

---

# **An Explainable Ensemble Convolutional Neural Network for Early Lung Cancer Prediction with Web Application**

---

By

Kazi Rifah Noor

Student ID: 0122320032

Submitted in partial fulfillment of the requirements  
of the degree of Master of Science in Computer Science and Engineering

November 16, 2025



Department of Computer Science and Engineering  
United International University



## Approval Certificate

This thesis titled "**An Explainable Ensemble Convolutional Neural Network for Early Lung Cancer Prediction with Web Application**" submitted by **Kazi Rifah Noor**, Student ID: 0122320032, has been accepted as Satisfactory in fulfillment of the requirement of the degree of Master of Science in Computer Science and Engineering on Date....

### Board of Examiners

1. \_\_\_\_\_ Supervisor

Dr. Mohammad Nurul Huda  
Professor  
Department of Computer Science & Engineering  
United International University (UIU)  
Dhaka, Bangladesh

2. \_\_\_\_\_ Head Examiner

Dr. Dewan Md. Farid  
Professor  
Department of Computer Science & Engineering  
United International University (UIU)  
Dhaka, Bangladesh

3. \_\_\_\_\_ Examiner-I

Dr. Riasat Azim  
Assistant Professor  
Department of Computer Science & Engineering  
United International University (UIU)  
Dhaka, Bangladesh

4. \_\_\_\_\_ Examiner-II

Dr. Muhammad Nomani Kabir  
Professor & UG Program Coordinator  
Department of Computer Science & Engineering  
United International University (UIU)  
Dhaka, Bangladesh

5. \_\_\_\_\_

Dr. Khondaker Abdullah Al Mamun  
Professor & Director – MSCSE  
Department of Computer Science & Engineering  
United International University (UIU)  
Dhaka, Bangladesh

## Declaration

This is to certify that the work entitled " **An Explainable Ensemble Convolutional Neural Network for Early Lung Cancer Prediction with Web Application**" is the outcome of the research carried out by me under the supervision of Dr. Mohammad Nurul Huda, Professor of the Dept. of CSE.

---

Kazi Rifah Noor

ID: 0122320032

MSCSE Program

United International University

Dhaka, Bangladesh

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

---

Dr. Mohammad Nurul Huda

Professor

Department of Computer Science & Engineering

United International University (UIU)

Dhaka, Bangladesh

---

## Abstract

---

Lung Cancer is one of the deadliest forms of cancer significantly contributing to the rising mortality rates globally. The high mortality rate associated with lung cancer can largely be attributed to its late detection, as symptoms often do not appear until the disease has reached advanced stages. Early detection of anomalies in medical imaging, particularly at the initial stages, is crucial for advancing both quantitative image analysis and patient care. In this context, our research introduces a fully automated web application to predict lung cancer early on CT scan images. This application leverages the power of a weighted average ensemble-based deep learning framework that combines multiple neural network architectures to enhance the reliability of automated classification. The work highlights the need for interpretability in clinical decision support systems, going beyond classification. Our proposed approach is organized into three independent phases. First, we performed image augmentation as part of the preprocessing stage analyzing the IQ-OTH lung cancer dataset which implies that the model is trained on diverse and enriched input data. The system incorporates ResNet50, VGG16, and a custom CNN, all of which provide complementary feature-learning capabilities that improve overall predictive dependability. The ensemble model clearly outperformed the individual networks, achieving an accuracy of 92.28% on the IQ-OTH lung cancer dataset. The visual indications and regions that most influence our model's decisions are highlighted using Explainable Artificial Intelligence techniques, particularly LIME and SHAP. By making the model's inner workings more understandable, these explanations hope to boost medical practitioners' confidence.

Finally, the ensemble model was integrated into a user-friendly web application allowing users to upload CT scan images, which are then analyzed to classify the images into normal, benign, or malignant categories. Furthermore, our web application gives confidence levels for each prediction increasing the credibility of its results. Our system provides medical practitioners with a smooth and effective tool by integrating the precision of an ensemble model with the flexibility of a web application.

**Index Terms**—Web application, lung cancer prediction, CT image, healthcare, deep learning, explainable AI

---

## Acknowledgement

---

I want to express my sincere gratitude towards all the teachers of the United International University for giving me the opportunity to pursue my master's degree and conduct this study. My sincere gratitude goes out to everyone who has encouraged and helped me along the entire way.

I intend to express my gratitude to my esteemed academic supervisor, Prof. Dr. Mohammad Nurul Huda, Professor of the Dept. of CSE for his unwavering guidance and mentorship. His unwavering support and perceptive feedback have been crucial in enabling me in accomplishing this research in the thrilling domains of deep learning and artificial intelligence.

Additionally, I would like to thank United International University (UIU) for facilitating an environment that is favorable to research and creativity. Finally, I acknowledge my family for their constant affection and assistance, which have been my biggest source of courage along this journey.

---

## Publication List

---

The main contributions of this research are either published or accepted or in preparation in journals and conferences as mentioned in the following list:

### Conference Papers

- i. K. R. Noor, S. Islam, N. Tasnim, M. N. Huda, (2024). DeepWeb: A Web Application for Advanced Lung Cancer Identification using Novel CNN Model. International Conference on Innovations in Science, Engineering and Technology 2024 (ICISSET 2024). DOI: <https://doi.org/10.1109/ICISSET62123.2024.11003916>
- ii. N. Tasnim, K. R. Noor, M. Islam, M. N. Huda, I. H. Sarker (2024). A Deep Learning based Image Processing Technique for Early Lung Cancer Prediction. 2024 International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS). DOI:10.1109/ICETISIS61505.2024.10459494

# Table of Contents

INTRODUCTION .....	1
1.1 Motivation.....	1
1.2 Cancer Detection Using Imaging Techniques.....	2
1.3 Explainable Artificial Intelligence in Cancer Detection .....	3
1.4 Problem Statement .....	4
1.5 Research Objective .....	5
1.6 Thesis Organization .....	5
BACKGROUND AND LITERATURE REVIEW.....	7
2.1 Background.....	7
2.2 Related Works.....	7
2.2.1 Related Works with the Same Dataset .....	8
2.2.2 Related works with Different Dataset .....	10
2.2.3 Related Works on Explainable Artificial Intelligence .....	15
METHODOLOGY .....	17
3.1 Dataset Description.....	17
3.2 Image Augmentation.....	18
3.3 Dataset Preprocessing .....	21
3.4 Lung Cancer Classification using Deep Learning Methods.....	22
3.5 Lung Cancer Classification using Weighted Average Ensemble Methods.....	30
3.6 Model Performance Evaluation Measures.....	31
3.7 Explainable AI .....	32
3.7.1 LIME Implementation.....	33
3.7.2 SHAP Implementation .....	33
WEB APPLICATION DEVELOPMENT .....	35
4.1 Web Application Development Process .....	35
4.2 Lung Cancer Prediction with the Web Application .....	37
EXPERIMENT & RESULT ANALYSIS .....	41

5.1 Performance Analysis of Different Deep Learning Approaches .....	41
5.1.1 Convolutional Neural Network.....	41
5.1.2 VGG-16.....	43
5.1.3 ResNet50.....	44
5.2 Performance Analysis of Weighted Average Ensemble Model.....	46
5.3 Performance Analysis of our Proposed Model with Existing Works .....	49
5.4 Explainable Artificial Intelligence .....	51
5.4.1 LIME Technique .....	51
5.4.2 SHAP Technique.....	52
5.5 Computational Complexity Analysis of the Proposed System .....	57
5.6 Cost Analysis of the Proposed System .....	58
CONCLUSION & FUTURE WORKS.....	59
6.1 Conclusion .....	59
6.2 Future Scope.....	60
References.....	61

## List of Tables

Table 1. Comparison table of related works on the same dataset .....	9
Table 2. Comparison table of related works on different dataset.....	15
Table 3. Class distribution after augmentation .....	19
Table 4. Performance evaluation of the three deep learning models .....	45
Table 5. Performance Comparison of our proposed method and previous research on same Dataset .....	50

## List of Figures

Figure 1. Visual exploration of imaging techniques for lung cancer detection .....	2
Figure 2: XAI for the interpretability of classification model .....	4
Figure 3. CT scan images of lungs with three cases .....	18
Figure 4. Class distribution of our dataset after augmentation. ....	20
Figure 5. Preprocessed Images.....	20
Figure 6. Workflow diagram for the proposed method of Lung cancer prediction .....	22
Figure 7. Summary of the CNN model.....	24
Figure 8. Max-pooling Layer .....	25
Figure 9. Flattening layer.....	26
Figure 10. Structure of the CNN model.....	27
Figure 11. Residual block or skip connection on ResNet50. ....	28
Figure 12. Architecture of ResNet50. ....	29
Figure 13. VGG-16 architecture diagram .....	29
Figure 14. Weighted average ensemble model for final prediction. ....	31
Figure 15. Workflow of our web application.....	35
Figure 16. UI of our Web application. ....	37
Figure 17. Lung Cancer prediction with confidence score .....	38
Figure 18. ERD diagram for web application .....	40
Figure 19. Training accuracy vs validation accuracy curve of CNN.....	42
Figure 20. Training loss vs validation loss curve of CNN.....	42
Figure 21. Training accuracy vs validation accuracy curve of VGG16.....	43
Figure 22. Training loss vs validation loss curve of VGG16.....	44
Figure 23. Training accuracy vs validation accuracy curve of ResNet50.....	45
Figure 24. Training loss vs validation loss curve of ResNet50.....	45
Figure 25. Performance comparison of deep learning models.....	46
Figure 26. Training accuracy vs validation accuracy curve for ensemble model .....	47
Figure 27. Receiver Operating Characteristic (ROC) curve for ensemble model .....	47

