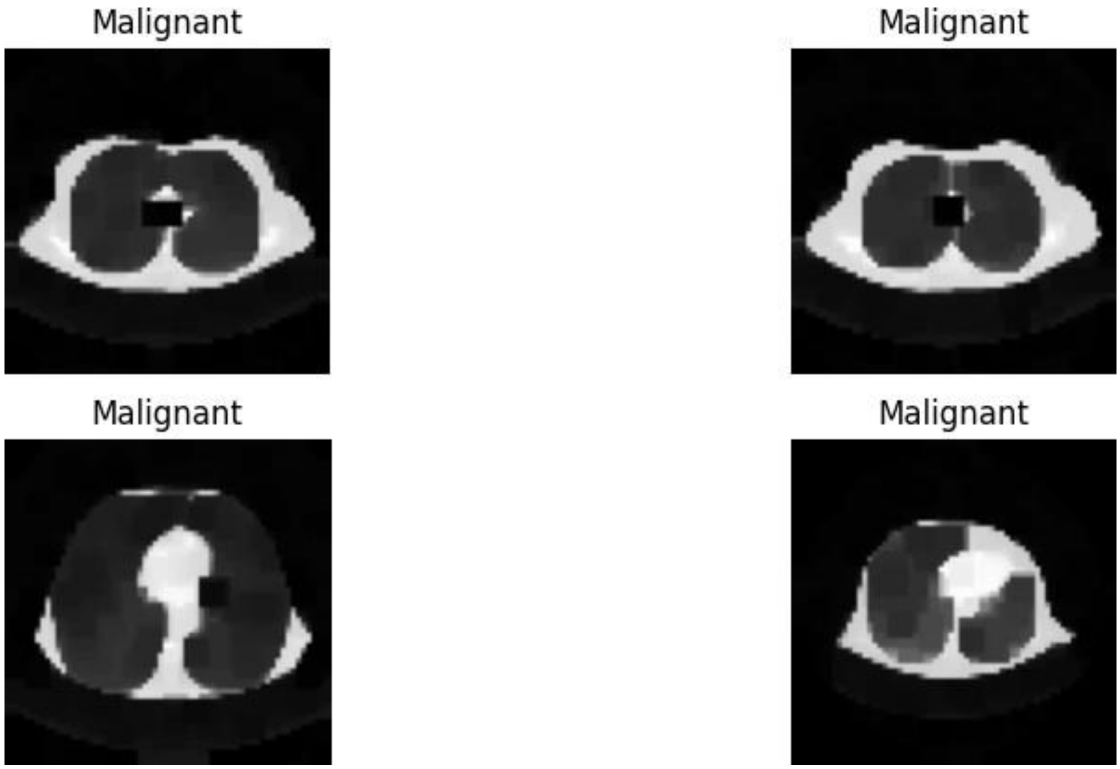
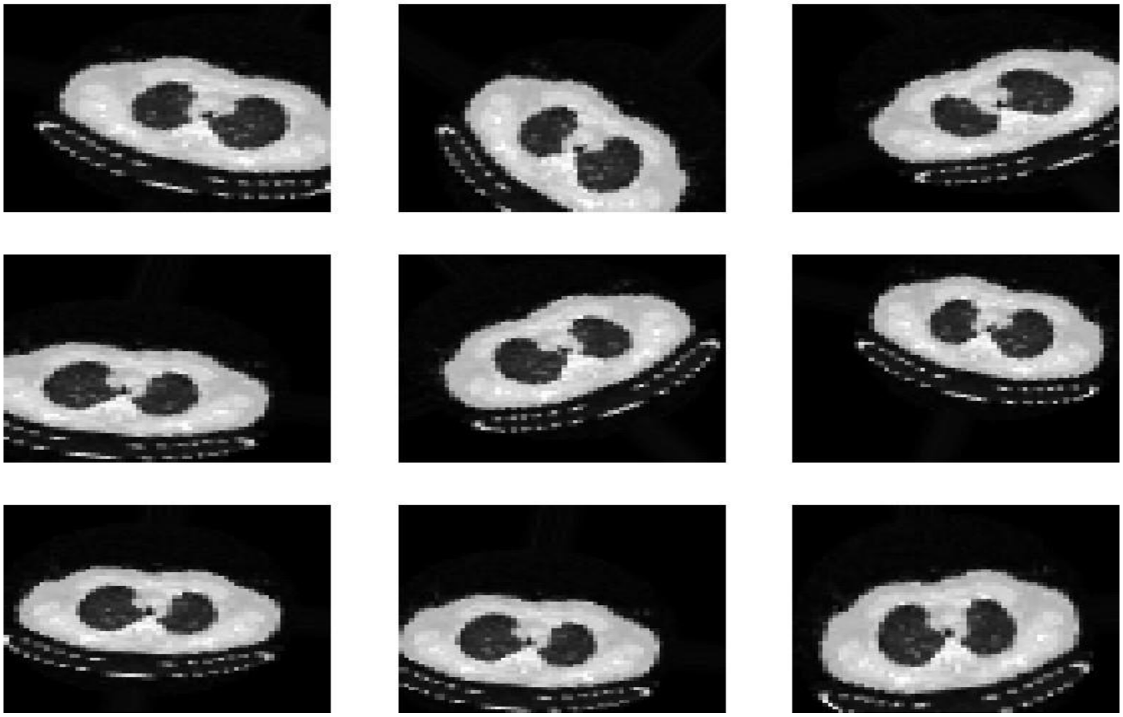


