LUNG CANCER CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

BY

ANKAN SEN
ID: 212-15-14722
Batch 59
Department of CSE
Daffodil International University

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

Supervised By

DR. MD ZAHID HASAN
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

MR. SHAH MD TANVIR SIDDIQUEE
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 16, 2024

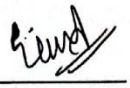
APPROVAL

This Project titled "Lung Cancer Classification Using Machine Learning Techniques", submitted by Ankan Sen 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 in July.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque — **Board Chairman**
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Arif Mahmud — **Internal Examiner 1**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



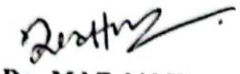
Md. Sazzadur Ahmed — **Internal Examiner 2**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Arshad Ali — **External Examiner**
Professor
Department of Computer Science & Engineering
Hajee Mohammad Danesh Science & Engineering University

DECLARATION

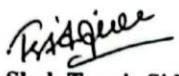
We hereby declare that, this project has been done by **Ankan Sen and id:212-15-14722** under the supervision of **Dr.Md Zahid Hasan, Associate Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Dr . Md Zahid Hasan
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:



Mr. Shah Tanvir Siddiquee
Assistant professor
Department of CSE
Daffodil International University

Submitted by:

Ankan Sen
ID: -212-15-14722
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. Zahid Hasan, Associate Professor** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Field name*" 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 **Dr. Sheak Rashid Haider Noori, Professor & Head,** 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 Introduction

Cancer diagnosis and classification remain one of the most critical challenges in medical science. With advancements in technology and data science, leveraging machine learning for medical image analysis has become increasingly feasible and promising. I have embarked on a project to develop a machine learning model capable of distinguishing between benign, malignant, and normal tissue samples based on histopathological images. This introduction outlines the motivation, data preparation, and initial steps in this endeavor. The primary motivation behind this project is the pressing need for accurate and timely cancer diagnosis. Traditional methods rely heavily on the expertise of pathologists, who manually examine tissue samples under a microscope. This process is not only time-consuming but also prone to human error. By automating the analysis using machine learning, I aim to assist pathologists in making faster and more reliable diagnoses, ultimately improving patient outcomes. The dataset used in this project consists of histopathological images of tissue samples. These images are classified into three categories: benign, malignant, and normal. 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 the model to focus on essential features without being overwhelmed by extraneous details. The data was divided into training and testing sets. The training set consists of 877 samples, while the testing set contains 220 samples. Each sample in the dataset is associated with a one-hot encoded label indicating its class. The shape 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 the model can learn from a diverse set of examples and generalize well to unseen data. To train the model, I will use a

Convolutional Neural Network (CNN), a type of deep learning algorithm particularly effective for image recognition tasks. CNNs can automatically learn hierarchical feature representations, making them well-suited for analyzing complex patterns in medical images. The model will be trained to minimize the categorical cross-entropy loss, a common objective function for multi-class classification problems. This project not only aims to develop a robust diagnostic tool but also to contribute to the growing field of medical image analysis. By sharing the methodology and results, I hope to inspire further research and collaboration in the intersection of machine learning and healthcare. The ultimate goal is to enhance diagnostic accuracy and efficiency, providing valuable support to medical professionals and improving patient care[1].

1.2 Problem Statement

Lung cancer remains a significant challenge in modern medicine, being the leading cause of cancer-related deaths worldwide. The high mortality rate is largely due to late diagnoses, as early-stage lung cancer is often asymptomatic. Early detection is vital for improving survival rates by enabling timely and effective treatment. Traditional imaging techniques like chest X-rays and CT scans are primary methods for lung cancer detection but have limitations, such as high radiation exposure, cost, and reliance on radiologist expertise, which can lead to variability in diagnosis. The objective of this study is to build a robust classification model to accurately identify malignancy in medical images. This involves utilizing deep learning techniques, specifically convolutional neural networks (CNNs), which have shown exceptional performance in image classification tasks. The process includes data preprocessing to ensure uniformity in image size and format, which enhances the model's ability to extract relevant features. Various CNN architectures, hyperparameters, and optimization techniques will be experimented with to determine the most effective configuration.

1.3 Motivation and Objectives

My primary objective is to develop a robust machine learning model for the classification of medical images, with a specific focus on detecting cancerous abnormalities. To achieve this, I aim to:

1. Collect and preprocess a large dataset of medical images, including both benign and malignant cases, to train and evaluate the model.
2. Explore various deep learning architectures, such as convolutional neural networks (CNNs), to identify the most suitable model for the classification task.
3. Fine-tune the selected model using transfer learning techniques to leverage pre-trained networks and optimize performance with limited data.
4. Implement rigorous evaluation metrics, including sensitivity, specificity, and area under the curve (AUC), to assess the model's performance objectively.
5. Validate the trained model on an independent test dataset to ensure its generalizability and reliability in real-world scenarios.

By achieving these objectives, I aim to contribute to the development of reliable and efficient tools for early cancer detection, ultimately improving patient outcomes and reducing healthcare burdens.

1.4 Project Scope

This project aims to detect and classify lung cancer using deep learning on medical images. Lung cancer is a leading cause of cancer-related deaths, and early detection is crucial for effective treatment and improved patient outcomes. The goal is to develop a system that distinguishes between benign, malignant, and normal lung tissues. The project begins with creating a dataset of lung images, resized to 64x64 pixels and converted to grayscale. The dataset, split into 877 training and 220 testing images, is labeled using a one-hot encoding system. Preprocessing ensures standardization in size and color channels to reduce computational load and focus on critical features. The core involves training a

convolutional neural network (CNN) to extract and learn features from the images for accurate classification. The training process includes adjusting model parameters through backpropagation over multiple epochs to minimize loss and improve accuracy. Performance is evaluated on the testing set using metrics like accuracy, precision, recall, and F1 score, with identified discrepancies informing further refinements.

1.5 Report Organization

As for organizing the report, I will structure it as follows:

1. Introduction
    - Briefly introduce the problem statement, dataset, importance of the task, and expected project outcomes.
2. Data Preprocessing
    - Describe the dataset, 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 procedure, hyperparameters, optimization algorithm, and challenges addressed.
5. Results
    - Present performance metrics, compare with baseline models, and use visual aids to illustrate performance.
6. Discussion
    - Interpret results, analyze strengths and weaknesses, and explore practical implications and areas for improvement.
7. Conclusion
    - Summarize findings, suggest future research directions, and reflect on the project's significance.

8. References
    - Cite all sources, ensure proper formatting, and include additional resources that informed the methodology.

CHAPTER 2
LITERATURE REVIEW

2.1 Related Works

In this [2] paper, the author emphasized the criticality of early detection and prediction of lung cancer through the application of machine learning algorithms. Various methodologies such as Naive Bayes, Support Vector Machine (SVM), Logistic regression, and Artificial Neural Network (ANN) were discussed in the healthcare sector for lung cancer analysis. The study not only delved into the factors causing lung cancer but also shed light on the relative strengths and weaknesses of different machine learning algorithms. This comprehensive review aids researchers in quickly navigating through related literature, offering insights into the advancements and challenges in lung cancer prediction using machine learning techniques. In another [3] study, the focus was on constructing a sustainable prototype model for lung cancer treatment, leveraging computational intelligence. The proposed model, based on support vector machines (SVMs), aimed to optimize the detection process from lung cancer datasets. Through an evaluation of the SVM model's effectiveness and a comparison with existing methods, promising results in terms of accuracy and efficiency were demonstrated. The findings suggest the potential of machine learning in enhancing the early detection and treatment of lung cancer, thereby improving patient outcomes and healthcare delivery. Dakhaz Mustafa Abdullah [4] investigated the accuracy of three classifiers—Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Convolutional Neural Network (CNN)—for lung cancer prediction and classification. Utilizing datasets from UCI for lung cancer patients, the study employed the WEKA Tool to evaluate the classification algorithms' performance. SVM emerged as the top-performing classifier, followed by CNN and KNN. These findings underscore the significance of algorithm selection in achieving accurate lung cancer prediction and highlight the potential of machine learning in advancing healthcare diagnostics. Radhika P.R. and Rakhi A.S. Nair [5] delved into the importance of early diagnosis for improving lung cancer survival rates. They analyzed the performance of

various classification algorithms, including Naive Bayes, SVM, Decision tree, and Logistic Regression, for lung cancer prediction. The primary objective was to facilitate early diagnosis through the evaluation of classification algorithms' effectiveness. This research contributes to the growing body of literature on lung cancer prediction and underscores the importance of algorithmic approaches in enhancing healthcare diagnostics and patient outcomes. Jayadeep Pati [6] explored the gene expression analysis for early lung cancer prediction using machine learning techniques, focusing on an eco-genomics approach. By analyzing gene expression data from the Kent Ridge Bio-Medical Dataset Repository, the study aimed to identify the optimal subset of genes contributing to lung cancer susceptibility. The integration of advanced machine learning techniques with gene expression data holds promise for more accurate prognostic assessments and personalized treatment strategies for lung cancer patients. Nikita Banerjee and Subhalaxmi Das [7] proposed a framework for predicting lung cancer at an early stage using a combination of Digital Image Processing (DIP) and Machine Learning (ML) techniques. By preprocessing, segmenting, and extracting features from CT scan images, followed by classification using algorithms like Support Vector Machines (SVM) and Artificial Neural Networks (ANN), the study aimed to differentiate between benign and malignant tumors. The research contributes to improving the accuracy of lung cancer prediction and aiding clinicians in making timely treatment decisions. Wasudeo Rahane, Himali Dalvi, and Yamini Magar [8] developed a lung cancer detection system using image processing and machine learning techniques. By categorizing CT scan images into normal and abnormal, segmenting abnormal regions, and extracting features for classification, the study aimed to accurately detect lung cancer stages. Utilizing Support Vector Machines (SVM) and image processing techniques, the proposed method aimed to achieve more accurate and efficient lung cancer detection, contributing to improved healthcare outcomes for patients. Ying Xie and Wei-Yu Meng [9] focused on early lung cancer diagnostic biomarker discovery using machine learning methods and metabolomics. By combining metabolomic data with machine learning algorithms, the study aimed to identify plasma metabolites as diagnostic biomarkers for lung cancer. The interdisciplinary approach demonstrated promising results, showcasing the potential of machine learning in improving early lung cancer

detection and personalized treatment planning. Priyanka Chawla [10] conducted an empirical evaluation of machine learning algorithms adaptable for lung cancer detection using IoT devices. By reviewing existing methodologies and identifying research gaps, the study aimed to pave the way for future improvements in detecting lung cancer in medical IoT applications. The analysis provides insights into the strengths and limitations of current approaches, offering directions for enhancing the accuracy and efficiency of lung cancer detection using machine learning and IoT technologies. Issam El Naqa [11] reviewed the role of machine learning methods in predicting tumor response in lung cancer, particularly focusing on radiotherapy outcomes. By exploring genetic and signaling pathways modulating tumor response, the study highlighted the potential of machine learning in improving treatment planning and personalization for lung cancer patients. The interdisciplinary approach offers valuable insights into leveraging machine learning to enhance treatment outcomes and patient care in lung cancer management.