Figure 28. Precision-Recall (PR) curve for ensemble model.....	48
Figure 29. Test accuracy comparison of the individual models and ensemble model.....	49
Figure 30. Images interpreted through LIME technique.....	51
Figure 31. Feature before scaling.....	53
Figure 32. Feature after scaling.....	54
Figure 33. Output with SHAP value .....	55

# CHAPTER 1

## INTRODUCTION

One of the deadliest diseases in the world today is lung cancer, which is defined as tumors resulting from uncontrolled cell development in the human lung causing death over the past decades. Lung tumors can either be cancerous which are called malignant or noncancerous known as benign. Primarily, tobacco usage is the main contributor to lung cancer [1]. While pulmonary nodules are common among smokers, occurring in roughly half of them, and multiple nodules are found in about a quarter, the vast majority of these nodules are benign, with a malignancy rate below 1% [2]. A wide range of factors can also contribute, including demographic characteristics (age, gender, race, socioeconomic status), environmental exposures (occupational, air pollution, secondhand smoke), health conditions (chronic lung disease, obesity), lifestyle (dry cough, alcohol consumption, diet), and genetic predisposition [3]. Even this cancer is spread by habits adopted by individuals [4]. This illness progresses rapidly; among cancer cells, 49% to 53% are found in stage 4, the last stage of the disease, when the survival rate is less than 6% [5]. Those who have lung cancer have an average lifespan after five years of about 14%, but if the tumor is discovered early, the survival rate can increase to 60–70% [6]. Thus, automated early identification and care of lung cancer by effective screening methods are important for improving patient outcomes. The most prevalent assessment method for lung cancer is computed tomography (CT) imaging detection [7].

Many medical practices have been able to grow and thrive as a result of recent advancements that have had a significant impact on the healthcare sector [8]. In modern times, the deep learning model is frequently employed to identify, evaluate as well as categorize vital medical interventions [9]. Smart healthcare makes it possible to spot cancer considerably sooner than with conventional techniques, giving patients the best chance of a full recovery and enabling them to begin treatments as soon as possible. [10].

### 1.1 Motivation

In addition to a number of other aspects that hamper early detection of the lung cancer, such as insufficient information for diagnosis, the primary reason is that doctors cannot identify lung cancer manually quickly from CT images for the poor quality of medical images [11]. In this scenario, systemic challenges, such as physician stress, the complexity of diagnoses, and

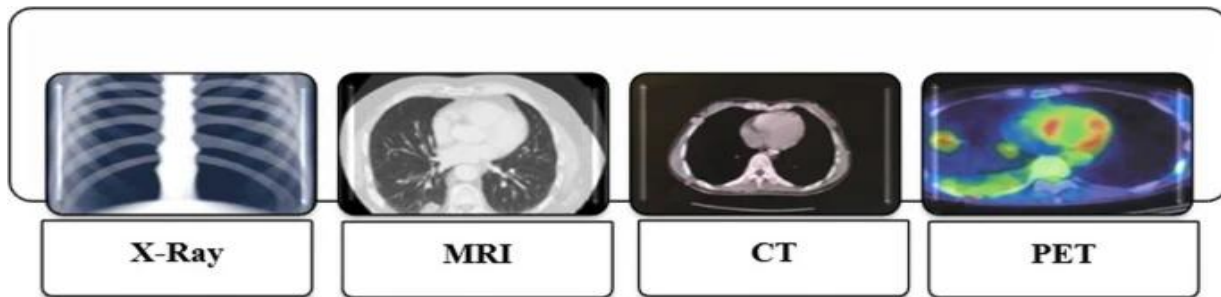
breakdowns in patient-physician communication, contribute to inappropriate care, driving up medical costs and threatening patient health. It is probable that the malignant disease has advanced to severe and challenging-to-treat stages throughout this period.

A fast and efficient cancer diagnosis process can greatly lower a patient's chance of death from lung cancer. Certain decision-making procedures related to cancer detection and diagnostic parameter setting can be automated with the help of image processing. It lessens the possibility of human error, which can be very important when evaluating circumstances that could be fatal.

In this study, deep learning models have been deployed on CT scan images rather than machine learning models to predict lung cancer. Because machine learning algorithms require manual feature extraction before segmentation or classification [12]. Conversely, deep learning-based algorithms automatically extract important features from input images via convolution [13].

## 1.2 Cancer Detection Using Imaging Techniques

Medical professionals can examine the body to determine potential health issues with the aid of imaging technology, which is a powerful diagnostic tool [14]. Figure 1 illustrates some imaging modalities for lung cancer detection. [15].



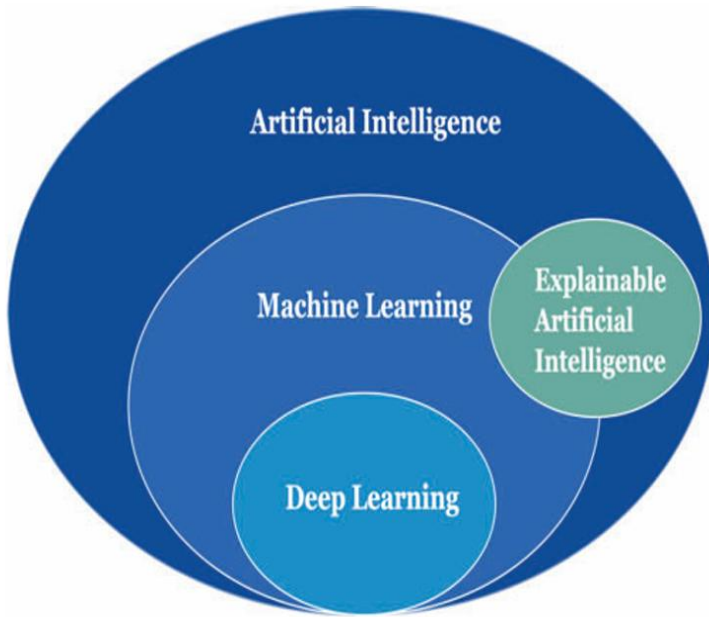
**Figure 1. Visual exploration of imaging techniques for lung cancer detection**

Employing cutting-edge imaging technology, including Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scanners, this technique produces intricate visual scenarios of internal organs and tissues. These images can be used to identify anomalies, including cancers, and assist medical professionals in reaching precise diagnoses. To identify particular types of cancer and identify changes in the body, other imaging methods like X-rays, ultrasounds, along

with Positron Emission Tomography (PET) scans are also frequently performed. Furthermore, there is growing evidence that low-dose CT (LDCT) screening of the chest, in addition to standard chest radiography, can help detect lung cancers early [16]. While chest radiography remains a most widely used tool for chest diseases, studies have not demonstrated it to reduce mortality when used for lung cancer screening. Low-dose computed tomography (LDCT) has shown greater effectiveness in this regard [17]. Thus, LDCT is the main technique for cancer screening, and CT is carried out if abnormalities are seen on chest radiographs. At the prior step of classification process, these images should be preprocessed for more accurate result. Image processing is a means to transform an image into digital form and execute certain operations on it to achieve some results, and based on these findings, a decision- making process is performed in the medical area [18].

### **1.3 Explainable Artificial Intelligence in Cancer Detection**

Explainable AI (XAI) is artificial intelligence designed to gain trust by explaining its objectives, justifications, and decision-making process in a manner that ordinary people recognize easily. It is typically necessary to understand the logic behind an AI model's decision-making. As a result, the demand for eXplainable AI (XAI) strategies to increase confidence in AI models has emerged [19]. This proposed framework enhances the interpretability and explainability of AI models, making their decision-making processes more transparent to clinicians and patients. It is growing more significant as AI algorithms and models get more complicated and difficult to understand. This might promote confidence as well as adoption of AI in the healthcare industry. Figure 2 shows the XAI for the interpretability of classification model [20].



**Figure 2: XAI for the interpretability of classification model**

Explainability techniques can enhance the accessibility and functionality of an ML-based system [21]. When an AI model is continuously giving inaccurate diagnoses, XAI techniques can assist in determining the characteristics the model is concentrating on and provide changes to increase the model's accuracy. Explainable Artificial Intelligence seeks to realize AI's full potential by increasing its interpretability and transparency while making sure that its decisions are consistent with moral standards, legal constraints, and human values.