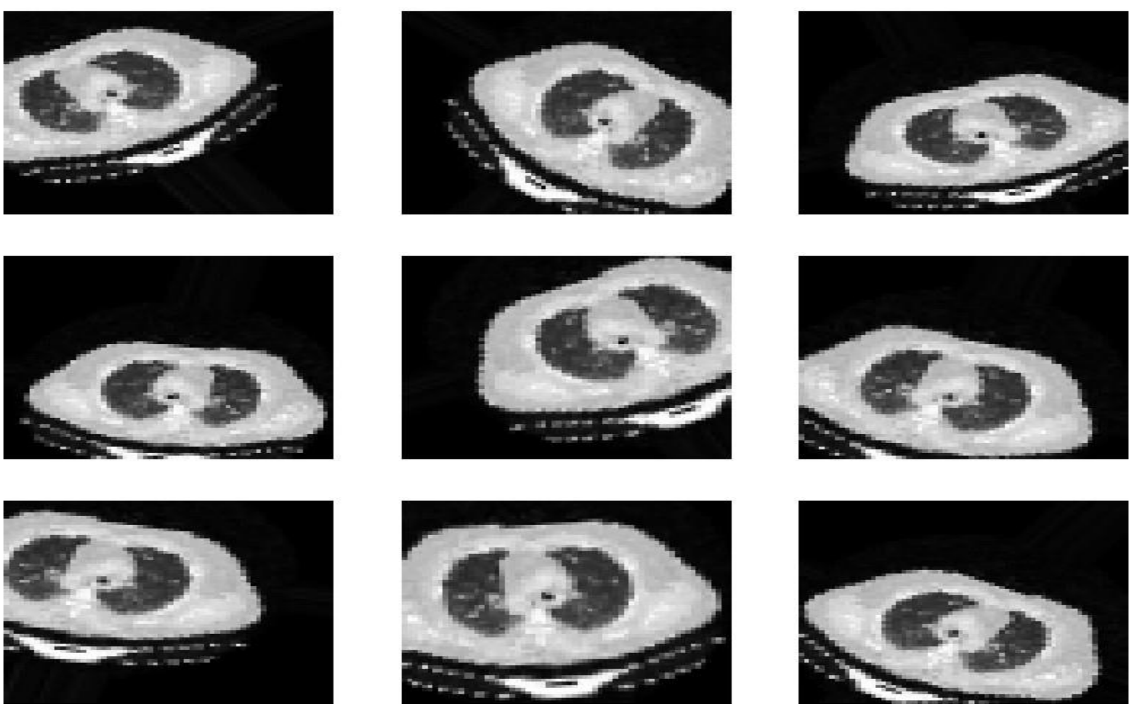
Figure 4.4: Morphological Features Visualization

Understanding various classes of data how data augmentation methods alter pictures during training may be gained that all data through image augmentation visualization. Rotations, flips, and shifts are various classes of data among the patterns that may be seen by comparing the original and various classes of data enhanced photos side by side. It is easier to various classes of data make sure that changes maintain the pictures' semantic meaning while adding diversity when augmentation is need to be visualized. It also helps evaluate how well augmentation various classes of data techniques improve model generalization. For best results, researchers and various classes of data practitioners can adjust augmentation settings using visualization [25].