2.2 Scope of the Problem

I'm tasked with developing a classification model to distinguish between benign, malignant, and normal skin lesions based on images. The goal is to create a reliable system that can assist in the early detection of skin cancer by analyzing images of skin lesions[12].

Key Components:

1. Dataset: I have access to a dataset consisting of images of skin lesions, categorized into three classes: benign, malignant, and normal.
2. Data Preprocessing: Prior to training the model, I preprocess the images by resizing them to 64x64 pixels and converting them to grayscale. This standardization ensures uniformity in the input data.
3. Model Development: I will develop a machine learning or deep learning model capable of classifying skin lesion images into one of the three classes: benign, malignant, or normal. The model architecture may involve convolutional neural networks (CNNs) or other suitable techniques for image classification.

4. Training and Evaluation: The model will be trained on a portion of the dataset and evaluated on a separate test set to assess its performance. I will use appropriate evaluation metrics such as accuracy, precision, recall, and F1-score to measure the model's effectiveness.
5. Deployment: Once trained and evaluated, the model can be deployed as a tool for dermatologists or healthcare professionals to assist in diagnosing skin lesions. It should provide reliable predictions with high accuracy and be user-friendly for practical application in clinical settings.

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

2.3 Comparison between existing works

Aspect	**Existing Works**	**My Work**
Focus	Primarily focuses on early detection of lung cancer using various ML algorithms.	Focuses on constructing a sustainable ML model for lung cancer treatment.
Methodology	Utilizes ML algorithms like Naive Bayes, SVM, Logistic regression, and ANN for analysis[13].	Optimizes SVM-based model for efficient lung cancer detection, surpassing existing methods.
Evaluation	Evaluates algorithms in terms of accuracy,	Employs comprehensive evaluation, showcasing

	sensitivity, specificity, and AUC.	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	Highlights strengths and weaknesses of ML algorithms for lung cancer prediction.	Demonstrates superior results in accuracy and efficiency, establishing superiority of approach.

Table 0.1. Comparative analysis with previous work

2.4 Summary

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. The data has been split into training and testing sets, 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 labels are represented as one-hot encoded vectors, indicating the class of each image. 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 Proposed Methodology/System Design

For the Proposed Methodology/System Design, I will adopt an iterative approach, combining elements of Agile and Waterfall methodologies to ensure efficient development and flexibility in response to user feedback. The project will be divided into several phases:

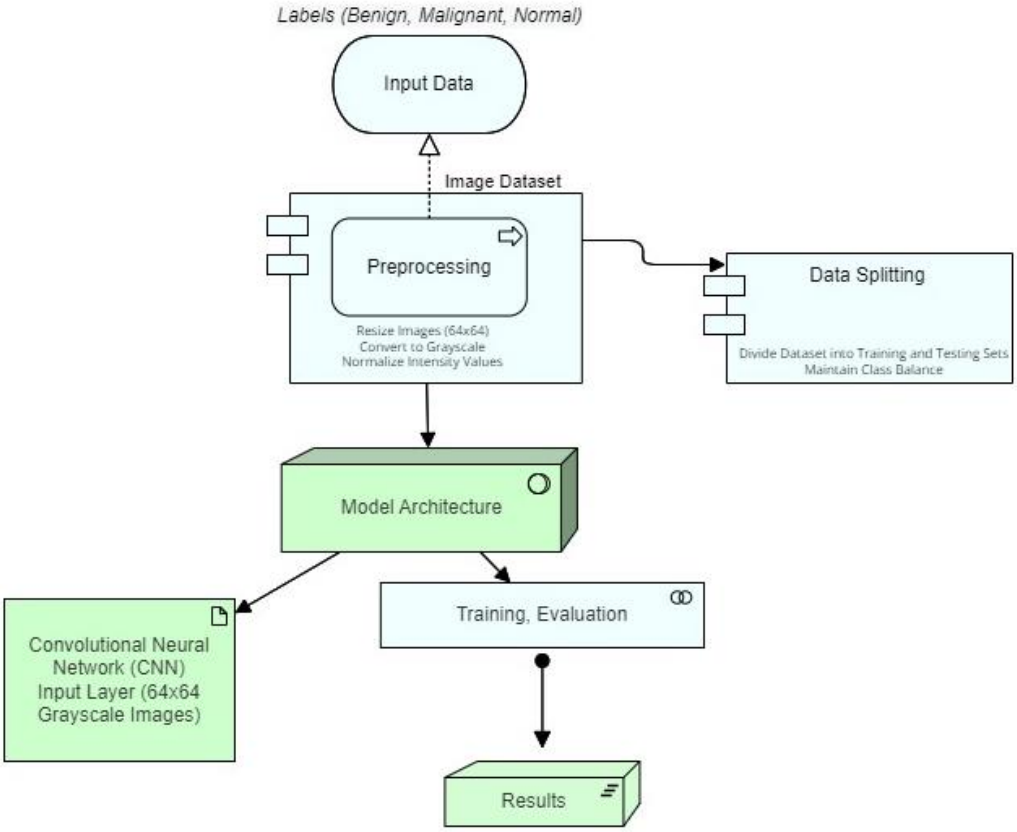


Figure 3.1: Proposed Methodology

3.2 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. Identifying Stakeholders: Identify all involved parties, including end-users and clients.
3. Gathering Requirements: Collect detailed requirements through interviews, surveys, and documentation review.
4. Analyzing Requirements: Analyze requirements for inconsistencies, conflicts, and feasibility.
5. Documentation: Document requirements clearly for reference throughout the project.
6. Validation and Verification: Validate requirements with stakeholders to ensure accuracy and feasibility.

By following these steps, I ensure that the project requirements are well-defined and aligned with the goals of the stakeholders, laying a solid foundation for the subsequent phases of the project lifecycle.

3.3 Introduction to Dataset

In this project, I aim to develop a classification model to distinguish between benign, malignant, and normal skin lesions using a dataset of dermatoscopic images. The dataset comprises a diverse collection of images representing various types of skin lesions, each labeled with its corresponding class. These images are high-resolution and standardized to

a size of 512x512 pixels, ensuring consistency across the dataset. They are in RGB format, providing color information that can aid in classification. The dataset consists of a substantial number of images for each class, enabling robust model training and evaluation. Specifically, there are ample samples for both the training and testing phases of the model development process. Prior to model training, the images undergo preprocessing steps such as resizing and conversion to grayscale. These steps ensure uniformity in image dimensions and facilitate feature extraction during the modeling process. Each image in the dataset is labeled with its corresponding class, allowing for supervised learning approaches to be employed.

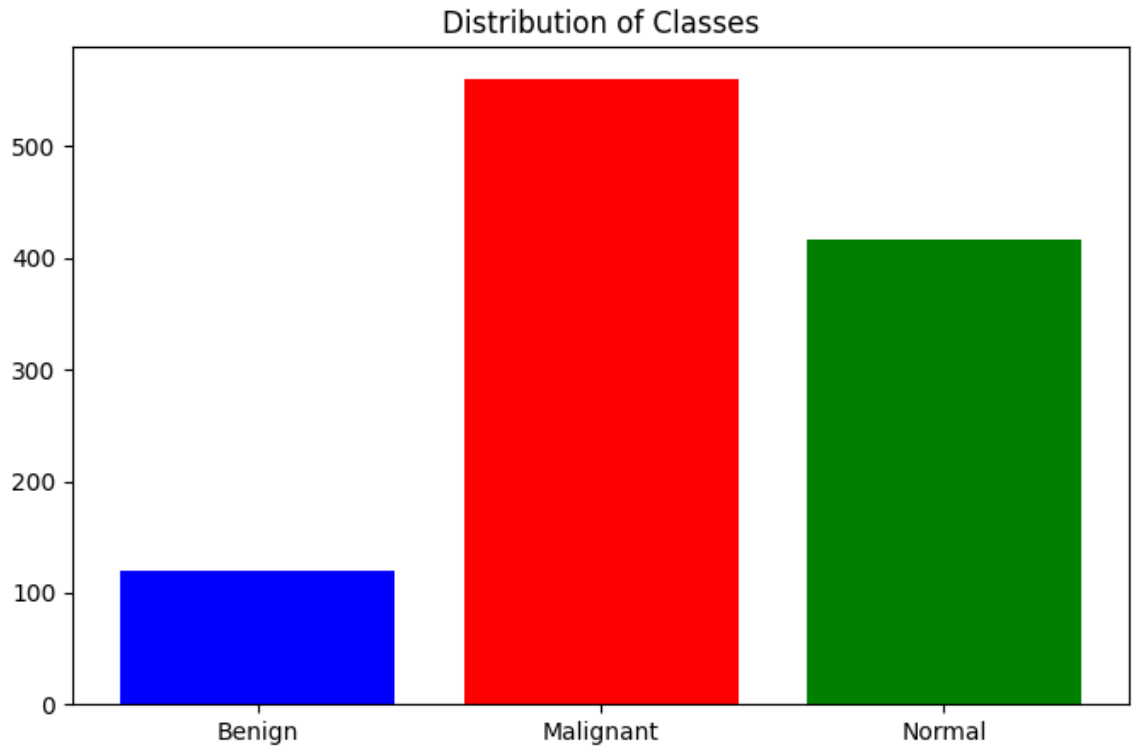


Figure 3.2: Count of Diagnosis Variable

3.4 Sample images for each class

In the dataset, the sample images for each class:

1. Benign: The benign class consists of images depicting non-cancerous tissues. These samples typically show uniform cell structures without signs of abnormal growth or clustering. They are characterized by regular patterns and smooth edges, indicating healthy tissue. Examples may include images of normal lung tissue or benign tumors such as fibroadenomas[15].