### **1.4 Problem Statement**

Healthcare professionals require a system that can deliver precise outcomes and diagnoses. Medical imaging methods cannot automatically identify abnormalities and have a significant false positive rate, among other drawbacks. Radiologists generally find it difficult to identify nodules and distinguish between benign and malignant nodules manually in clinical practice. One major constraint in this study is the need for a substantial quantity of high-quality data. Expanding the number of datasets, including medical photographs, is often more difficult than in other sectors. Although automated methods have been developed for these datasets, more accurate results are still required. Therefore, by providing an efficient method for precisely identifying and categorizing lung cancer, we want to contribute to the objective.

## 1.5 Research Objective

The novel aspects and primary contributions of our study are outlined below:

1. **Lung cancer classification:** To classify lung cancer using the weighted average ensemble learning model with certain suitable image preprocessing steps.
2. **Model's interpretability:** To improve the transparency and trustworthiness of the proposed weighted average ensemble model by adopting explainable artificial intelligence (XAI) techniques.
3. **Web application development:** To deliver our suggested model available as a user-friendly and accessible web tool for practical clinical applications and real-time forecasts.

## 1.6 Thesis Organization

This thesis adheres to a meticulously structured framework that ensures a logical and comprehensible flow of ideas. Chapter 2 provides comprehensive background information that is crucial to the study and serves as a foundational chapter. This chapter also includes a comprehensive review of the literature, looking at current studies on early lung cancer detection devices. Additionally, Chapter 2 explores the concept of explainable artificial intelligence (XAI), going over ideas and approaches from past studies that aimed to increase the interpretability and transparency of AI models, particularly in the context of medical diagnosis.

In Chapter 3, the basic research methodology is highlighted with great detail on how the lung imaging datasets are collected and prepared for additional analysis. The chapter then goes into detail on how to apply data augmentation techniques, which are essential for enhancing deep learning models' robustness and generalizability by artificially growing the training dataset. Chapter 3 then explores applying an ensemble model to the preprocessed data using several deep learning algorithms. Additionally, by applying XAI techniques into the study, this chapter addresses the critical issue of model interpretability that enable users to comprehend the model's decision-making process more thoroughly and increases confidence in its predictions.

A user-friendly web application is created in Chapter 4 to demonstrate the way of research findings putting into practice. In order to assist with automated lung cancer screening, this application provides users with real-time projections and associated confidence levels. The development

process is described in detail in this chapter, along with the technology and approaches used to create the web application. This covers things like user interface design, front-end and back-end development, and incorporating the trained ensemble model into the application's workflow.

Chapter 5 presents the scientific evaluation of this research, which carefully examines the experimental findings. In order to provide a thorough assessment of the models' efficacy in identifying lung cancer, this chapter focuses on evaluating their performance and accuracy. Furthermore, the impact of various similarity algorithms on the system's overall performance is evaluated through a comparative analysis in this chapter.

Finally, Chapter 6 concludes the dissertation, offering a concise yet insightful summary of the research findings. It covers the developments in automated lung cancer diagnosis while reviewing the study's contributions. The chapter acknowledges the limitations of the current work suggesting potential avenues for future investigation and investigating new strategies.

## **CHAPTER 2**

### **BACKGROUND AND LITERATURE REVIEW**

This chapter describes the key concepts and prior studies on lung cancer detection and contemporary imaging-based screening techniques. It looks at the developments in existing approaches, the challenges they still face, and the justification for using ensemble and deep learning techniques. By examining these earlier investigations, the chapter defines the background for the current study and clarifies the research gap it aims to fill.

#### **2.1 Background**

The primary reason for the relatively high death rate from lung cancer is the significant delay in diagnosis, even if it is just four weeks, and the subsequent initiation of treatment [22]. This crucial delay emphasizes the urgent necessity for cutting-edge diagnostic procedures. In response, the scientific community is actively investigating the potential of cutting-edge technologies, namely image processing and artificial intelligence (AI), to revolutionize early lung cancer detection. For the evaluation of medical pictures including X-rays, CT scans, and PET scans, image processing techniques offer strong tools.

Furthermore, extensive medical image datasets can be used to train AI systems to identify sophisticated patterns and characteristics similar to both diseased and non-cancerous lung tissue. These technologies have the potential to dramatically change the method in which lung cancer is treated by helping with early detection, delivering more precise treatments, and ultimately saving countless lives.

Most automated models currently in use depend on simple machine learning algorithms that aren't sophisticated enough to adequately examine complex visual patterns. Our study investigates the critical issue of enhancing the precision and dependability of lung cancer diagnosis in medical imaging using state-of-the-art deep learning algorithms and displaying the prediction results.

#### **2.2 Related Works**

The research community has worked hard to find an efficient way to detect lung cancer. In recent years, researchers have worked hard to find a solution to this challenging problem. The remarkable

advancements in machine learning, primarily deep learning, during the last ten years have led to the development of several innovative methods which have shown great promise for improving lung cancer nodule classification accuracy and successfully identifying between benign and malignant cases.

This study includes a comparative analysis of previous research projects to provide a comprehensive understanding of the recent state of the art. It encompasses a diverse range of relevant studies conducted on various datasets, as well as a focused examination of existing research utilizing the specific dataset employed in our own investigation.

### **2.2.1 Related Works with the Same Dataset**

Recently, many approaches have been investigated in the field of medical imaging analysis of lung cancer in order to improve the precision and effectiveness of diagnosis. This section explores a number of well-known machine learning and deep learning methods used in the automated analysis of lung cancer images, outlining each method's benefits and drawbacks.

In 2024, S. Yogesh Kumaran et al. developed an integrated deep learning system to classify lung cancer into three classes i.e. benign, malignant, and normal. The IQ-OTH/NCCD dataset from online source Kaggle has been utilized and in the preprocessing steps, they applied SMOTE and Gaussian Blur techniques to address the class imbalance. They applied VGG16, Resnet50 and InceptionV3 where VGG16 achieved 98.18% accuracy [23].

However, with the same dataset, in 2022, M. Humayun et. al used an augmentation step following rescale, shear range, zoom range and horizontal flip to improve the size of the image dataset. They applied different deep neural networks VGG16, VGG 19, and Xception with 20 epochs, categorical cross entropy, an adam optimizer. Here, authors achieved 98.83% accuracy from VGG16 model to detect lung cancer [24].

Earlier in 2021, researchers also experimented with the same dataset of IQ-OTH/NCCD lung cancer to detect malignant and benign nodules in lung computed tomography images. Initially, Gabor filter is applied to preprocess the data and then with the advent of the transfer learning method GoogleNet, authors achieved 94.38% accuracy [25].

In 2021, A. Mohite classified lung cancer into three classes as benign, malignant and normal. They have collected 1190 image data from IQ-OTH/NCCD lung cancer dataset. A data augmentation technique involving rotations, width shifts, height shifts, and zooming has been utilized before training the data. They applied some deep learning techniques with 50 epochs and 10 batch size including MobileNet, VGG16, VGG 19, DenseNet201 and ResNet101 where DenseNet201 performed best achieving 53% accuracy [26].

The overall comparison of these research works has been summarized in Table 1.

Year	Paper	Objective	Dataset	Preprocessing Techniques	Deep Learning Models	Best Model & Accuracy
2024	[23]	Classify lung cancer into three classes (benign, malignant, normal)	IQ-OTH/NCCD	SMOTE, Gaussian Blur	VGG16, ResNet50, InceptionV3	<b>VGG16:</b> 98.18%
2022	[24]	Detect lung cancer	IQ-OTH/NCCD	Rescale, shear range, zoom range, horizontal flip (augmentation)	VGG16, VGG19, Xception (20 epochs)	<b>VGG16:</b> 98.83%
2021	[25]	Detect malignant and benign nodules	IQ-OTH/NCCD	Gabor filter	GoogleNet	<b>GoogleNet:</b> 94.38%
2021	[26]	Classify lung cancer into three classes (benign, malignant, normal)	IQ-OTH/NCCD	Data augmentation (rotation, width/height shifts, zooming)	MobileNet, VGG16, VGG19, DenseNet201, ResNet101	<b>DenseNet201:</b> 53%

**Table 1. Comparison table of related works on the same dataset**

### **2.2.2 Related works with Different Dataset**

In 2024, to improve the accuracy of lung nodule severity classification, Gautam et al. [27] introduced a new set of DL models that combine ensemble learning with deep transfer learning. They did this by utilizing three advanced CNNs: ResNet152, DenseNet169, and EfficientNetB7. A mechanism for giving each base model a weight was developed in order to improve the performance of the ensemble approach. After a thorough evaluation on the LIDC-IDRI CT scan dataset, the method achieved a 97.23% accuracy rate. The novel weight optimization approach, which dramatically reduced false negatives and raised sensitivity to 98.6%, is credited by the authors with the high accuracy.