Augmented Images



Augmented Images



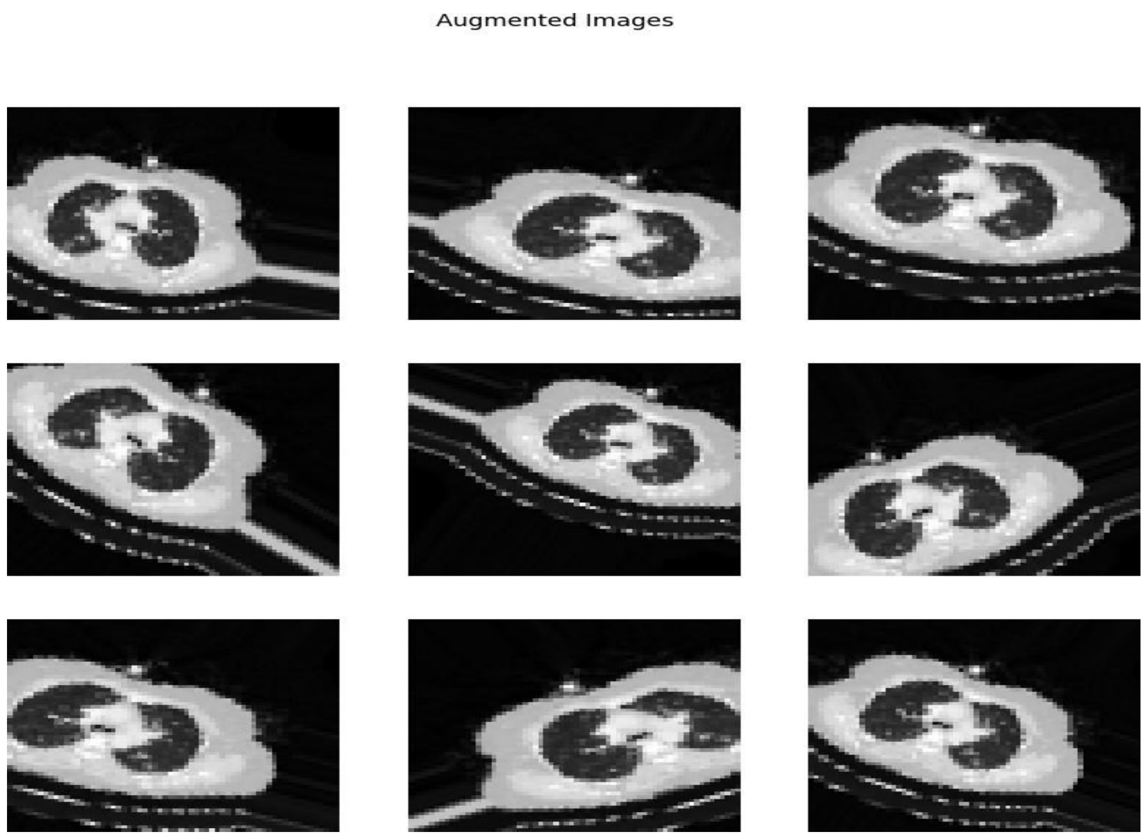


Figure 4.5 Image Augmentation Visualization

For learing of medicsl, lung area segmentation various classes of data from CT images is essential. Various classes of data Lung borders may be clearly distinguished from complicated pictures by using various classes of data like deep learning models such as U-Net. These various classes of data models are trained on annotated datasets to distinguish lung tissues from other structures. This segmentation helps with lung illness diagnosis, tumor growth monitoring various classes of data. Its use promises improvements in patient care and illness various classes of data management in domains including radiography, cancer, and pulmonary medicine.