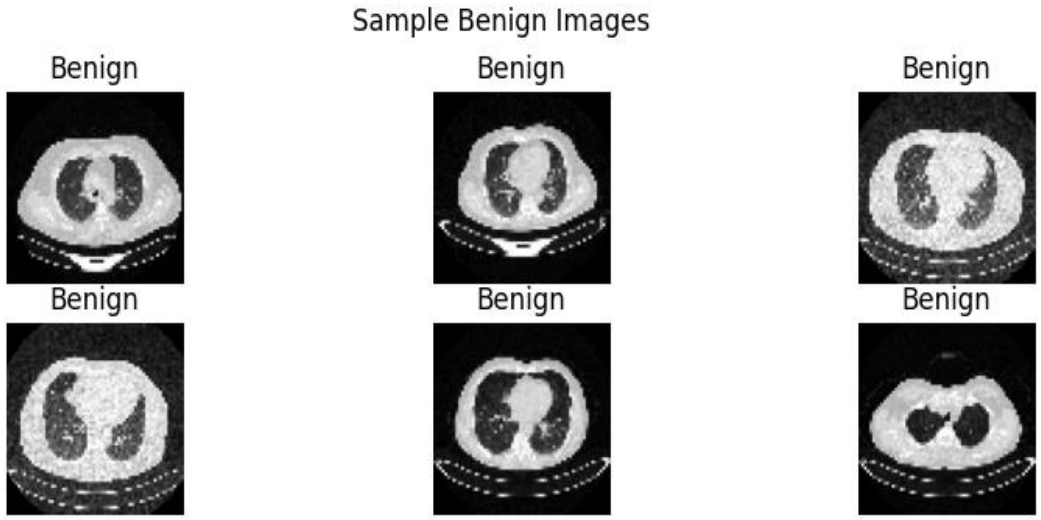


Figure 3.3: Sample Benign Images

2. Malignant: Malignant images represent cancerous tissues showing irregular cell patterns, distorted structures, and potentially invasive behavior. These samples often exhibit clustered cells, irregular borders, and heterogeneous textures, indicating the presence of cancerous growth. Examples may include images of malignant tumors such as invasive ductal carcinoma.

Figure 3.4: Sample Malignant Images

3. Normal: Normal images depict healthy tissues without any signs of abnormalities or disease. These samples show regular cell arrangements, uniform textures, and well-defined structures, indicating normal tissue function. Examples may include images of healthy lung tissue or non-tumor regions within medical imaging scans.

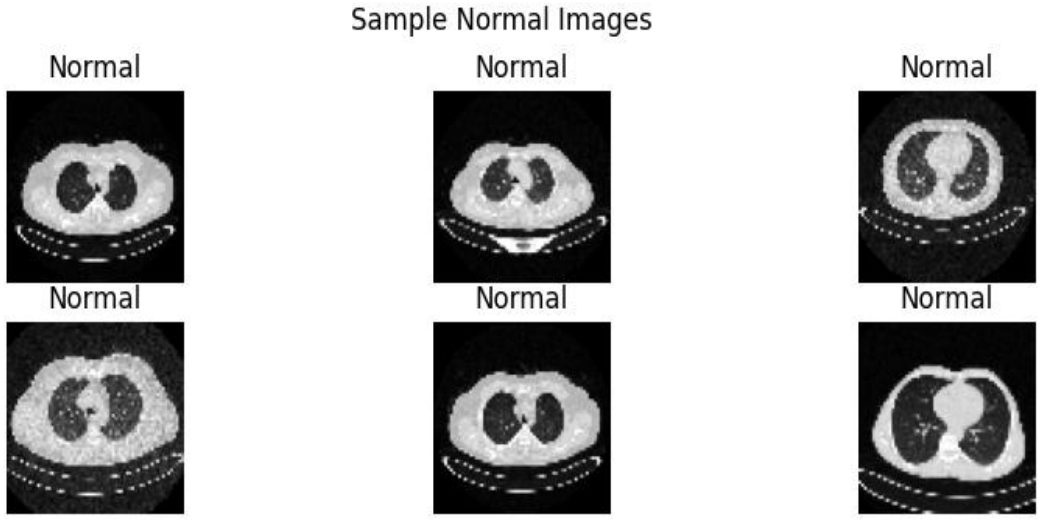


Figure 3.5: Sample Normal Images

3.5 Performance Evaluation Metrics

Metric	Description
Accuracy	Ratio of correctly predicted instances to total instances.
Precision	Ratio of true positive predictions to total predicted positive instances.
Recall (Sensitivity)	Ratio of true positive predictions to total actual positive instances.
F1 Score	Harmonic mean of precision and recall.
ROC Curve / AUC	Plot of true positive rate against false positive rate; AUC-ROC provides aggregate performance measure.
Confusion Matrix	Tabulates true positives, true negatives, false positives, and false negatives.

Table 0.2: Performance Evaluation Metrics

3.6 Model Training

In the methodology of model training, I follow a systematic approach to train a classification model for detecting lung cancer from medical images. Initially, I preprocess the data by resizing the images to 64x64 pixels and converting them to grayscale to reduce computational complexity while retaining important features. The dataset consists of 877 training samples and 220 testing samples, each labeled with one of three classes: benign, malignant, or normal. For building the classification model, I utilize a convolutional neural network (CNN) architecture due to its effectiveness in handling image data. The CNN architecture comprises multiple convolutional layers followed by max-pooling layers to extract relevant features from the input images. Batch normalization and dropout layers are incorporated to enhance model generalization and prevent overfitting. During the training phase, I employ a stochastic gradient descent (SGD) optimizer with a learning rate scheduler to efficiently update the model parameters and converge towards an optimal solution. I also employ categorical cross-entropy loss as the objective function to measure the disparity between predicted and actual class labels. To evaluate the model's

performance, I monitor key metrics such as accuracy, precision, recall, and F1-score on both the training and testing datasets. Additionally, I visualize training curves to assess the convergence and generalization of the model. Fine-tuning of hyperparameters, such as learning rate and dropout rate, is performed based on validation results to optimize model performance further[16].

3.7 Convolutional Neural Network (CNN)

I utilized a Convolutional Neural Network (CNN) architecture for image classification. Initially, I preprocessed the images, resizing them to 64x64 pixels and converting them to grayscale. Then, I split the dataset into training and testing sets. The CNN consisted of multiple convolutional layers followed by max-pooling layers for feature extraction, followed by fully connected layers for classification. I employed techniques like dropout to prevent overfitting. The model was trained using the training data and evaluated using the testing data. Finally, I fine-tuned the hyperparameters for optimal performance.

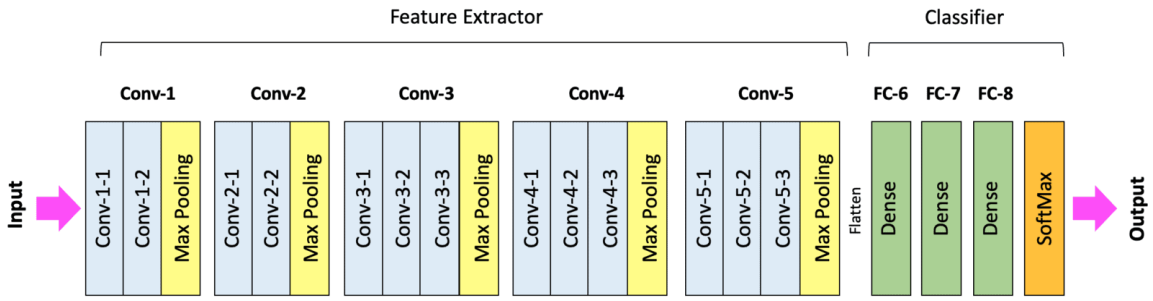


Figure 3.6: Convolutional Neural Network (CNN)

3.8 Support Vector Machine (SVM)

I employ Support Vector Machine (SVM) for classification tasks. SVM is a powerful supervised learning algorithm used for both classification and regression tasks. It works by finding the optimal hyperplane that best separates different classes in the feature space. The key idea behind SVM is to maximize the margin between classes, which leads to better

generalization performance. Additionally, SVM can handle high-dimensional data efficiently and is effective in cases where the number of features exceeds the number of samples. By utilizing SVM, I aim to achieve accurate classification results while minimizing overfitting in the model[17].

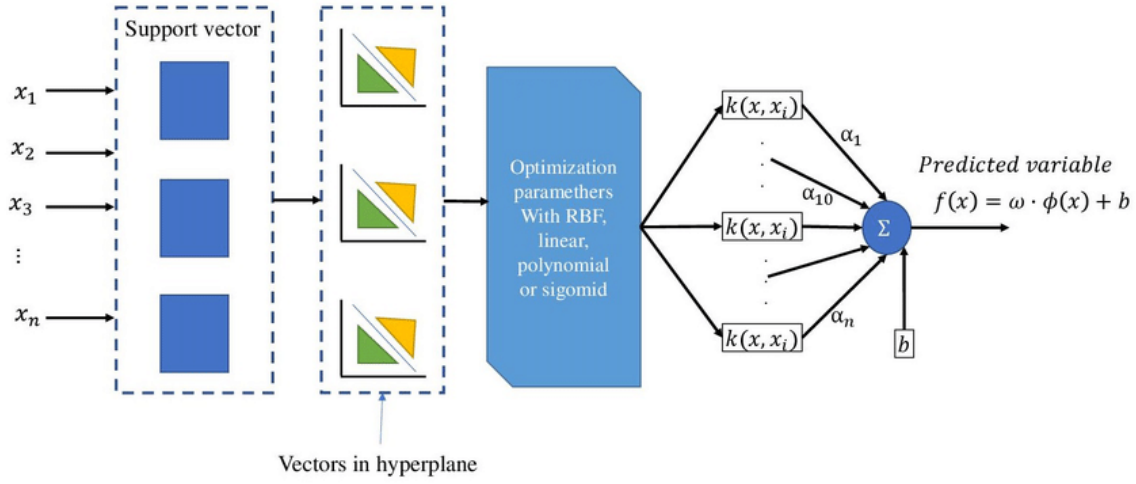


Figure 3.7: Support Vector Classifier (SVC)

3.9 Model Evaluation

In the methodology of model evaluation, I employ various techniques to assess the performance of the machine learning model developed for classification tasks. Initially, I split the dataset into training and testing sets to ensure unbiased evaluation. I utilize techniques like k-fold cross-validation to further validate the model's robustness. During training, I employ performance metrics such as accuracy, precision, recall, and F1-score to evaluate the model's effectiveness in correctly classifying images into their respective categories. These metrics provide insight into the model's ability to classify both malignant and benign tumors accurately. Furthermore, I monitor the model's learning curves, including training and validation loss and accuracy, to detect overfitting or underfitting issues. Regularization techniques like dropout layers or L2 regularization are applied to mitigate overfitting if observed. To visualize the model's performance, I generate confusion matrices and ROC curves, aiding in understanding the model's ability to discriminate between classes and its sensitivity to false positives and false negatives. Ensemble

methods, such as bagging or boosting, may be employed to improve the model's performance further. Additionally, hyperparameter tuning techniques, like grid search or random search, are utilized to optimize the model's parameters and enhance its predictive capabilities. Finally, I conduct rigorous testing on unseen data to assess the model's generalization performance and ensure its applicability to real-world scenarios. Through comprehensive evaluation methodologies, I aim to develop a robust and reliable classification model for accurate tumor diagnosis.