An ensemble CNN method was introduced by Madan et al. [28] to detect lung cancer. CT scans and X-ray pictures from two datasets were employed by this technique. Images were annotated first, then data sampling and rescaling have been done in the preprocessing steps. The accuracy of this approach, which used a dataset of 1623 photos, was about 93%, which was less than the accuracy of the CAD technique suggested in this article. In summary, this article's suggested approach yields 99.42% accuracy, which is superior than the findings in [27]. Additionally, recall, precision, and the F-score were all attained at 99.76, 99.88, and 99.82%. In this investigation, VGG-19 and LSTMs were adapted for use.

In 2023, M. Mamun et al. gathered The Cancer Imaging Archive dataset and classified Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma differentiating from the normal lung images. In the preprocessing steps images were resized and then noise removal and segmentation have been done. Different deep learning models i.e. CNN, CNN GD, Inception V3, Resnet-50, VGG-16, and VGG-19 have been utilized and CNN GD gave the highest accuracy of 97.86% [29].

The authors employed deep learning methods in [30] on 613 images from Kaggle Chest CT Scan images dataset. They classified lung cancer into four classes Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma as well normal lung. They offered deep learning techniques such as InceptionV3, Resnet50, and VGG16. The results indicated that the suggested CNN model achieves 94.2% higher accuracy than VGG16, 88.5% higher accuracy, 91.4% higher accuracy for

InceptionV3, and 91.4% higher accuracy for Resnet50. In CNN, the skip connections were used to add output from the previous layer to the next layer which reduce the vanishing gradient problem.

In 2020, QINGHAI ZHANG et al. [31] suggested an automated technique for lung nodule identification design system. A public data set LIDC / IDRI, series of 1018 CT scan images have been gathered for this experiment to classify benign and malignant nodule. To achieve accurate and rapid nodule classification, four distinct Convolutional Neural Network (CNN) architectures were developed and evaluated. These models were trained on datasets comprising pairs of images, each representing a candidate nodule from two distinct classes. The classification was performed across four defined nodule severity levels.

To identify lung cancer, in 2020, Mesut Togacar et al. [32] suggested a CNN-based method. They have collected 100 pictures from 69 distinct patients from the Cancer Imaging Archive, 50 of which are cancerous and the other 50 are not. Augmentation was utilized to obtain an optimal dataset and the study included the usage of CNNs from AlexNet, LeNet, and VGG-16. Weights for each training set were updated using the optimization technique of stochastic gradient descent (for AlexNet and VGG-16). Following CNN designs are several more conventional machine learning models, including LR, LDA, SVM, KNN, and DT. The Principal Component Analysis approach was used to improve the performance. The accuracy of 99.51 was achieved by selecting KNN with CNN & mRMR.

In order to help radiologists diagnose lung cancer, Masood et al. [33] created a sophisticated computer-aided decision support system for lung nodule identification that makes use of a 3D deep convolutional neural network (3DDCNN). The system used multi-region proposal network (mRPN) for automated region-of-interest identification and median intensity projection to take use of three-dimensional information from CT images. The LUNA16, ANODE09, and LIDC-IDR datasets were used to train and verify their CAD system, which showed excellent performance with 98.4% sensitivity, 92% specificity, 96% AUROC, and 98.51% accuracy at 2.1 false positives per scan. The study, however, indicated a restriction in the precision of identifying micronodules with a diameter of less than 3 mm, which called for more research and advancement.

Moreover, to detect early lung cancer and classify its four classes as adenocarcinoma, large cell carcinoma, small cell carcinoma and normal, researchers [34] proposed a deep learning framework using Inception V3, Xception, and ResNet-50 models with 50 epochs and 13 batch size where CNN outperformed the other models with an accuracy of 92%. They collected 967 CT scan image data from online source Kaggle in DCM format. In the pre-processing steps, they resized the image and removed noise.

In 2020, Pang et al. [35] examined CT scans of lung cancer from the Shandong Provincial Hospital using a CNN model based on DenseNet and an adaptive boosting approach for classification. To balance the training data, they employed a number of picture translation and alteration methods. The dataset has 2222 CT images divided into three types: adenocarcinoma, squamous cell carcinoma and small cell carcinoma. After image augmentation procedure, they got 4000 cases. They were able to outperform DenseNet without Adaboost, ResNet, VGG16, and AlexNet with an accuracy rate of 89.85%.

For lung CT scans, Lakshmanprabu et al. [36] presented a novel automated diagnostic technique. In order to evaluate the CT images, it used linear discriminant analysis (LDA) and optimal deep neural networks (ODNN). To identify nodules as benign or malignant, LDR was used to extract and decrease the deep features from CT lung images. Then, in order to classify lung cancer, the ODNN was improved using a modified gravitational search algorithm (MGSA). The accuracy of 94.56%, specificity of 94.2%, and sensitivity of 96.2% were attained by the suggested classifier.

The overall comparison of the above research works has been summarized in Table 2.

<b>Year</b>	<b>Paper</b>	<b>Objective</b>	<b>Dataset</b>	<b>Preprocessing Techniques</b>	<b>Deep Learning Models</b>	<b>Best Model &amp; Accuracy</b>
2024	[27]	Improve lung nodule severity classification	LIDC-IDRI	Weight optimization for ensemble learning	ResNet152, DenseNet169, EfficientNetB7	Ensemble: 97.23%

2024	[28]	Detect lung cancer	Two datasets (CT & X-rays)	Data sampling, rescaling.	Ensemble CNN, VGG19, LSTMs	Ensemble CNN: 99.42%,
2023	[29]	Classify Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and normal lungs	Cancer Imaging Archive	Resizing, noise removal, segmentation	CNN, CNN GD, InceptionV3, ResNet50, VGG16, VGG19	CNN GD: 97.86%
2023	[30]	Classify Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and normal lungs	Kaggle Chest CT Scan Images	Rescale, skip connections to reduce vanishing gradients	CNN, InceptionV3, ResNet50, VGG16	CNN: 94.2%
2020	[31]	Detection of four-stage nodules by integrating two image Scenes.	LIDC/IDR I	Rescale, normalization	A four-channel CNN model	Increased accuracy significantly reducing false positives

2020	[32]	Classify cancerous and non-cancerous lungs	Cancer Imaging Archive	Augmentation, PCA, mRMR, optimization with SGD	AlexNet, LeNet, KNN, LR, SVM	KNN with CNN & mRMR: 99.51%
2020	[33]	Identify lung nodules	LUNA16, ANODE09, LIDC-IDR	Median intensity projection, multi-region proposal network (mRPN)	3D Deep CNN	3DDCNN: 98.51%
2020	[34]	Classified four classes as adenocarcinoma, large cell carcinoma, small cell carcinoma and normal	967 CT scan image data	Image resizing, noise removal	Inception V3, Xception, and ResNet-50, CNN	CNN: 92%
2020	[35]	Classify adenocarcinoma, squamous cell carcinoma, and small cell carcinoma	Shandong Provincial Hospital	Augmentation (translation and alteration methods)	DenseNet with Adaboost, ResNet, VGG16, AlexNet	DenseNet with Adaboost: 89.85%
2019	[36]	Classify benign and	CT Lung Images	Linear Discriminant	Optimal Deep Neural	ODNN with MGSA: 94.56%

		malignant nodules		Analysis (LDA).	Networks (ODNN)	
--	--	-------------------	--	-----------------	-----------------	--

**Table 2. Comparison table of related works on different dataset**

### 2.2.3 Related Works on Explainable Artificial Intelligence

Using MRI and ultrasound images, Hassan et al. [37] investigated the classification of prostate cancer. They used a XAI approach based on deep learning to automatically separate and classify cancers in MRI and ultrasound images of the prostate. By providing physicians with more precise information, the deep learning model was taught to efficiently detect and categorize cancerous areas in the pictures, improving treatment options.

As remarked by Dindorf et al. [38], diagnosing spinal disorders is a major burden for physicians. To create pathology-independent classifiers that can distinguish between abnormal and normal spine postures, deep learning models are used. XAI tools are then used to fully understand the categorization process. By assisting doctors learn about the deep learning models' decision-making process, XAI approaches help them understand the reasoning behind particular outcomes. Furthermore, XAI approaches offer the ability to examine any errors that could occur within the prediction process.

LimeforTime did the lowest at recognizing the important aspects, whereas SHAP variations yielded the best explanations, according to [39]. Decision support systems in delicate fields, such as the marine industry, can be interpreted by XAI. For time series classifiers, the heatmap explanation offers more information than the feature significance plot.

In 2023, Sabbir et al. [40] interpret lung cancer diagnostic system using SHAP and LIME methods. Initially they predicted lung cancer through Machine Learning models, Logistic regression and Random-forest classifiers which scored the highest accuracy of 97%. This study is conducted using the "Lung Cancer Detection" Kaggle dataset.

In 2020, Meske et al. [41] explored how explainability affects AI trust and its potential applications in medical settings. CNN and multi-layer perceptron (MLP) models were employed in their study to Identify malaria based on thin blood smear slide photos.

Chen et al. [42] introduced an Explainable Deep Network (EDN) architecture for image recognition. This concept is inspired by how categorization reasoning involves comparing distinct images components to learnt component prototypes. The proposed EDN performed better than black-box DNNs in two picture classification challenges.