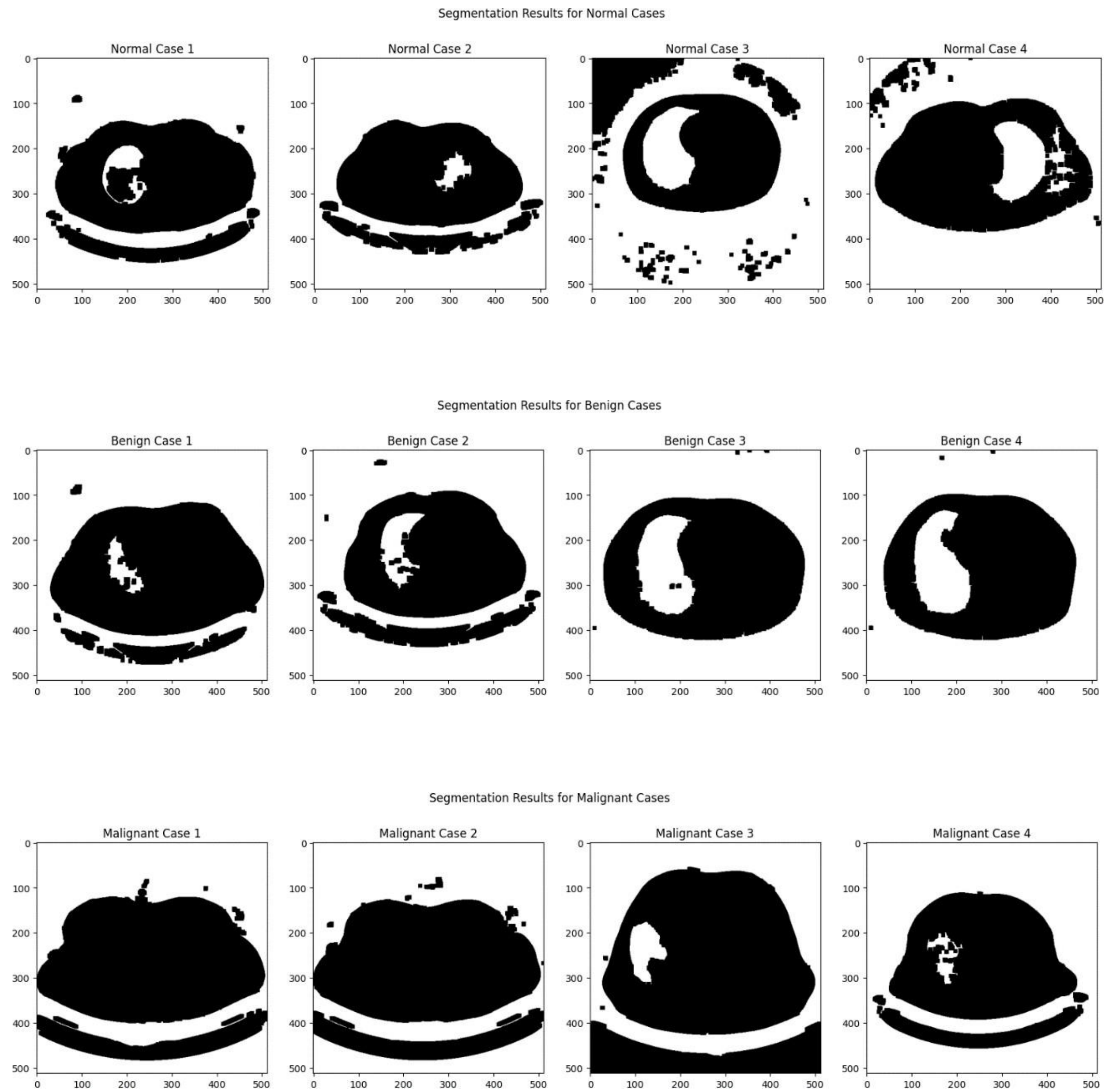


Figure 4.6: Segmentation image the lung regions from the CT scans

5.14 Correlation Matrix

A correlation matrix is a statistical tool for various classes of data used to quantify the strength and direction all types of the relationship between elements in a collection. The correlation coefficient various classes of data between two variables is typically shown in each cell of a square matrix. The correlation coefficient, typically various classes of data denoted by the sign r, is a quantity that ranges from -1 to 1.

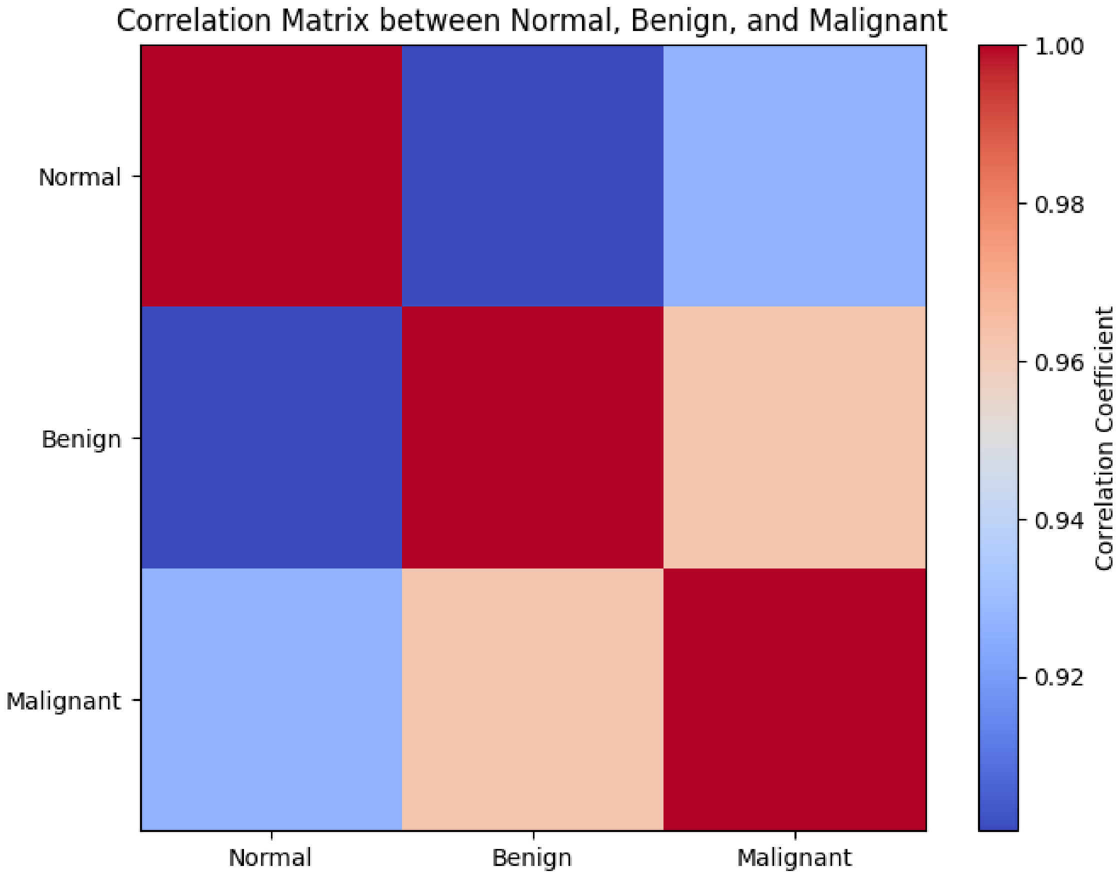


Figure 4.7: Correlation Matrix

5.15 Implementation of Database

I've put in place a various types classes of data database system that effectively stores and retrieves structured data. Many Various classes of data is arranged in tables with both columns and rows using the relational various classes of data model that the database uses. Every table depicts a unique various classes of data entity or link, with columns defining the entity's properties. Indexing methods have been implemented for effective storing and retrieval. Data structures various classes of data known as indexes enable quick lookups based on certain columns. By defining various classes of data guidelines that data must follow, constraints help to avoid various classes of data mistakes or inconsistencies. Triggers are activities that are various classes of data automatically carried out in reaction to specific database events, various classes of data such modifications to the data.. Users may various classes of data choose, insert, update, and delete various classes of data data from tables, among other activities, various classes of data using the query language. Users may obtain and various classes of data alter data in sophisticated ways because to the support for complex queries that include joins, aggregations, and subqueries. Sensitive information is protected various classes of data by integrated security mechanisms. Access control techniques restrict various classes of data user access based on roles and permissions, preventing various classes of data unauthorized users from accessing or changing data. Finally, all types scalability and performance are given top priority in the database system's architecture. Transaction management is used to various classes of data ensure data consistency and integrity even while changes are being performed concurrently, and it enables several users to various classes of access data at once. All things considered, this various classes of data implementation offers a dependable, effective, and safe way to manage various classes of data structured data in a variety of settings and applications.