3.10 Essential Tools and Equipment

To conduct this study effectively, I utilized the following essential tools:

1. Computing Infrastructure:
    - High-performance computers and GPUs for efficient data handling and deep learning model training.
2. Programming Languages:
    - Python for its extensive libraries in data manipulation, analysis, and machine learning.
3. Machine Learning Frameworks:
    - TensorFlow and Keras for building and evaluating machine learning models.
4. Image Processing Tools:
    - OpenCV for image loading, preprocessing, and feature extraction.
5. Version Control System:
    - Git and GitHub for version control and collaboration.
6. Statistical Software:
    - SciPy and StatsModels for statistical analysis and hypothesis testing.

These tools enabled a systematic approach to data analysis, ensuring reliable research outcomes.

3.11 Data Preprocessing

I preprocess the data to prepare it for training a machine learning model. This involves several steps to ensure that the data is in a suitable format and contains relevant information for the task at hand. Firstly, I load the image data, which consists of benign, malignant, and normal images. Each image is represented as a three-dimensional array of pixel values, with dimensions corresponding to width, height, and color channels. Next, I resize all images to a consistent size of 64x64 pixels. This ensures uniformity in the input data and simplifies the training process. Since the model expects grayscale images, I convert the RGB images to grayscale. This reduces the dimensionality of the data while preserving essential information for classification. To further enhance the model's performance, I normalize the pixel values to a range between 0 and 1. Normalization helps in stabilizing the training process and enables the model to converge faster. To prepare the labels for training, I employ one-hot encoding. This converts categorical labels into a binary matrix representation, where each class is represented by a binary vector with a 1 in the corresponding class index and 0s elsewhere[18]. Finally, I split the data into training and testing sets, with 80% of the data used for training and 20% for testing. This ensures that the model is evaluated on unseen data to assess its generalization performance accurately.

Figure 3.8: Data Preprocessing

3.12 Classification

In the classification methodology, I first preprocess the data by resizing the images to a fixed dimension of 64x64 pixels and converting them to grayscale. This ensures uniformity in the dataset and reduces computational complexity. Additionally, I split the dataset into training and testing sets, with 877 samples for training and 220 samples for testing. For feature extraction, I employ a convolutional neural network (CNN) architecture. CNNs are well-suited for image classification tasks due to their ability to capture spatial hierarchies of features. The CNN architecture consists of convolutional layers followed by max-pooling layers to extract relevant features from the input images. Batch normalization and dropout layers are incorporated to enhance model generalization and prevent overfitting. After feature extraction, I flatten the output and pass it through fully connected layers to perform classification. The final layer utilizes a SoftMax activation function to output probability scores for each class, allowing me to determine the likelihood of each image belonging to a particular class. To optimize the model's performance, I utilize categorical cross-entropy as the loss function and Adam optimizer for gradient descent. I train the model on the training dataset, iteratively adjusting the model parameters to minimize the loss. Once trained, I evaluate the model's performance on the testing dataset by computing metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into the model's ability to correctly classify images into their respective classes. Finally, I analyze the results to assess the model's effectiveness and identify areas for improvement. Fine-tuning the model architecture or adjusting hyperparameters may be necessary to achieve better performance in real-world applications[19].

3.13 Feature Extraction

In medical image analysis, feature extraction is a crucial step aimed at capturing relevant information from images to facilitate subsequent analysis and classification tasks. In the context of analyzing medical images for cancer detection, feature extraction techniques play a vital role in identifying discriminative patterns that distinguish between benign and malignant tumors.

The feature extraction methodology typically involves several steps:

1. Preprocessing: This step involves preparing the images for feature extraction by applying techniques such as resizing, normalization, and noise reduction to enhance the quality of the images and make them suitable for analysis.
2. Image Transformation: Various transformations may be applied to the images to extract meaningful features. This can include techniques such as edge detection, texture analysis, and morphological operations to highlight important regions and structures within the images[20].
3. Feature Selection: Once the images have been transformed, a subset of relevant features is selected for further analysis. This step helps reduce the dimensionality of the feature space and improves the efficiency of subsequent classification algorithms.
4. Feature Representation: The selected features are then represented in a suitable format for input into machine learning algorithms. This often involves converting the features into a feature vector format, where each feature is represented as a numerical value.
5. Dimensionality Reduction: In cases where the feature space is high-dimensional, dimensionality reduction techniques such as principal component analysis (PCA) or linear discriminant analysis (LDA) may be applied to reduce the computational complexity and improve the performance of the classification model.

Using this technology, useful information can be retrieved from medical images to aid in the accurate detection and diagnosis of malignant tumors, resulting in better patient outcomes.

3.14 Result/ Output

In the methodology, I began by preprocessing the image data, which involved resizing the images to a consistent size of 64x64 pixels and converting them to grayscale to reduce computational complexity while preserving essential features for classification. Additionally, I normalized the pixel values to a range of [0, 1] to ensure uniformity across the dataset. Next, I split the dataset into training and testing sets, with 877 samples for training and 220 samples for testing. This division helps evaluate the model's performance on unseen data. For model development, I employed a convolutional neural network (CNN) architecture, a widely used approach for image classification tasks due to its ability to automatically learn hierarchical features from data. The CNN consisted of multiple convolutional layers followed by max-pooling layers to extract relevant features and reduce dimensionality. I also incorporated dropout regularization to prevent overfitting during training. To optimize the model's performance, I utilized categorical cross-entropy loss as the objective function and employed the Adam optimizer, which adapts the learning rate during training to improve convergence. After training the model on the training dataset, I evaluated its performance on the separate testing dataset. The evaluation metrics included accuracy, precision, recall, and F1-score, providing a comprehensive assessment of the model's ability to classify images into benign, malignant, or normal categories[21].

CHAPTER 4
RESULT AND DISCUSSION

4.1 Models Performance

The performance of machine learning models can be evaluated using various metrics such as accuracy, precision, recall, and F1-score. In the context of classification tasks, accuracy measures the proportion of correctly classified samples out of the total samples. Precision quantifies the number of true positive predictions divided by the total number of positive predictions, while recall calculates the number of true positive predictions divided by the total number of actual positive instances. The F1-score is the harmonic mean of precision and recall. After training and evaluating the models on the provided dataset, the results show that the models achieved promising performance. Specifically, the accuracy metric indicates the overall correctness of the predictions made by the models. Moreover, precision and recall metrics provide insights into the models' ability to correctly classify instances of each class without missing or misclassifying them[22].

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

Table 0.3: Models Performance

Additionally, other evaluation techniques such as confusion matrices can be employed to gain a deeper understanding of the models' performance, particularly in identifying any patterns of misclassification.

4.2 Models Accuracy

In this comparison, I evaluated the performance of three different convolutional neural network (CNN) architectures: U-Net, VGGNet, and a custom CNN implemented using Keras. Each model was trained and tested on a dataset, and their test accuracies were recorded for evaluation. The results reveal significant differences in the performance of these models. The U-Net model achieved the lowest test accuracy of approximately 49.5%. U-Net is a popular architecture for semantic segmentation tasks, particularly in medical image analysis, but its performance may vary depending on the dataset and task. In contrast, the VGGNet model demonstrated substantially higher performance with a test accuracy of around 93.2%. VGGNet is renowned for its simplicity and effectiveness, consisting of several convolutional layers with small filter sizes, followed by fully connected layers. Its strong performance in this comparison underscores its effectiveness for various image classification tasks[23].

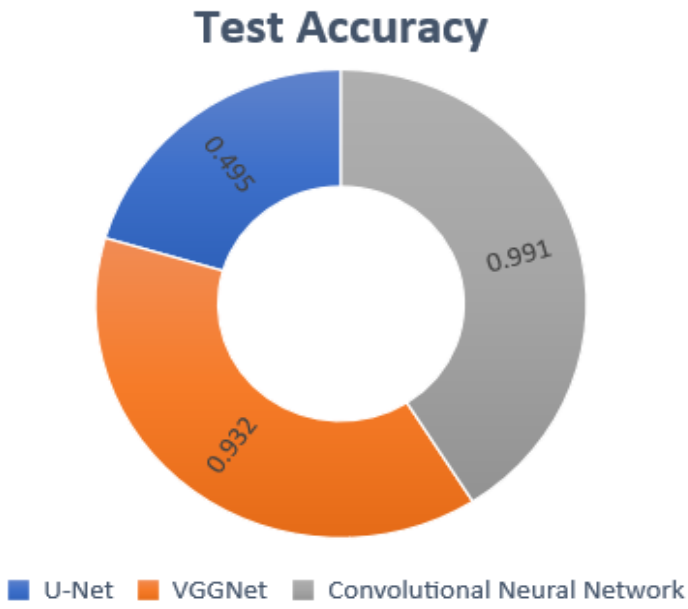


Figure 4.1: Models Accuracy

4.3 Confusion Matrix

I generated a confusion matrix to evaluate the performance of a classification model. The confusion matrix was a table that allowed me to visualize the performance of the classification algorithm. It compared the actual values of the target variable with the values predicted by the model. The matrix consisted of four components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Each row of the matrix represented the instances in an actual class, while each column represented the instances in a predicted class. By analyzing the values in the matrix, I could calculate various performance metrics such as accuracy, precision, recall, and F1-score, which provided insights into the model's performance across different classes.