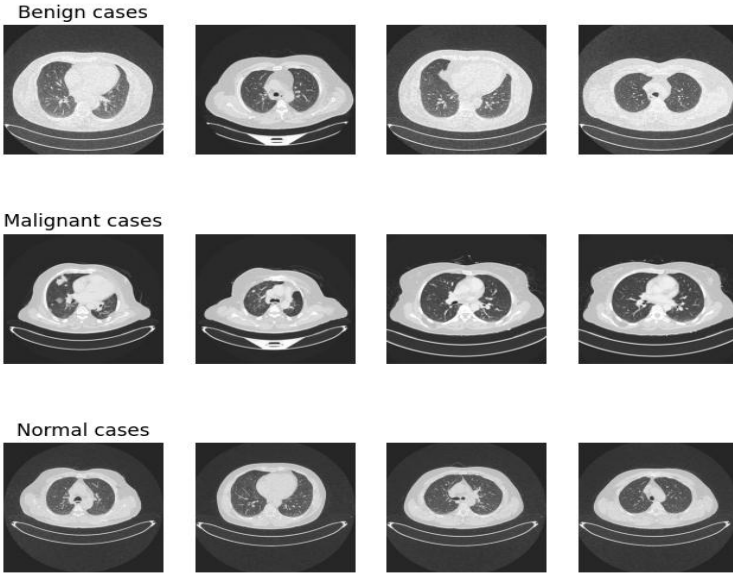
## CHAPTER 3

### METHODOLOGY

This chapter outlines the steps followed to develop the proposed lung cancer prediction system which began with preparing and processing the CT scan dataset, ensuring that the images were suitable for training deep learning models. After this, three individual networks were built and trained, and their outputs were combined through an ensemble approach to achieve more stable and accurate results. The chapter also explains how interpretability methods were applied to understand the model's decisions and how the final model was incorporated into a web application.

#### 3.1 Dataset Description

The Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset is used for our research which was gathered in the Fall of 2019 throughout three months in the specialized hospitals. Approximately 110 cases of lung disease patients who may differ in terms of living conditions, age, gender, and level of education are involved in this dataset. The number of CT scan images gathered in this dataset is 1190 images in total. The images have been divided into three categories including normal, benign, and malignant, and of these, 40 are diagnosed as malignant cases; 15 are benign and 55 are normal cases. Initially, the CT scans were captured in DICOM format where every scan comprises various slices varying in quantity from 80 to 200, each illustrating a different aspect and angle of a human chest. In this scenario, the actual dimensions of 120 benign case pictures are  $512 \times 512$ , 501 malignant cases and 415 normal cases have similar dimensions. However, only one malignant instance had a dimension of  $404 \times 511$ , while 3 malignant cases had a dimension of  $512 \times 623$ . A sample image visualization of every class is presented in Figure 3.



**Figure 3. CT scan images of lungs with three cases**

The proposed methodology of the DeepWeb system, a systematic approach to advanced lung cancer identification is divided into three main phases. The initial phase is a pre-processing phase, which includes various image augmentation techniques. The next phase proceeds by feeding the pre-processed images into our proposed CNN model. Finally, the testing phase identifies lung cancer into different categories.

### **3.2 Image Augmentation**

The process of creating new photos from preexisting ones in order to artificially increase the dataset's size is known as image augmentation. This lessens overfitting and enhances the model's ability to generalize. The measures we took for this procedure are listed below:

- **Rotation:** The rotational augmentation involves rotating the CT scan images by 90, 180, or 270 degrees to change the orientation of the lung structures.
- **Horizontal Flipping:** Flipping the lung pictures in CT scans gives the model additional anatomical arrangement possibilities. Fifty percent of the photos have been flipped horizontally.
- **Affine Transformations:** Translation and scaling are examples of affine transformations that replicate changes in scanner settings and location. While scaling replicates changes in the field of

view, translation adds shifts to the image. The model's ability to adapt to various clinical circumstances is facilitated by these modifications. The photos were separately translated by -10 to 10 pixels on the x and y axes.

- **Cropping and Padding:** Spatial variations are added by randomly cropping or padding the photos by a predetermined pixel range. In medical imaging, where patient postures and scan framing might differ significantly, this is very important. Up to 10 pixels were added to each side of the cropped photos. The value is sampled once and applied to all sides of each image.

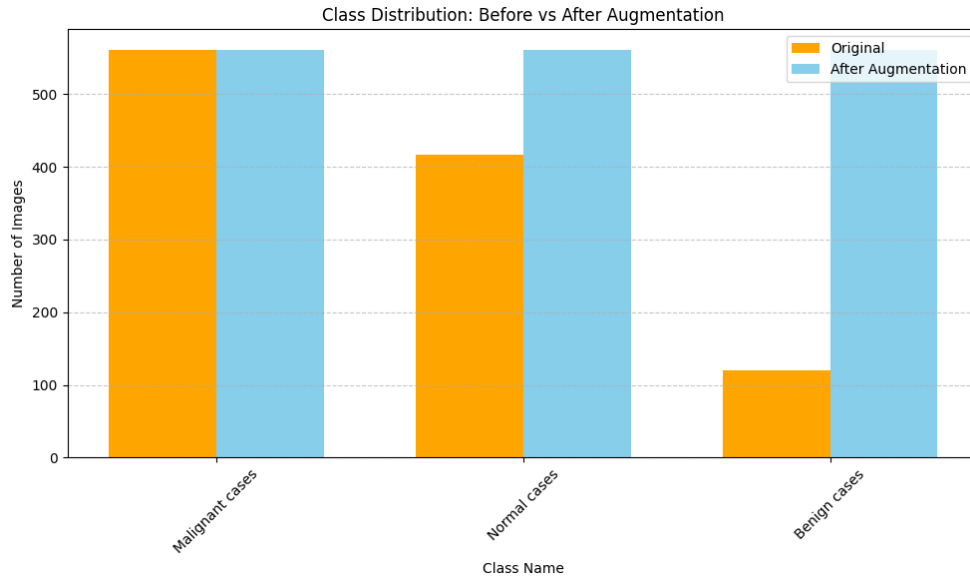
- **Gaussian Noise:** Gaussian noise is added to medical photographs to replicate their inherent noise. This phase strengthens the model's resistance to noisy datasets and improves its capacity for adapting to real-world situations.

- **Linear Contrast Adjustment:** The model can adjust to different lighting situations by adjusting the linear contrast of the images. This is necessary to guarantee the model's resilience in the face of variations in brightness and contrast between various CT images.

Here, Table 3 represents the class distribution of the dataset after the augmentation technique, and Figure 4 represents the bar chart of three class distributions.

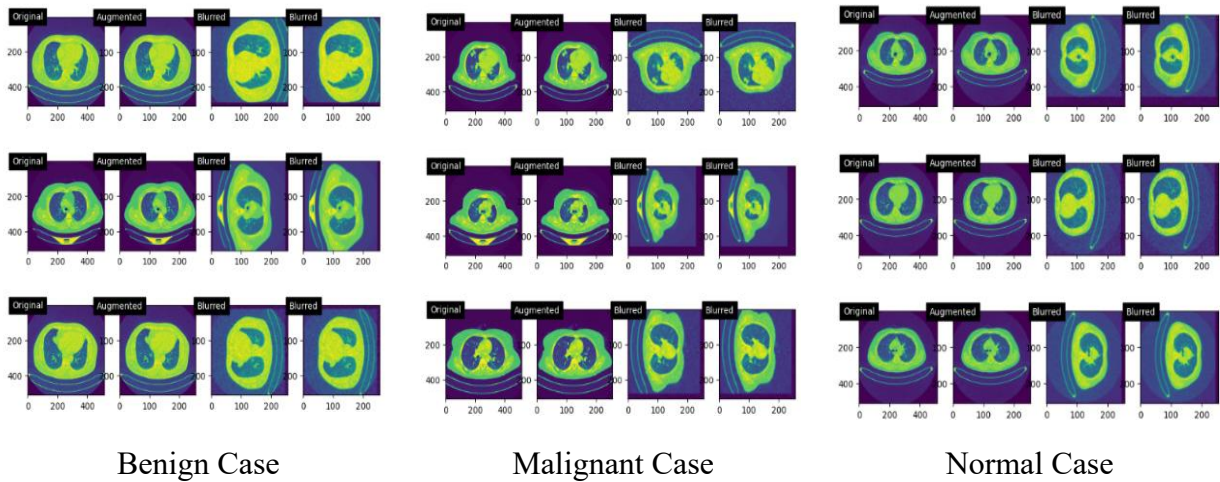
<b>Class</b>	<b>Before Augmentation</b>	<b>After Augmentation</b>
Benign	120	561
Malignant	561	561
Normal	416	561

**Table 3. Class distribution after augmentation**



**Figure 4. Class distribution of our dataset after augmentation.**

The newly created image collection of benign, malignant, and normal instances following augmentation is shown in Figure 5 here.