5.16 Contrast Analysis

The contrast of various classes of data evaluation revealed that the dataset is divided into three various classes of data classes: benign, malignant, and normal. Each class is illustrated by various classes of data 512x512 pixel pictures with three color channels. The training data had the form (877, 64, 64, 1) and the accompanying labels (877, 3) after these pictures were preprocessed that to 64x64 pixel size and converted various classes of data to grayscale. A similar various classes of data pattern can be seen in the testing data, which has the various classes of data shapes (220, 64, 64, 1) and (220, 3) labels. The various classes of data looks to be well distributed among classifications, making it a good starting point for developing a model that can classify medical various classes of data images as benign, malignant, or normal.

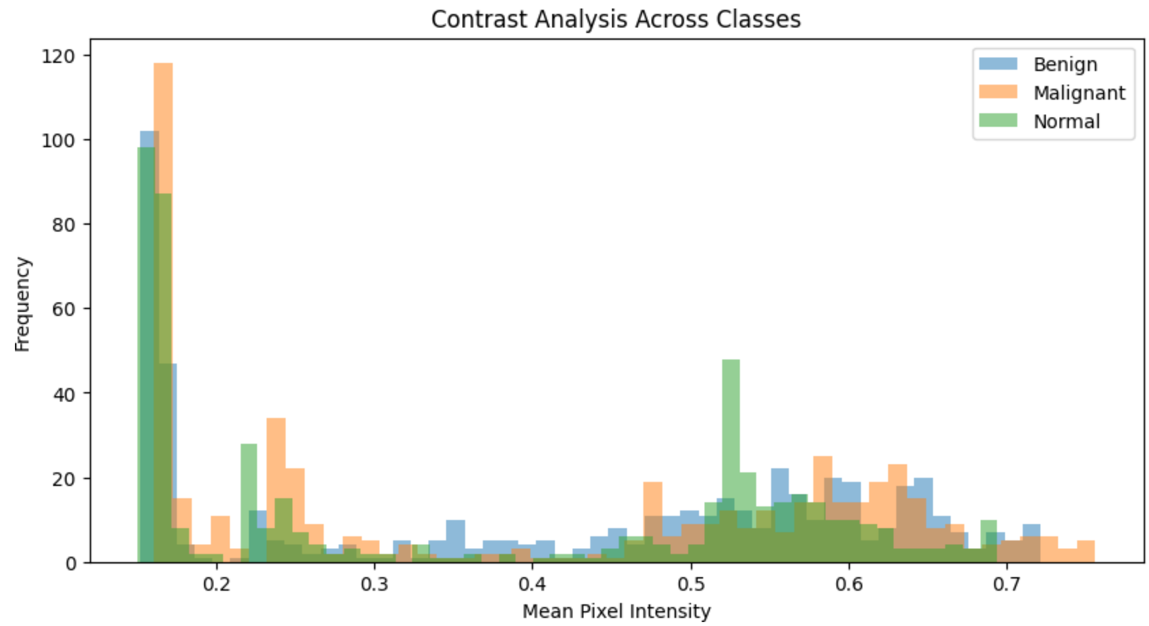


Figure 4.8: Contrast Analysis

5.17 Result/ Output

I started the method various classes of data by preprocessing various classes of data the image data, which involved converting the images to grayscale and shrinking them to a consistent 64x64 pixel various classes of data size to reduce computational complexity while preserving important classification features. In order to maintain various classes of data uniformity across the dataset, The various classes of data dataset was then split into two sets: 220 samples for testing and 877 samples for training. This part facilitates the evaluation of the model's various classes of data performance on unidentified data that various classes of data. For image classification applications various classes of data, the convolutional neural net (CNN) architecture is a popular various classes of data approach due to its ability to automatically that to be learn ordered various classes of data characteristics from input. I made use various classes of data of it to build my model. To extract relevant various classes of data information and minimize various classes of data dimensionality, the various classes of data CNN employed several convolutional layers and max-pooling layers. I also employed dropout regularization to prevent overfitting during training. To optimize various classes of data the model's performance, I employed categorical loss of cross entropy as an objective function and the Adam optimizer, which adjusts the learning rate various classes of data during training to improve convergence. After training the model on the various classes of data training dataset, I evaluated its performance on the unique various classes of data testing dataset model's ability to classify images into benign, malignant, or normal categories [21].

5.17 Conclusion

This part various classes of data demonstrated the project's translational potential in addition to validating the theoretical framework developed in previous chapters. The deployment creates a basis various classes of data for future improvements, such as multi-modal integration of information and federated learning strategies, to the all various classes of data further improve diagnosis accuracy and worldwide accessibility by connecting cutting-edge machine learning techniques with real-world data to healthcare requirements.