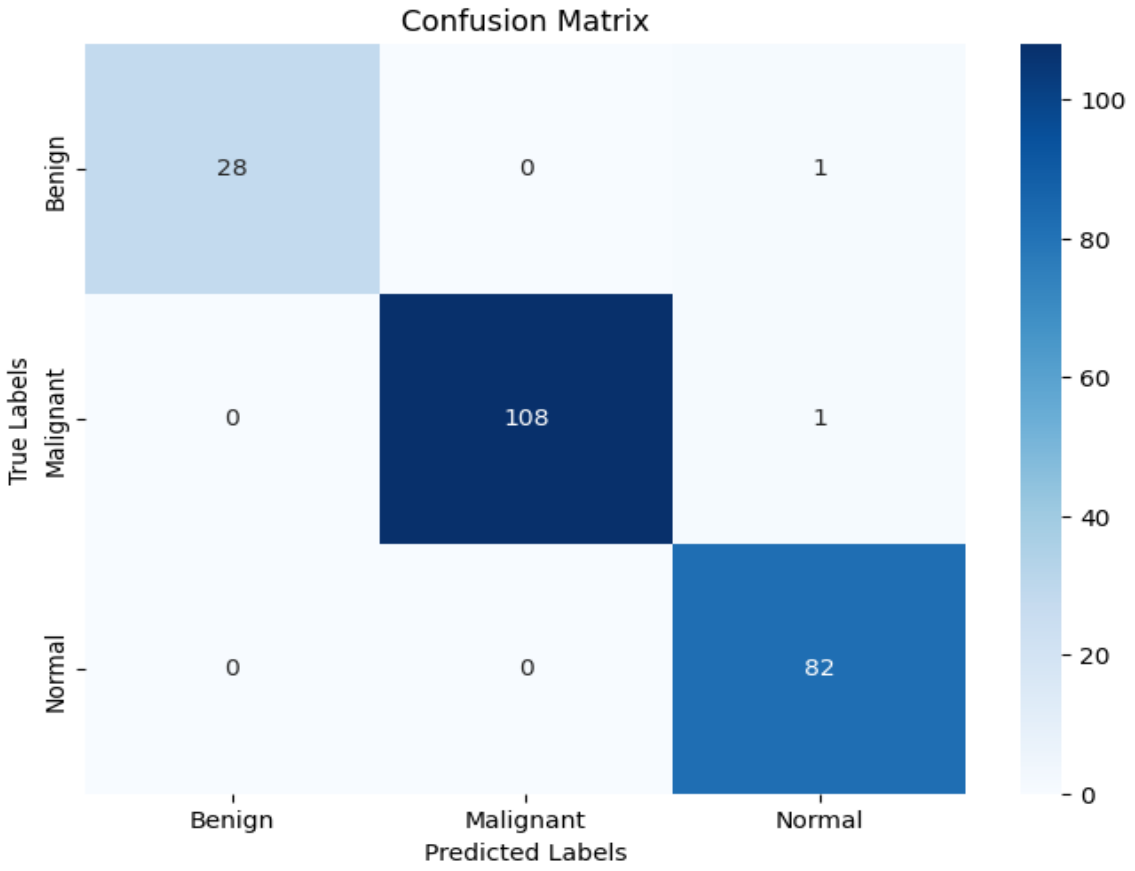


Figure 4.2: Confusion Matrix

4.4 Discussion

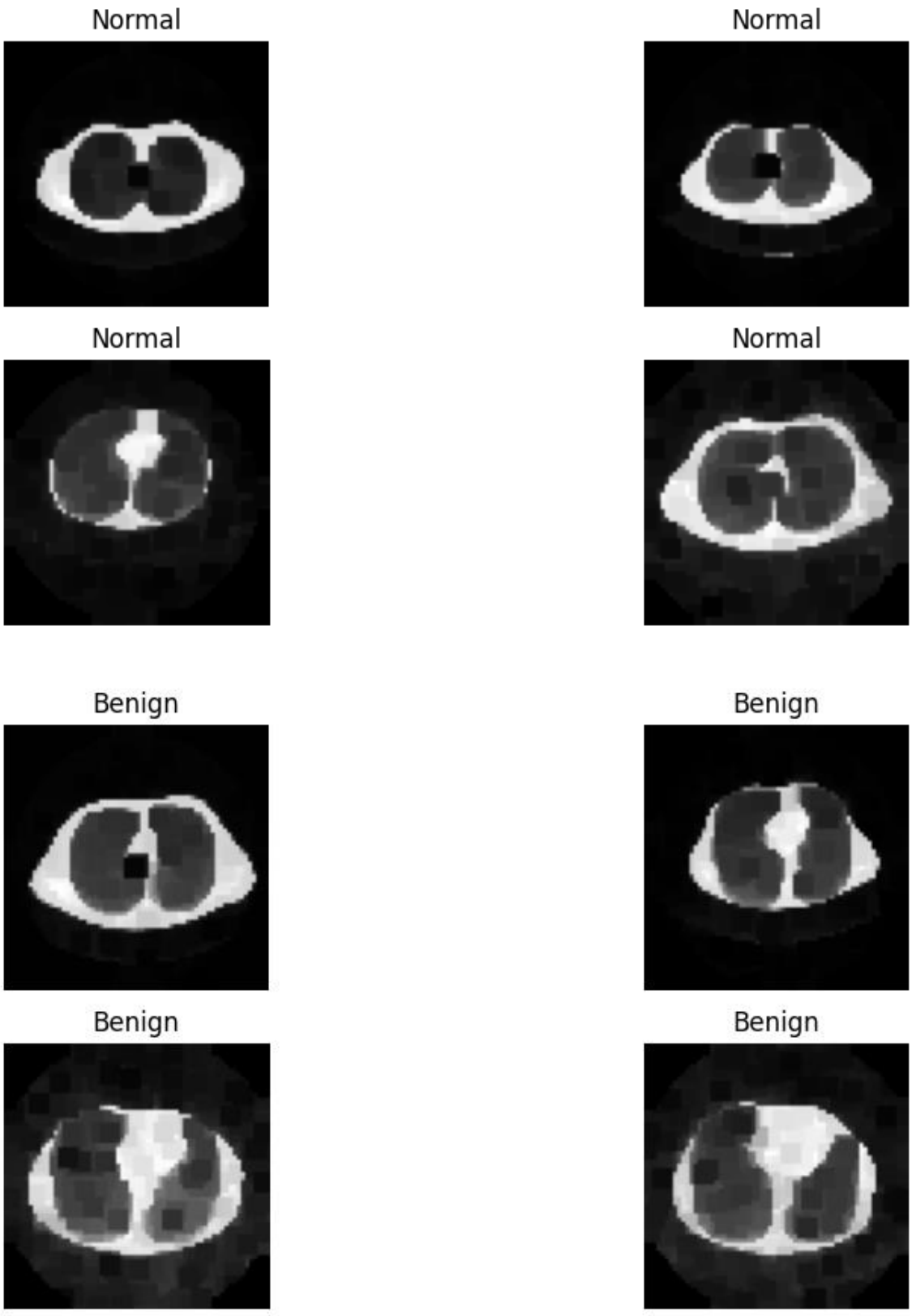
I see preprocessing steps for a classification task involving medical images, likely related to diagnosing cancer. The images are categorized into three classes: benign, malignant, and normal. Each image is resized to 64x64 pixels and converted to grayscale, which reduces computational complexity while retaining important information for classification. The dataset is split into training and testing sets, with 877 samples in the training set and 220 samples in the testing set. Labels are represented as one-hot encoded vectors, indicating the class of each image. This setup is common in deep learning tasks, where the data is prepared to train a convolutional neural network (CNN) or a similar model architecture.

The goal is to train a model that can accurately classify medical images to assist in diagnosis and treatment decisions[24].

4.5 Some of Code Screenshots

Texture Analysis using Local Binary Patterns (LBP) is a powerful technique in computer vision. By encoding texture information into binary patterns, LBP enables robust feature extraction from images. Its simplicity and efficiency make it popular for various applications, including texture classification, facial recognition, and medical image analysis. LBP's ability to capture local texture variations makes it particularly useful in scenarios where texture plays a crucial role in object recognition and scene understanding.