**Figure 5. Preprocessed Images**

For this, the original grayscale images are represented in the first column of the subplot, the resized images to the specified size are represented in the second column, and the images after Gaussian blur are applied to the resized images are represented in the third and fourth columns.

### 3.3 Dataset Preprocessing

Preprocessing images involves converting unstructured picture data into a format that is fit for computer vision applications. This important stage seeks to demonstrate important traits and eliminate inappropriate anomalies. In order to prepare our image data for input into deep learning models, preprocessing is an essential fundamental step. We have used the following techniques to pre-process our data:

- **Grayscale Conversion:** By converting all images to a conventional grayscale structure, which also maintains consistency and simplifies the amount of input parameters, the model may focus on crucial structural elements rather than irrelevant color information. Images are read in grayscale using `cv2.imread(filepath, 0)`. This reduces the number of channels from 3 (for color images) to 1.
- **Resizing:** Each image is resized to a fixed size of 256x256 pixels using `cv2.resize(img, (img_size, img_size))`. Resizing all images to a uniform dimension ensures compatibility with the network architecture while preserving spatial relationships within the image.
- **Data Structuring:** The resized grayscale images and their corresponding class labels are stored as pairs in a list called data.
- **Shuffling:** The data list is randomly shuffled to ensure that the training and validation sets are representative of the entire dataset and to prevent the model from learning the order of the data.
- **Separating Features and Labels:** The image data (features) and the labels are separated into two lists, X and y.
- **Converting to NumPy Arrays:** The lists X and y are converted into NumPy arrays. X is also reshaped to have the shape `(-1, img_size, img_size, 1)`, which is the expected input format for a convolutional neural network (CNN) with a single channel.
- **Normalization:** To help the model learn more easily, a picture scaling technique, normalization is done in the initial step of preprocessing data with the help of the mean and standard deviation values that bring the image pixel values to a standard scale.

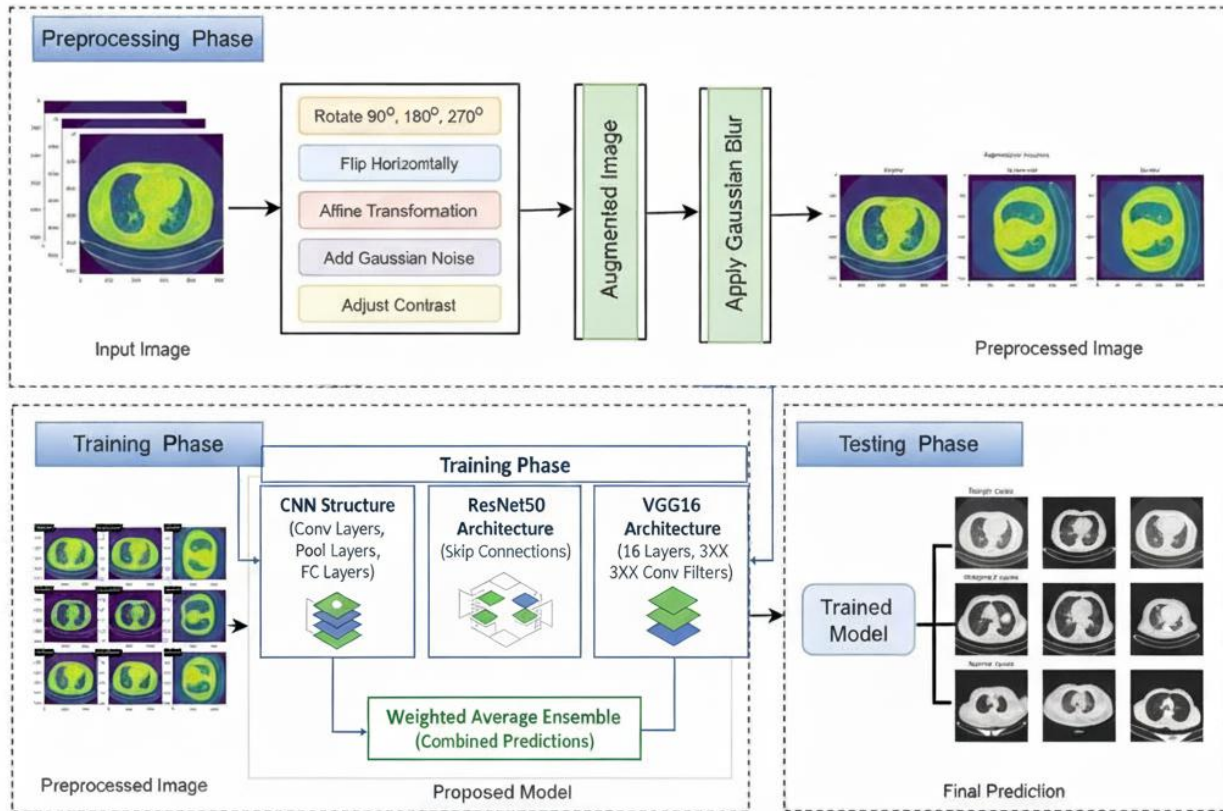
The pixel values in X are normalized by dividing by 255.0. This scales the pixel values to be between 0 and 1, which can help improve model training. This reduces the magnitude

of input values, preventing large gradient updates and improving convergence during training.

These steps prepare the image data from the augmented dataset for use in training a machine learning model.

### 3.4 Lung Cancer Classification using Deep Learning Methods

This study aims to diagnose lung cancer with its three classes benign, malignant, and normal. We have utilized several deep learning models including ResNet50, VGG16, and finally Convolutional Neural Network. A workflow diagram of our proposed method is presented in Figure 6.

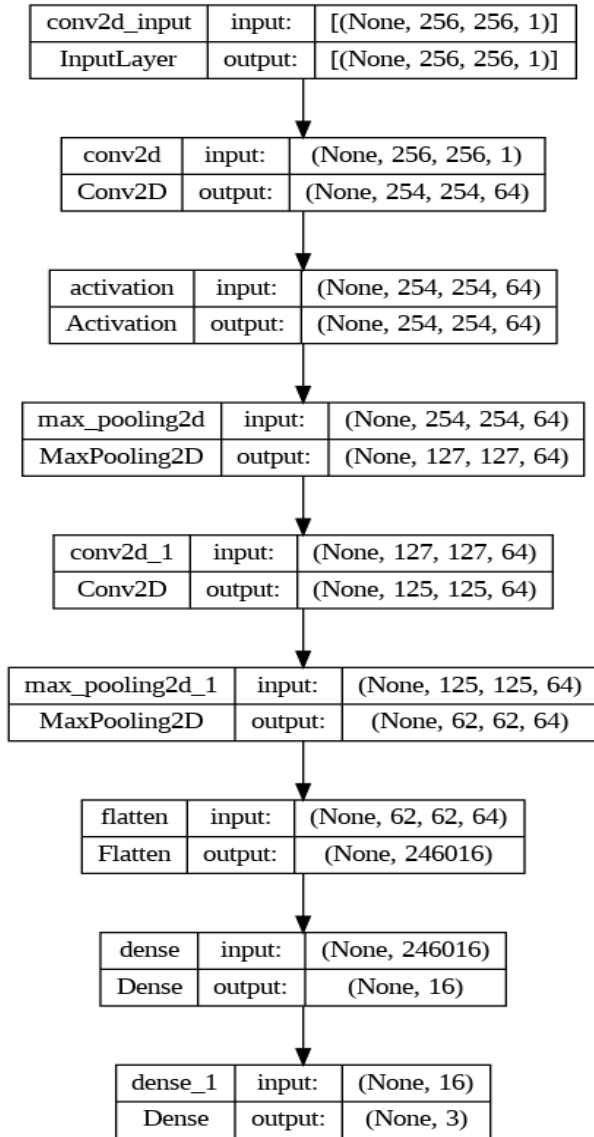


**Figure 6. Workflow diagram for the proposed method of Lung cancer prediction**

At the very beginning, we partitioned the dataset into test, train, and validation parts. Then we created our augmented training dataset by incorporating some preprocessing steps. Finally, with

this newly generated dataset, our proposed model has predicted lung cancer successfully classifying them into benign, malignant, and normal classes we trained our model and then the performance of our model was evaluated with the help of a confusion matrix.

**1) CNN:** Initially, a Convolutional Neural Network (CNN) with the depth of two convolutional layers each followed by a Max-Pooling layer is applied to our dataset. However, it shows satisfactory outcomes with its great efficiency and low computational cost. It carries out parameter sharing and makes use of unique convolution and pooling methods. This makes CNN models widely captivating by allowing them to function on any device. The model architecture is built using TensorFlow and Keras compiled with Adam optimizer which has consistent efficacy in managing complex medical imaging datasets, and categorical cross-entropy for the multi-class classification. We took 30 epochs for our experiment.