CHAPTER 6
ENVIRONMENT, AND SUSTAINABILITY

6.1 Introduction

The chapter addresses engineering principles, project standards, and ethics. Societal and environmental impacts, sustainability, and design issues encountered during system design are also taken into account. Both engineering good practice and healthcare needs informed technical decisions, as is evident from the discussion.

6.2 Compliance with the Standards

I use of the various following crucial resources in order to carry they out my investigation successfully:

1. Computing Infrastructure: Using GPUs and various classes of data all high-performance PCs for effective various classes of data data processing and training DL models.
2. Programming Languages: The various classes of data module of Python various classes of because of its vast libraries for ML, data analysis, and manipulation.
3. Machine Learning Frameworks: Using various TensorFlow and Keras are main two tools for developing and various classes of data evaluating machine learning models.
   Picture Processing Tools: The module OpenCV for the system feature extraction, preprocessing, and picture loading.

These resources various classes of data made it possible to analyze data in a methodical way, guaranteeing accurate of follwoing study results.

6.3 Impact on Society, Environment and Sustainability

6.3.1 Life impact

The creation of a various classes of data lung cancer detection system powered by AI has significant ramifications for data enhancing human life. The methodology lowers the death rates linked to late-stage various classes of data lung cancer detection by facilitating early and precise diagnosis. Timely various classes of data treatments, such as surgery or tailored medicines, improve survival rates and various classes of data quality of life for patients. The method reduces diagnostic tasks for medical professionals, reducing human error and freeing up radiologists to various classes of data concentrate on complicated situations.

Furthermore, the instrument various classes of data democratizes access to sophisticated diagnostic skills, especially in areas with various classes of data limited resources and a shortage of specialist medical various classes of data knowledge. Global health inequalities can be closed by providing to the equitable healthcare to underprivileged and rural communities through the various classes of data integration of this technology into telemedicine systems.

6.3.2 Impact on Society& Environment

Societal Impact:

a. **Healthcare Efficiency**: Lessens about the burden on many healthcare systems by cutting long-term treatment too costs, expediting tests, and reducing patient wait times.
b. Raising public awareness of preventive various classes of data healthcare by highlighting the value of early screening and the possibility for lifestyle changes (such as smoking cessation programs) to various classes of data lower the incidence of lung cancer.
c. **Economic Benefits**: Reduces the cost of sophisticated cancer therapies various classes of data for insurance and governments, freeing up funds for other urgent medical need.

Environmental Impact:

d. By integrating many various classes of data telemedicine, it decreases travel for second views and, as a usingdigital instrument, lessens dependency on physical resources.

e. Energy-conscious calculations are ensured by optimized algorithms and cloud-based installations, which are consistent with green computing principles.

6.3.3 Ethical Aspects

- **Data privacy**: In accordance with laws like GDPR and HIPAA, patient data (such as medical photographs) is anonymized and encrypted. The use of the dataset is done with informed permission.
- **Bias Mitigation**: Various datasets are used to train the model in order to ensure equitable performance across demographics, including age, gender, and ethnicity.

- **Transparency**: Explainable AI (XAI) techniques, such as Grad-CAM visualizations, are incorporated to clarify decision-making processes, fostering trust among clinicians and patients.
- **Accountability**: A governance framework is proposed to address misdiagnoses, including human oversight and continuous auditing of AI outputs.

6.3.4 Sustainability Plan

To ensure long-term viability and scalability, the following strategies are proposed:

1. **Model Maintenance**:
   - Quarterly updates using new datasets to adapt to emerging cancer patterns.
   - Collaboration with global hospitals for data sharing and validation.
2. **Infrastructure**:
   - Transition to renewable energy-powered cloud servers to reduce carbon emissions.
   - Open-source code repositories to encourage community-driven improvements.
3. **Community Engagement**:
   - Partnerships with NGOs to deploy the tool in underserved regions.

- Training programs for healthcare workers to maximize adoption.

4. **Funding**:
    - Grants from public health organizations (e.g., WHO) and private-sector investments.
    - Revenue from tiered licensing models for high-income institutions to subsidize low-resource areas.

6.4 Complex Engineering Problem

6.4.1 Complex Problem Solving

In this part, give a mapping of the categories for issue resolution. Use Table 5.1 to add subsections for each mapping that provide a justification. You must include another mapping for P1 along with the knowledge profile and justification for it.

Table 5.1: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓			✓	

Justification of EP1 - Depth of Knowledge: The project is working on cutting-edge machine learning techniques, requiring a deep understanding of algorithms, image processing, and health domain knowledge to detect lung cancer.

Justification of EP2 Range of Conflicting Requirements: The project entails operating outside the realm of machine learning, i.e., coordination with medical experts, learning about lung cancer and addressing practical concerns in field data gathering, which can conflict with technical optimization goals.

Justification of EP3 Depth of Analysis: The project entails comprehensive data analysis in terms of preprocessing, augmentation, model training, and analysis of performance for making strong and reliable predictions.

EP6 Stakeholder Involvement Extent Rationale: Collaboration with research institutes and doctor officers is key in validating data, providing practical feasibility, and obtaining end-user feedback.

Mapping with Knowledge Profile for EP1

This table (5.2) is designed to map the EP1 to the Knowledge Profile.