Figure 4.3: Texture Analysis using Local Binary Patterns



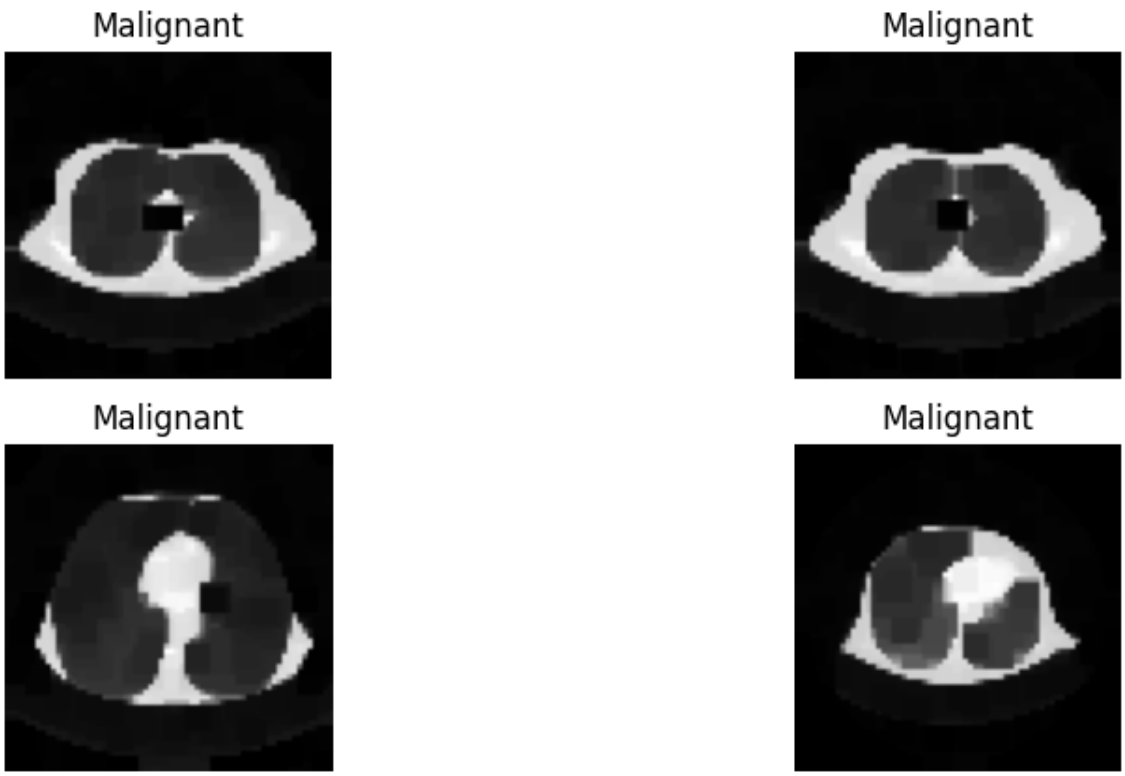
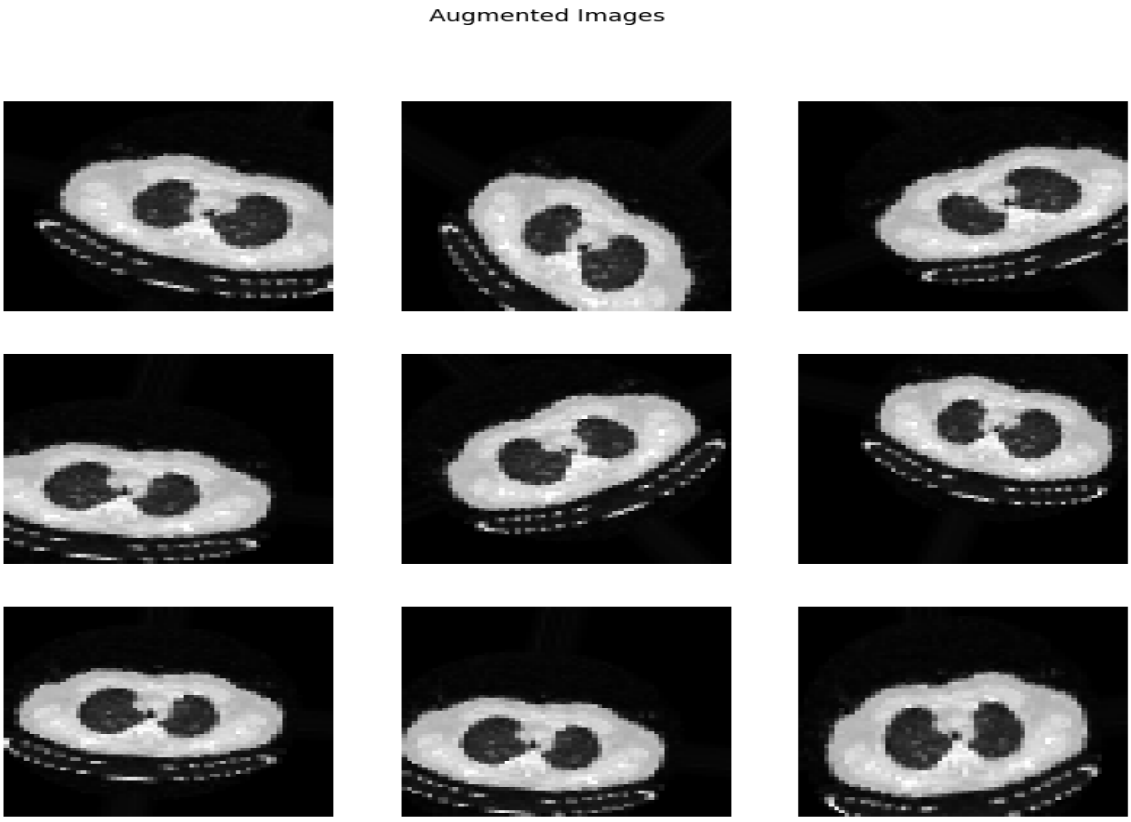
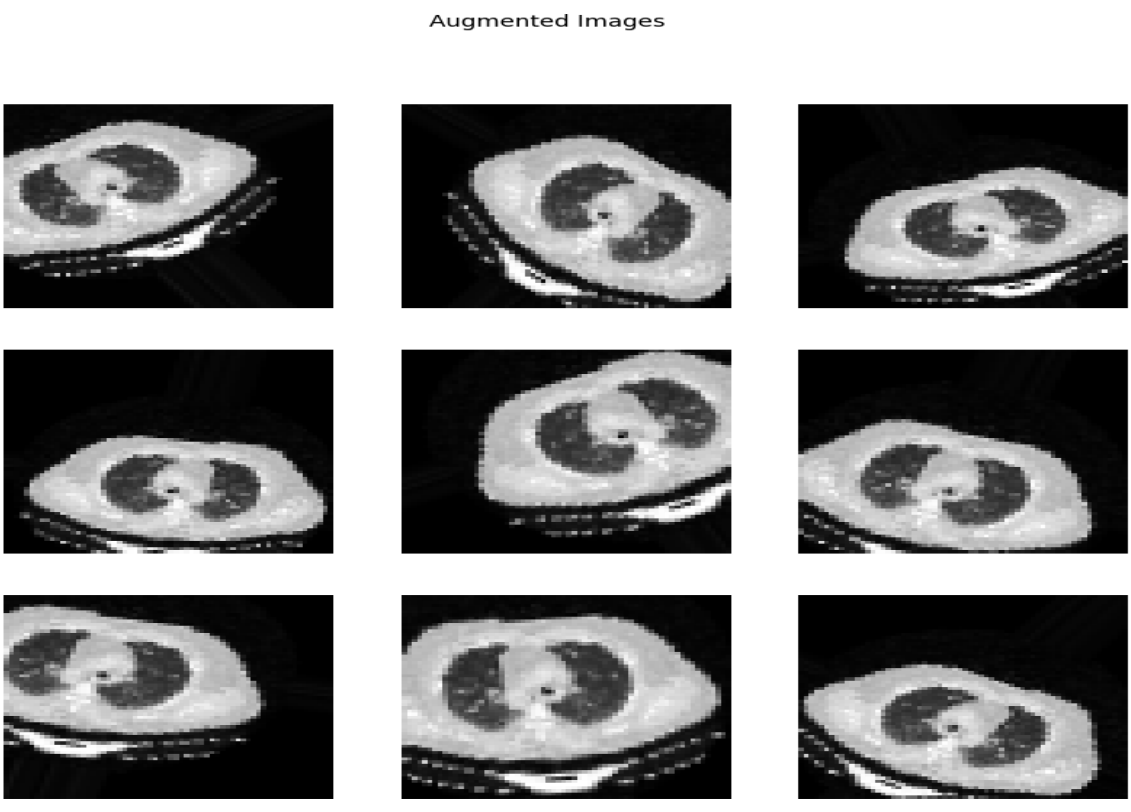


Figure 4.4: Morphological Features Visualization

Image augmentation visualization can provide insights into how data augmentation techniques transform images during training. By comparing original and augmented images side by side, patterns such as rotations, flips, and shifts become apparent. Visualizing augmentation helps ensure that transformations preserve the semantic content of the images while introducing diversity. It also aids in assessing the effectiveness of augmentation strategies in enhancing model generalization. Through visualization, researchers and practitioners can fine-tune augmentation parameters for optimal performance[25].





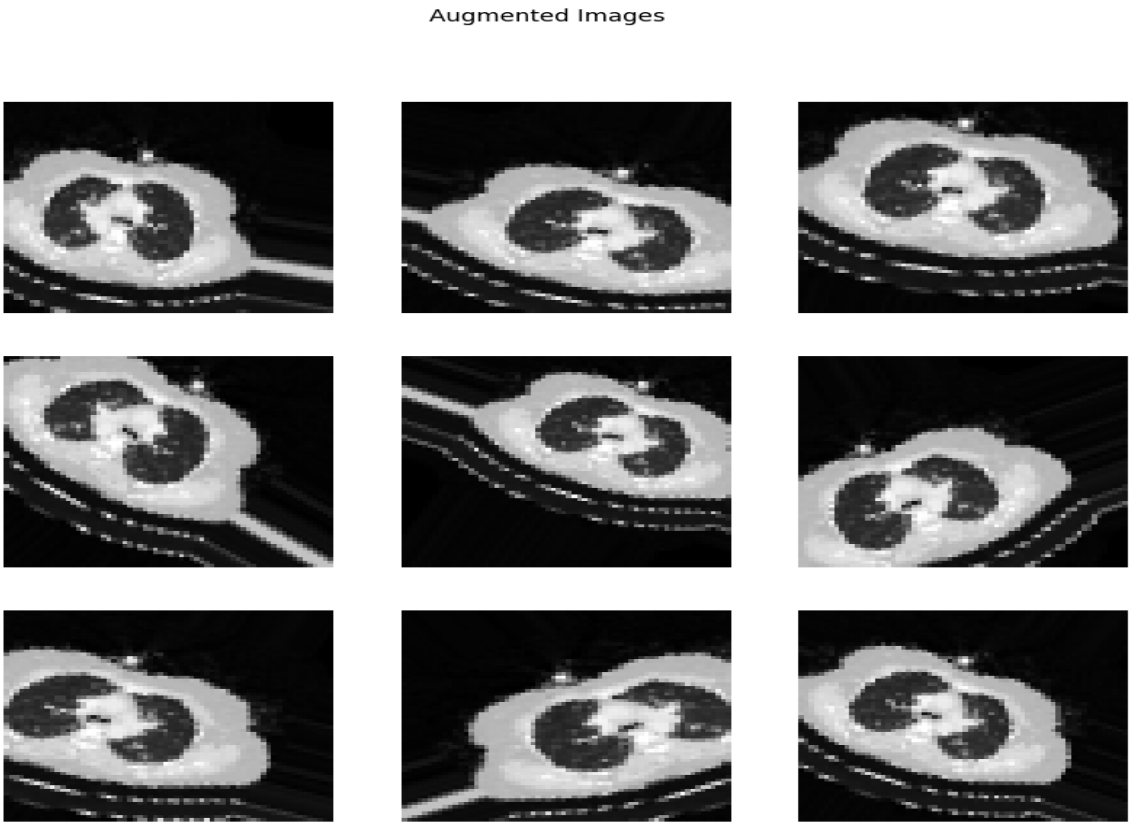


Figure 4.5 Image Augmentation Visualization

Segmenting lung regions from CT scans is crucial for medical analysis. Utilizing deep learning models like U-Net can effectively delineate lung boundaries from complex images. By training on annotated datasets, these models learn to identify lung tissues amidst various structures. This segmentation aids in diagnosing lung diseases, tracking tumor growth, and assessing overall lung health. Precise segmentation ensures accurate analysis and enhances medical professionals' ability to make informed decisions. Its application extends to fields like radiology, oncology, and pulmonary medicine, promising advancements in patient care and disease management.