**Figure 7. Summary of the CNN model**

Figure 7 represents the summary of the model CNN. The sequential stack of layers in our CNN model is intended to extract and learn feature hierarchies in the following manner gradually:

- **Convolutional layers:** This layer extracts the feature from the input dataset. The first required Conv2D parameter is the number of filters that the convolutional layer will learn. A total of 32 filters is learnt here. For inputs to the CNN, the depth is the number of channels in the image. In our deep convolutional network structure, a 3x3 kernel extracts various features from the input

images in each convolutional layer. The output, known as the feature map, provides us with details about the image, including its edges and corners.

- **Activation Function:** It is one of the most important parameters of the CNN model to become familiar with and estimate any type of complex, continuous relationship between network variables. An activation function rectified linear unit (ReLU) is employed to the output of the previous layer. By introducing non-linearity, it allows the model to extract complex patterns. This activation function is structured as:

$$f(x) = \max(0, x)$$

It sets the input threshold at zero, returning the input itself for positive values and zero for negative ones. ReLU behaves as a linear function with a gradient of one when inputs are greater than zero. This signifies that it permits the gradient to go through unaltered during backpropagation and does not change the scale of positive inputs. This characteristic is required for addressing the issue of the vanish gradient. Even though ReLU is linear for half of its input space, it is technically a non-linear function because it has a non-differentiable point at  $x=0$ , where it abruptly changes from  $x$ .

- **Max-Pooling Layers:** In this layer, the highest value within a specific region of a feature map is identified. It takes the feature map from the convolutional layer. By selecting these maximum values, the resulting feature map highlights the most important characteristics of the previous layer. Here, 2x2 pool size is applied and the window is shifted by stride 2 which means it determines the maximum pixel value and reconsiders it as an output when stride 2 slides the filter. The output matrix reduced spatial dimensions as shown in Figure 8.

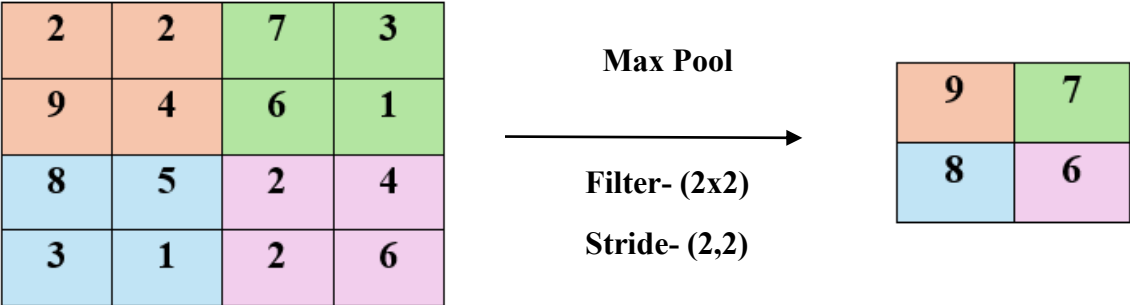


Figure 8. Max-pooling Layer

For a feature map having dimensions  $n_h \times n_w \times n_c$ , the dimensions of output obtained after a pooling layer is:

$$(n_h - f + 1) / s * (n_w - f + 1) / s * n_c$$

Here,

->  $n_h$  - height of feature map

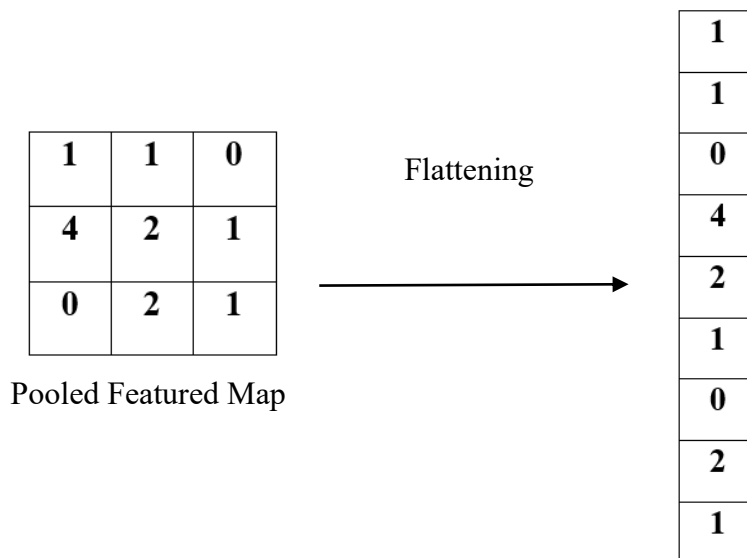
->  $n_w$  - width of feature map

->  $n_c$  - number of channels in the feature map

->  $f$  - size of filter

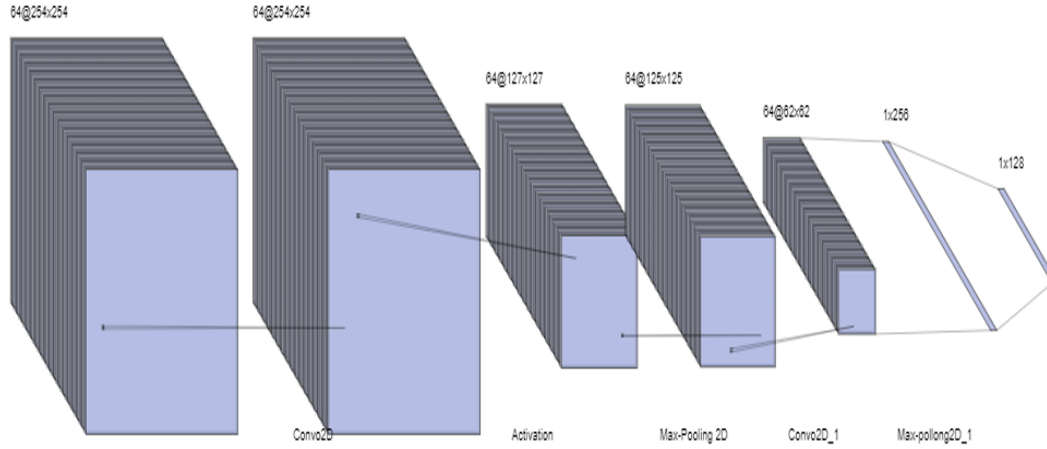
->  $s$  - stride length

• **Flatten and Fully Connected Layer:** The neurons are connected between two different layers via the structure of the Fully Connected (FC) layer, which also includes weights and biases. By reducing the dimensionality of the data, the Flatten layer can help lowering the number of parameters in the fully linked layers that follow. The flatten layer transforms the multidimensional input image array into a 1D array fed to the FC layer as well as the classification procedure gets started. An example has been represented in Figure 9.



**Figure 9. Flattening layer**

Each neuron in one layer is connected to every other layer's neuron through fully connected layers. After being selectively initiated by the pooling layers, the convolutional layers learned representations are combined to create predictions. In this case, flatten layer reshapes the feature map from input (62, 62, 64) to a 1D vector with  $(62 \times 62 \times 64) = 246016$  elements.



**Figure 10. Structure of the CNN model**

In Figure 10, the structure of the Convolutional neural network is shown where all the above-mentioned layers are described. Unlike the other deep learning model, the proposed CNN architecture does not include any dropout layer as there is no overfitting in training and validation accuracy.

One of the key equations in this module is the computation of the output image:

$$X_1 = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} \sum_{c=0}^{C-1} X(i+m, j+n, c) \cdot W(m, n, c) + b \quad (1)$$

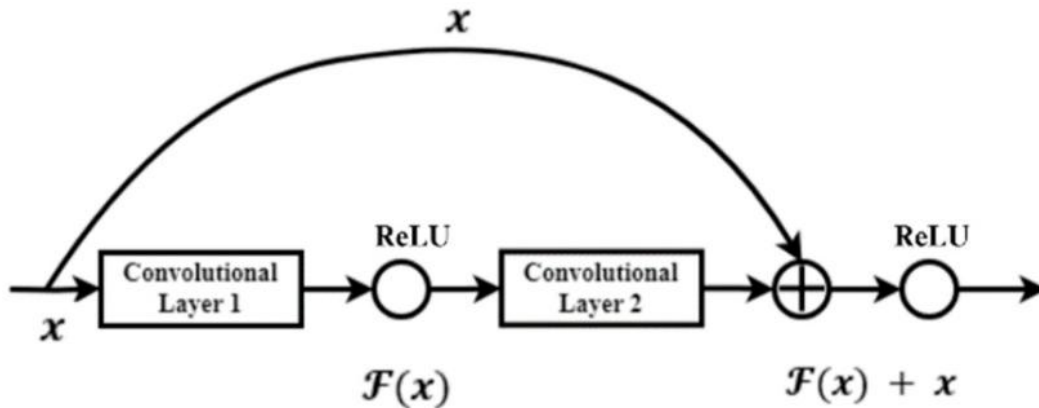
Eq. (1) here works for evaluating the convolution layer.  $X_1$  denotes the output for this layer. It takes the input as  $X(i,j)$  as the pixel value at position  $(i, j)$ .  $W(m,n)$  is the weight value at position  $(m,n)$ ,  $b$  is the bias term and  $c$  is the number of input channels. In our case, the number of channels is 3.