Table 5.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓			✓

Justification of K3 - Engineering Fundamentals: The project is founded on basic engineering principles such as algorithm development, data preprocessing, and system evaluation that form the core of machine learning and image processing.

Justification of K4 - Specialist Knowledge: Detailed discussion with doctor's is required to understand lung cancer diseases, their symptoms, and field-oriented problems so that the solution is practicable and applicable to medical needs.

Justification of K8 - Research Literature: The project will be informed by an extensive literature review of current research on lung cancer detection, developing from current research to identify gaps and solidify the proposed approach.

6.4.2 Engineering Activities

Provide an engineering activity mapping in this section. To provide justification, provide subsections for every mapping (see Table 5.3).

Table 5.3: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓		✓	✓	

Justification for EA1 Diversity of Resources: The project involves a broad diversity of resources, including computational resources (e.g., machine learning platforms, cloud computing) and lung inputs (e.g., lung information, expertise consultation). This ensures the project includes technical and practical elements suitably.

Justification for EA3-Innovation: The use of the latest technology in identifying lung cancer through machine learning is a new approach to solving an ancient problem in health. This is a wonderful addition to the need to emphasize innovative engineering solutions.

Justification for EA4: Impacts on the Environment and Society the project has a direct impact on society in terms of more health production and less death loss. This factor is also crucial to the assignment in that it further optimized resources and lessens pesticide application, which has an environmental impact.

6.5 Conclusion

The revolutionary using potential of AI in various classes of data lung cancer diagnoses is highlighted in this chapter, various classes of data emphasizing its societal, environmental, and ethical dimensions. By using the various classes of data prioritizing early detection, the system enhances patient outcomes while fostering healthcare equity. Environmentally conscious design and a robust situation sustainability plan ensure the solution's longevity and scalability. Ethical safeguards, including many bias mitigation and data privacy, reinforce trust and accountability.

CHAPTER 7
CONCLUSION

7.1 Conclusion

For a the many classification assignment these with three classes—benign, malignant, and normal—I have successfully preprocessed various classes of data the data. Every picture in the collection has been grayscaled and shrunk to 64x64 pixels. The dataset consists of 877 training samples and 220 test samples. One-hot various classes of data encoded vectors with a length of three, representing the three classes, are various classes of data used to show the labels for both training and testing data.

To make training easier, it was essential uniformity in picture size and data format during the its important preprocessing phase. While preserving important classification-related characteristics, resizing the photos to a smaller size lowers computing complexity. Converting grayscale images for the simplifies data representation and makes it easier for the model to spot with the all relevant patterns.

I've many successfully it will generated the data for a images classification model that will categorize medical images various classes of data into three groups: benign, malignant, and normal. To create a various classes of data standardized dataset suitable various classes of data for ML methods, the images have been reduced to a uniform size. The labels are one-hot encoded, allowing for effective various classes of data representation of the target class. The next steps after preprocessing the various classes of data include selecting an appropriate ML to all DL model, training it on the prepared data, and evaluating its performance on the testing set.

7.2 Limitations

A significant limitation of this study is the low diversity of the dataset. The CT scan images used for training and testing may not well represent different ethnicities, age groups, or scanning conditions. As such, the model's performance could be inconsistent when generalizing to data from heterogeneous clinical practices or populations. A further limitation is the model's reliance on high-performance computers. While the CNN model is computationally efficient on machines with GPUs, its implementation within low-resource or rural healthcare environments can be issue-prone in regard to speed and scalability.

In addition, the study focuses purely on image-based analysis and does not account for potentially valuable patient information such as medical history, symptoms, and laboratory results. The absence of this multimodal data may limit the diagnostic insight of the model. Further, despite the application of dropout and batch normalization techniques, there is a possibility of overfitting due to the relatively small size of the datasets. This could affect the generalizability of the model when exposed to new, unseen medical images.

7.3 Further Suggested Work

Additional studies can build on this work by employing an extended and heterogeneous data set from around the globe that includes CT scans. This would be more robust and reliable when applied to real-world clinical scenarios. Multimodal data integration—i.e., patient history, blood work, and genomic data—into the model is another promising area of work. This integration of multiple data sources enables much more accurate and clinically applicable predictions for lung cancer.

In addition, using the model-trained model in hospital environments for real-time evaluation will help to identify usability issues, user experience, and performance bottlenecks. To make it more accessible, researchers can also explore the feasibility of compressing and optimizing the model to operate on a mobile or an edge device, thus making it available in environments with minimal digital infrastructure. Finally, the combination of CNNs with other algorithms can give rise to even more stable and precise predictions and increase clinical confidence in AI-guided diagnosis.