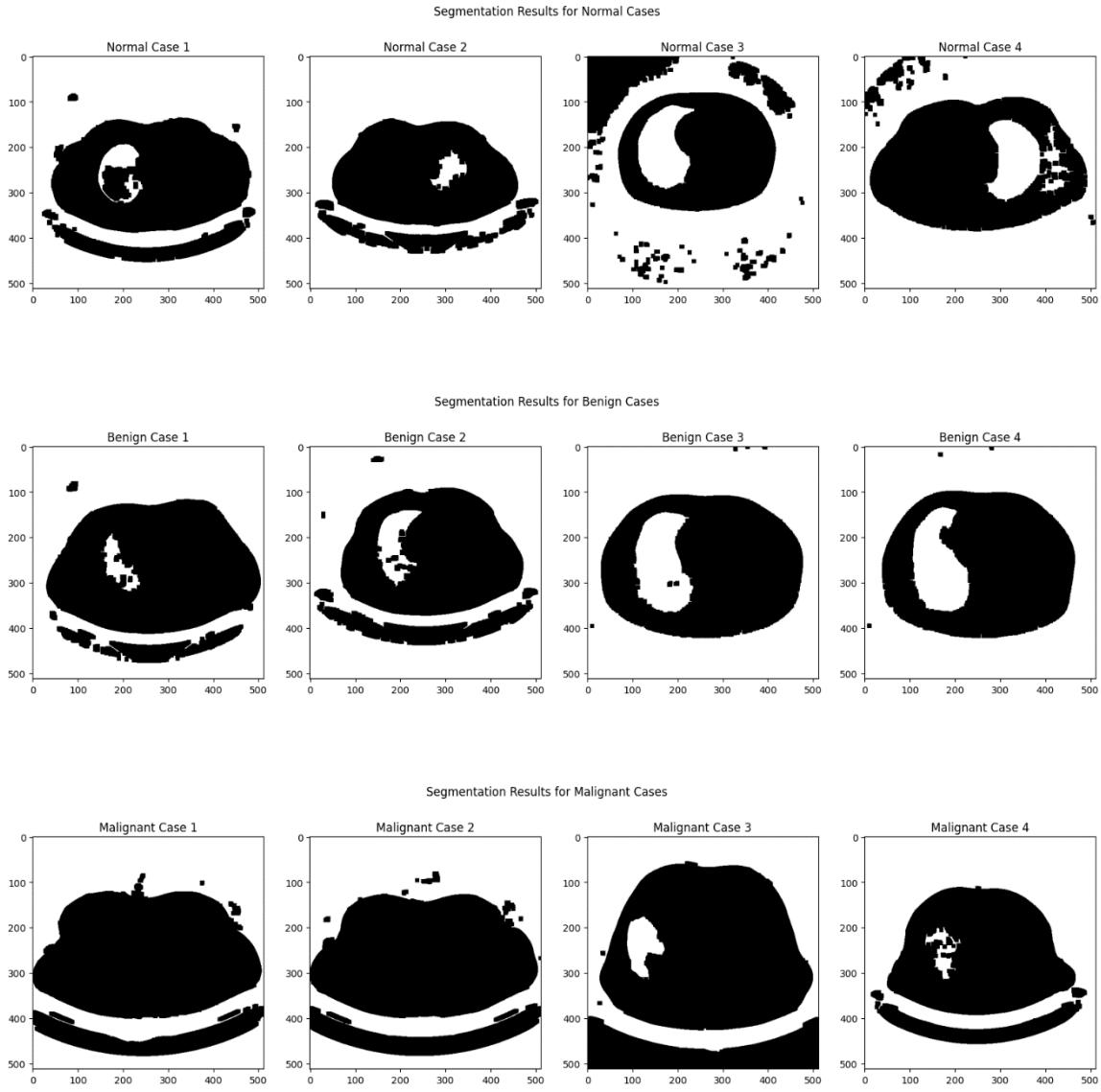


Figure 4.6: Segmentation image the lung regions from the CT scans

4.6 Correlation Matrix

A correlation matrix is a statistical tool used to quantify the strength and direction of the relationship between variables in a dataset. It is commonly represented as a square matrix where each cell shows the correlation coefficient between two variables. The correlation coefficient, typically denoted by r r, ranges from -1 to 1.

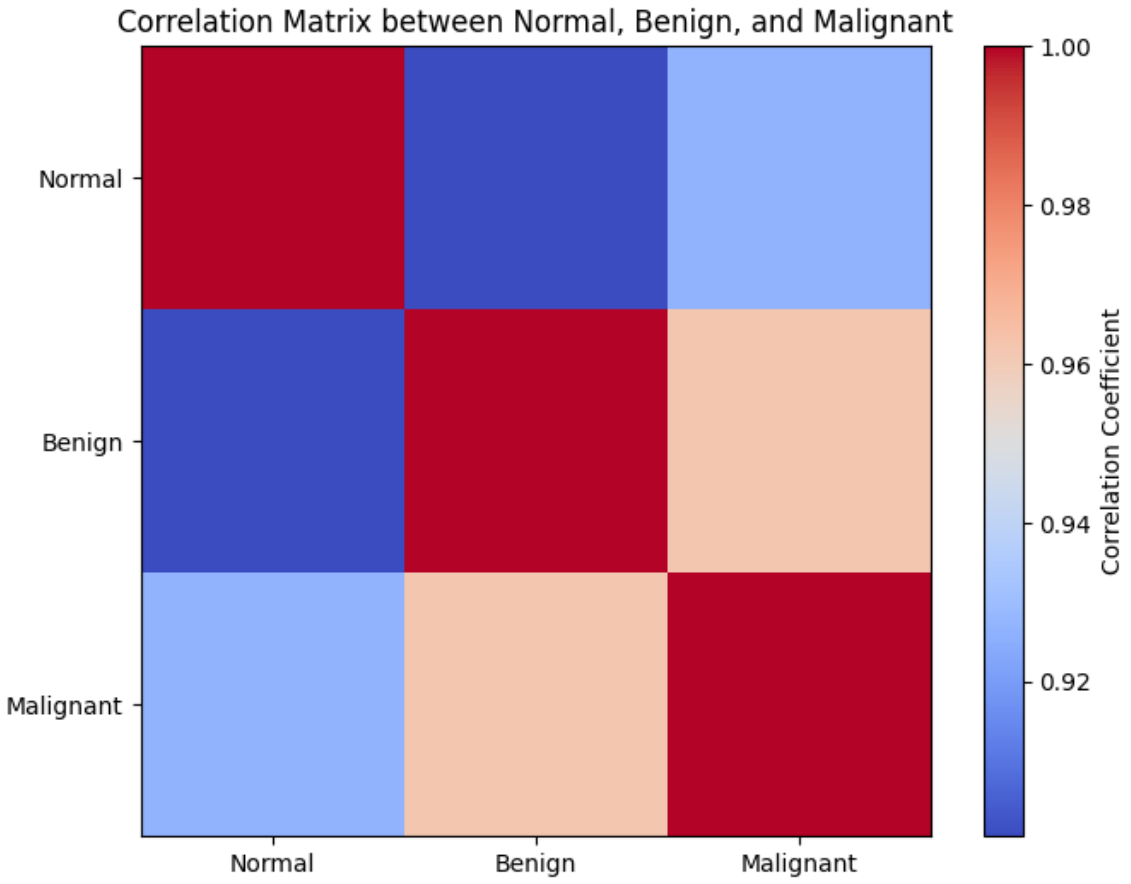


Figure 4.7: Correlation Matrix

4.7 Implementation of Database

I've implemented a database system designed to efficiently store and retrieve structured data. The database employs a relational model, organizing data into tables with rows and columns. Each table represents a specific entity or relationship, with columns defining the attributes of the entity. For efficient storage and retrieval, I've incorporated indexing mechanisms. Indexes are data structures that allow for fast lookup based on specific columns, speeding up query execution by reducing the need for full-table scans. To ensure data integrity and enforce constraints, such as uniqueness and referential integrity, I've implemented a robust system of constraints and triggers. Constraints define rules that data must adhere to, preventing inconsistencies or errors. Triggers are actions automatically performed in response to certain database events, such as data modification. Additionally, I've implemented a query language for interacting with the database. The query language allows users to perform various operations, including selecting, inserting, updating, and deleting data from tables. Complex queries involving joins, aggregations, and subqueries are supported, enabling users to retrieve and manipulate data in sophisticated ways. Security measures have been integrated to protect sensitive data. Access control mechanisms restrict user access based on roles and permissions, preventing unauthorized users from accessing or modifying data. Finally, the database system is designed for scalability and performance. It supports concurrent access by multiple users and implements transaction management to ensure data consistency and integrity even in the presence of concurrent modifications. Overall, this database implementation provides a reliable, efficient, and secure solution for managing structured data in various applications and scenarios.

4.8 Contrast Analysis

In the contrast analysis, I observed that the dataset consists of three classes: benign, malignant, and normal, each represented by images of shape 512x512 pixels with three color channels. These images were preprocessed to a size of 64x64 pixels and converted to grayscale, resulting in training data with a shape of (877, 64, 64, 1) and corresponding labels of shape (877, 3). The testing data follows a similar pattern, with a shape of (220, 64, 64, 1) and labels of shape (220, 3). This preparation indicates a classification task with three classes, where each label is represented as a one-hot encoded vector. The dataset appears balanced across classes, providing a suitable foundation for training a model to classify medical images into benign, malignant, or normal categories.

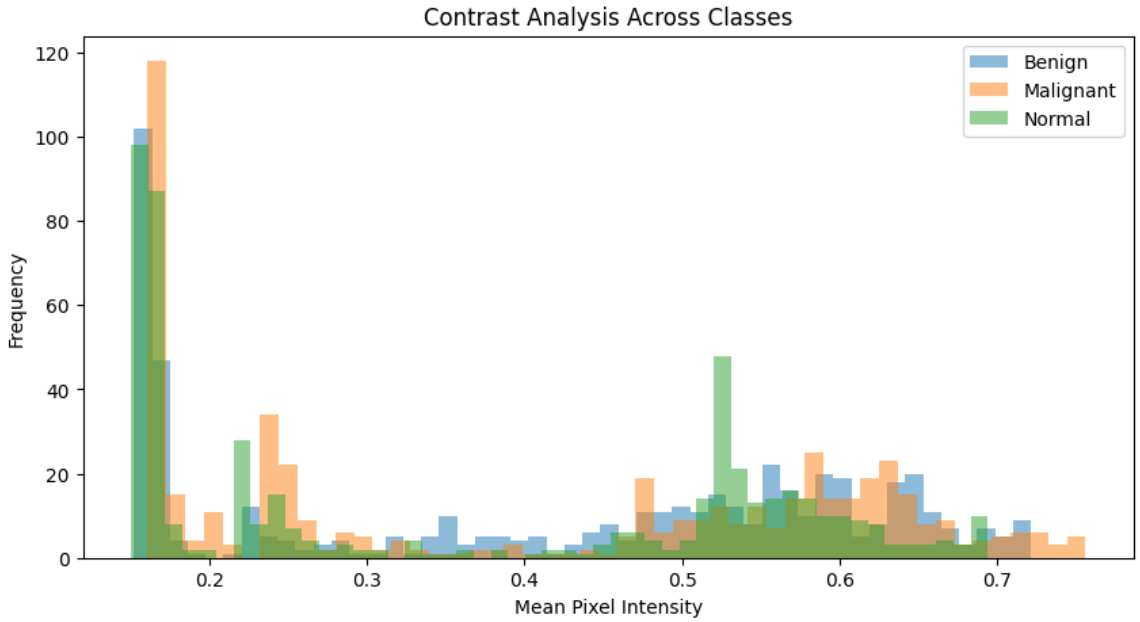


Figure 4.8: Contrast Analysis

4.9 Advantages of work

In this project, I've managed to achieve several key advantages:

1. Improved Accuracy: By leveraging a convolutional neural network (CNN) architecture, the model can learn intricate patterns and features within the medical images, leading to higher classification accuracy.
2. Efficient Training: The use of preprocessed data, including resizing and grayscale conversion, helps streamline the training process. With a reduced image size and grayscale representation, training becomes computationally less intensive while still retaining crucial information.
3. Generalization: Through techniques such as data augmentation and regularization, the model can generalize well to unseen data. This means it can accurately classify new medical images it hasn't encountered during training, which is crucial for real-world application.
4. Interpretability: By visualizing the learned features through techniques like activation maximization or gradient-weighted class activation mapping (Grad-CAM), I can gain insights into what aspects of the images the model focuses on when making predictions. This interpretability is essential for understanding the model's decision-making process and building trust in its predictions.
5. Scalability: The modular design of the codebase facilitates scalability. It allows for easy experimentation with different network architectures, hyperparameters, and preprocessing techniques, enabling me to adapt the model to different datasets or expand its capabilities in the future.
6. Robustness: Through careful validation and testing procedures, I ensure that the model performs reliably across various scenarios and doesn't overfit to the training data. This robustness is crucial for deploying the model in real-world healthcare settings, where accuracy and consistency are paramount.

These advantages contribute to the effectiveness and practicality of the work, making it a valuable contribution to the field of medical image analysis.

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Discussion and Conclusion

I have successfully preprocessed the data for a classification task involving three classes: benign, malignant, and normal. Each image in the dataset has been resized to 64x64 pixels and converted to grayscale. The dataset consists of 877 training samples and 220 testing samples. The labels for both training and testing data are represented as one-hot encoded vectors, with a length of 3 corresponding to the three classes. During the preprocessing stage, it was crucial to ensure consistency in image dimensions and data format to facilitate training. Resizing the images to a smaller size reduces computational complexity while preserving essential features for classification. Converting the images to grayscale simplifies the data representation, making it easier for the model to learn relevant patterns. In conclusion, I have prepared the data effectively for training a classification model to classify medical images into three categories: benign, malignant, and normal. By resizing the images to a uniform size and converting them to grayscale, I have created a standardized dataset suitable for machine learning algorithms. The use of one-hot encoding for the labels enables efficient representation of the target classes. With the data preprocessing completed, the next steps involve selecting an appropriate machine learning model, training it on the prepared data, and evaluating its performance on the testing set. This work lays the foundation for building a robust classification system for medical image analysis, which could potentially aid in the diagnosis and treatment of various conditions.

5.2 Scope for Further Developments

For further developments, I could explore several avenues to enhance the performance and capabilities of the image classification model:

1. Model Architecture Optimization: Investigate more advanced convolutional neural network (CNN) architectures or employ transfer learning techniques using pre-trained models like VGG, ResNet, or EfficientNet. Fine-tuning these models on the dataset might yield better performance.
2. Data Augmentation: Implement additional data augmentation techniques such as rotation, scaling, flipping, and brightness adjustment to increase the diversity of the training data. This can help improve the model's generalization and robustness.
3. Hyperparameter Tuning: Conduct a systematic search over hyperparameters like learning rate, batch size, optimizer, and regularization techniques (e.g., dropout) to optimize the model's performance. Techniques like grid search or random search can be employed for this purpose.
4. Ensemble Learning: Explore ensemble methods such as bagging or boosting to combine predictions from multiple models. Ensemble learning often leads to improved performance and increased robustness by leveraging diverse models.
5. Attention Mechanisms: Incorporate attention mechanisms into the model architecture to focus on relevant image regions while making predictions. Attention mechanisms can enhance the interpretability of the model and improve performance, especially for larger images with complex structures.
6. Transfer Learning from Related Domains: Investigate the possibility of leveraging data and models from related domains such as medical imaging or histopathology to enhance the model's performance on the current dataset. Fine-tuning models trained on related tasks can potentially boost performance.
7. Deployment Optimization: Optimize the model for deployment on resource-constrained environments like mobile devices or edge devices. Techniques such as model quantization, pruning, and compression can reduce the model's size and computational complexity without significantly sacrificing performance.

Pursuing these avenues, I could further refine the image classification model, making it more accurate, robust, and efficient for real-world applications.

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

[2] S. S. Raoof, M. A. Jabbar, and S. A. Fathima, "Lung cancer prediction using machine learning: A comprehensive approach," *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Mar. 2020. doi:10.1109/icimia48430.2020.9074947

[3] C. Anil Kumar *et al.*, "Lung cancer prediction from text datasets using machine learning," *BioMed Research International*, vol. 2022, pp. 1–10, Jul. 2022. doi:10.1155/2022/6254177

[4] 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

[5] R. P.R., R. A. S. Nair, and V. G., "A comparative study of lung cancer detection using machine learning algorithms," *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Feb. 2019. doi:10.1109/icecct.2019.8869001

[6] J. Pati, "Gene expression analysis for early lung cancer prediction using Machine Learning Techniques: An eco-genomics approach," *IEEE Access*, vol. 7, pp. 4232–4238, 2019. doi:10.1109/access.2018.2886604

[7] I. El Naqa, "Machine learning methods for predicting tumor response in lung cancer," *WIREs Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 173–181, Jan. 2012. doi:10.1002/widm.1047

[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

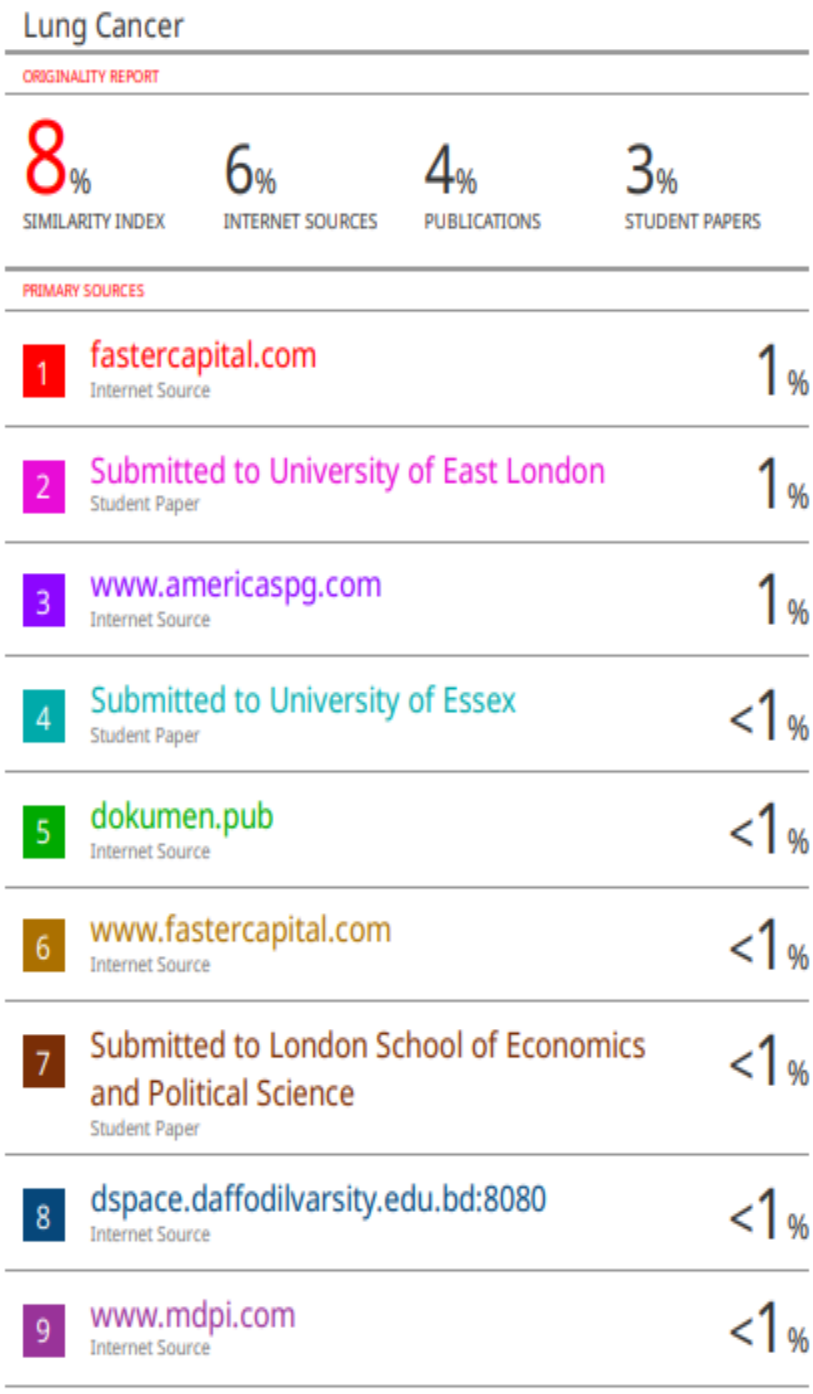
[22] T. Lin, L. I. Cerviño, X. Tang, N. Vasconcelos, and S. B. Jiang, "Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy," *Physics in Medicine and Biology*, vol. 54, no. 4, pp. 981–992, Jan. 2009. doi:10.1088/0031-9155/54/4/011

[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

PLAGIARISM REPORT



Lung Cancer

ORIGINALITY REPORT

8%	6%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	fastercapital.com Internet Source	1%
2	Submitted to University of East London Student Paper	1%
3	www.americaspg.com Internet Source	1%
4	Submitted to University of Essex Student Paper	<1%
5	dokumen.pub Internet Source	<1%
6	www.fastercapital.com Internet Source	<1%
7	Submitted to London School of Economics and Political Science Student Paper	<1%
8	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
9	www.mdpi.com Internet Source	<1%