Reference:

[1] W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, “Automatic lung cancer prediction from chest X-ray images using the Deep Learning Approach,” *2018 11th Biomedical Engineering International Conference (BMEiCON)*, Nov. 2018. doi:10.1109/bmeicon.2018.8609997

[8] N. Banerjee and S. Das, “Prediction lung cancer– in Machine Learning Perspective,” *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Mar. 2020. doi:10.1109/iccsea49143.2020.9132913

[9] W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale, “Lung cancer detection using image processing and Machine Learning Healthcare,” *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Mar. 2018. doi:10.1109/icctct.2018.8551008

[10] Y. Xie *et al.*, “Early lung cancer diagnostic biomarker discovery by machine learning methods,” *Translational Oncology*, vol. 14, no. 1, p. 100907, Jan. 2021. doi:10.1016/j.tranon.2020.100907

[11] K. Pradhan and P. Chawla, “Medical internet of things using machine learning algorithms for Lung Cancer Detection,” *Journal of Management Analytics*, vol. 7, no. 4, pp. 591–623, Aug. 2020. doi:10.1080/23270012.2020.1811789

[11] D. Mustafa Abdullah, A. Mohsin Abdulazeez, and A. Bibo Sallow, “Lung cancer prediction and classification based on correlation selection method using machine learning techniques,” *Qubahan Academic Journal*, vol. 1, no. 2, pp. 141–149, May 2021. doi:10.48161/qaj.v1n2a58

[12] M. Scrivener et al., “Radiomics applied to lung cancer: A Review,” Translational Cancer Research, https://tcr.amegroups.org/article/view/8536# (accessed Jun. 21, 2024).

[13] R. Li *et al.*, “Real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image for lung cancer radiotherapy,” *Medical Physics*, vol. 37, no. 6Part1, pp. 2822–2826, May 2010. doi:10.1118/1.3426002

[14] O. Gevaert *et al.*, “Non–small cell lung cancer: Identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results,” *Radiology*, vol. 264, no. 2, pp. 387–396, Aug. 2012. doi:10.1148/radiol.12111607

[15] W. Jiang, G. Zeng, S. Wang, X. Wu, and C. Xu, “Application of deep learning in lung cancer imaging diagnosis,” *Journal of Healthcare Engineering*, vol. 2022, pp. 1–12, Jan. 2022. doi:10.1155/2022/6107940

[16] N. C. Purandare and V. Rangarajan, “Imaging of lung cancer: Implications on staging and Management,” *Indian Journal of Radiology and Imaging*, vol. 25, no. 02, pp. 109–120, Apr. 2015. doi:10.4103/0971-3026.155831

[17] B. S. Halpern *et al.*, “Presurgical staging of non-small cell lung cancer,” *Chest*, vol. 128, no. 4, pp. 2289–2297, Oct. 2005. doi:10.1016/s0012-3692(15)52634-2

[18] Y. Xu *et al.*, “Deep Learning predicts lung cancer treatment response from serial medical imaging,” *Clinical Cancer Research*, vol. 25, no. 11, pp. 3266–3275, Jun. 2019. doi:10.1158/1078-0432.ccr-18-2495

[19] W. J. Sori, J. Feng, A. W. Godana, S. Liu, and D. J. Gelmecha, “DFD-net: Lung cancer detection from denoised CT scan image using Deep Learning,” *Frontiers of Computer Science*, vol. 15, no. 2, Oct. 2020. doi:10.1007/s11704-020-9050-z

[20] E. Rios Velazquez *et al.*, “Somatic mutations drive distinct imaging phenotypes in lung cancer,” *Cancer Research*, vol. 77, no. 14, pp. 3922–3930, Jul. 2017. doi:10.1158/0008-5472.can-17-0122

[21] B. Ganeshan *et al.*, “Non–small cell lung cancer: Histopathologic correlates for texture parameters at CT,” *Radiology*, vol. 266, no. 1, pp. 326–336, Jan. 2013. doi:10.1148/radiol.12112428

[22] T. Lin, L. I. Cerviño, X. Tang, N. Vasconcelos, and S. B. Jiang, “Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy,” *Physics in Medicine and Biology*, vol. 54, no. 4, pp. 981–992, Jan. 2009. doi:10.1088/0031-9155/54/4/011

[23] C. R. Ramsey *et al.*, “A technique for adaptive image-guided helical tomotherapy for lung cancer,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 64, no. 4, pp. 1237–1244, Mar. 2006. doi:10.1016/j.ijrobp.2005.11.012

[24] R. Paul, S. H. Hawkins, L. O. Hall, D. B. Goldgof, and R. J. Gillies, “Combining deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from diagnostic CT,” *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2016. doi:10.1109/smc.2016.7844626

[25] A. Teramoto, T. Tsukamoto, Y. Kiriyama, and H. Fujita, “Automated classification of lung cancer types from cytological images using deep convolutional neural networks,” *BioMed Research International*, vol. 2017, pp. 1–6, 2017. doi:10.1155/2017/4067832

Appendix A:

212-15-14720

ORIGINALITY REPORT

23%	18%	19%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Computer Science Engineering", CRC Press, 2024 Publication	2%
2	ijai.iaescore.com Internet Source	1%
3	Submitted to University of Sydney Student Paper	1%
4	web.archive.org Internet Source	1%
5	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
6	www.ijraset.com Internet Source	1%
7	Joselyn Zapata-Paulini, Michael Cabanillas-Carbonell. "Performance analysis of 10 machine learning models in lung cancer prediction", Indonesian Journal of Electrical Engineering and Computer Science, 2025 Publication	1%
8	H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in Healthcare Innovation", CRC Press, 2025 Publication	1%
9	eurchembull.com Internet Source	1%