$$X_2(i, j, c) = \max_{m,n} X_1(i.s + m, j.s + n, c) \quad (2)$$

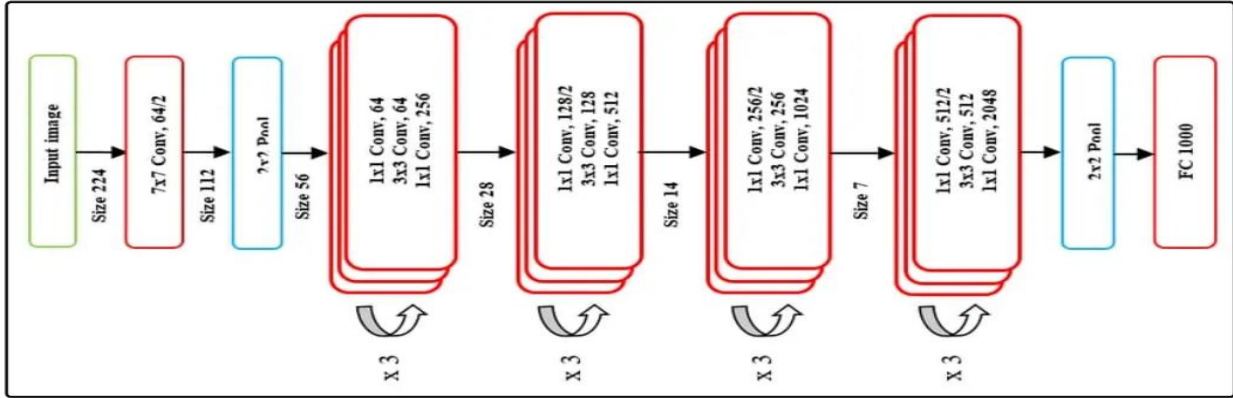
And the output of each max pooling layer is generated by Eq. (2). The output of the convo layer is used to analyze each max pooling level where  $X_1$  is the input image, and  $s$  is the stride, which represents the step size when moving the pooling window. Here  $\max(m,n)$  is used to find the maximum value within a specific region of the input feature map.

Though our proposed model CNN gave an outstanding performance, we have implemented two more deep learning models ResNet50 and VGG16 to compare them with CNN. The other two models have been described as below:

**2) ResNet50:** This model simplifies network training while reducing the issue of network deterioration. We have applied a pre-trained deep neural network ResNet50 that is 50 layers deep on our dataset. At the end of its construction, this model has 49 convolutional layers and one fully-connected layer. By introducing skip connections, often referred to as residual connections, this architecture enables the gradients to flow directly through other bypass channels and thus reduce the problem of vanishing gradients. We can train an effective deep neural network by using residual blocks. Figure 11 shows the Residual block or skip connection on ResNet [43] and Figure 12 demonstrates the architecture of ResNet50 [44].



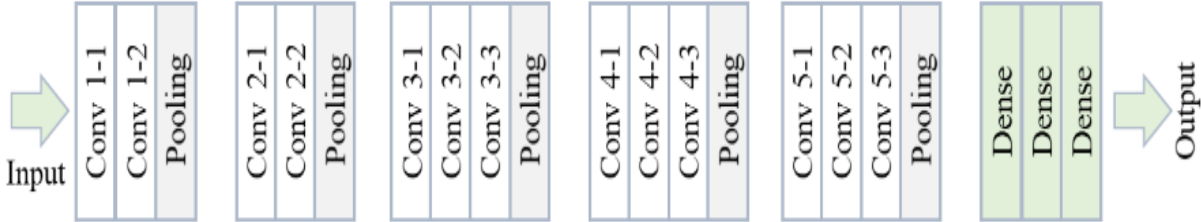
**Figure 11. Residual block or skip connection on ResNet50.**



**Figure 12. Architecture of ResNet50.**

During our experiment, we resized images in 224x224 as the input image. We have used the imagenet for weights, and Global Average Pooling to reduce the spatial dimensions of the output tensor. We have also applied the Adam optimizer and sparse categorical crossentropy. To train the model we took 30 epochs and batch size=16 and softmax activation function.

**3) VGG16:** Then we applied VGG16 another deep neural network with 16 layers including 13 convolutional layers and 3 fully connected layers where the input image size is 224x224. We have used and 3x3 filters in convolutional layers, the imagenet for weights, Global Average Pooling to reduce the spatial dimensions of the output tensor. We have also applied adam optimizer and sparse categorical cross-entropy. To train the model we took 30 epochs and batch size = 16 and softmax activation function. While VGG16 is very good at extracting low-level attributes from images, it is highly computational due to its enormous number of trainable parameters. Figure 13 shows the architecture diagram of VGG-16 [45].



**Figure 13. VGG-16 architecture diagram**

### **3.5 Lung Cancer Classification using Weighted Average Ensemble Methods**

In this research, a deep learning-based ensemble framework was designed with the goal of improving classification performance and achieving more consistent predictions across varying input samples. Instead of depending on a single network, the approach brings together several well-established convolutional neural network (CNN) architectures. Since individual models tend to learn different aspects of the data, combining them helps capture a wider range of features and lowers the likelihood of classification errors.

The ensemble consists of three models: a custom Convolutional Neural Network (CNN), VGG16, and ResNet50. The custom CNN provides a foundational feature extractor designed specifically for the target dataset. VGG16 contributes strong spatial feature-learning through its sequential convolutional layers, which capture fine-grained details. ResNet50 provides the benefit of broader representation learning using residual shortcuts, which solve vanishing-gradient concerns and aid in the successful training of deep structures. These three networks work together to provide a wide and thorough representation of features.

This fusion strategy assures that all models have an equal effect on the final conclusion, avoiding any one design from dominating the forecast. As a result, the ensemble produces more reliable and precise findings than any single model.

The procedure for ensemble prediction is as listed below:

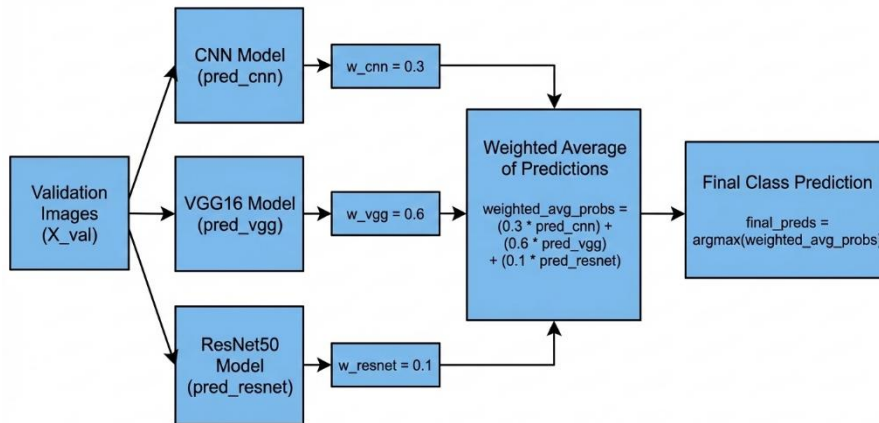
1. **Independent Predictions:** Each model produces a vector containing expected probability for the three classifications. These probabilities represent the model's confidence in each class for a certain image.
2. **Weighted Combination:** In order to adjust the variances in each model performance, a weighted average technique is used. Each model's output probabilities are multiplied by a weight based on relative validation performance and refined empirically to maximize overall accuracy while maintaining model diversity, which is a commonly adopted and effective strategy in applied medical image analysis. For example, a higher weight is given to the best-performing model (VGG16), a moderate weight to the custom CNN, and a lower weight to ResNet50.

Mathematically, for each image:

$$P_{ensemble} = w_{CNN} \cdot P_{CNN} + w_{VGG} \cdot P_{VGG} + w_{ResNet} \cdot P_{ResNet}$$

where  $P$  represents the predicted probability vector from each model and  $w$  the assigned weight.

3. Final Prediction: The combined probability vector is examined, and the class with the highest weighted probability is selected as the final prediction for that image. Figure 14 represents the weighted average ensemble model for our study.



**Figure 14. Weighted average ensemble model for final prediction.**

### 3.6 Model Performance Evaluation Measures

- Accuracy: It is the measure of entire correctness across all categories. This calculates the level of correctness or wrongness of the model's predictions by Eq. (3). Here TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative respectively

Accuracy = correct classifications / total classifications

$$\frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

- Precision: It estimates the frequency of accurate predictions for the positive class. This indicates the quality of a positive prediction and can be measured with Eq. (4).

Precision = correctly classified actual positives / everything classified as positives

$$\frac{TP}{TP+FP} \quad (4)$$

• Recall: It quantifies the model's ability to recognize every positive case in the dataset. It is also known as sensitivity and measures actual true positive rate by the Eq. (5)

RE = correctly classified actual positives / all actual positives

$$\frac{TP}{TP+FN} \quad (5)$$

• F1 Score: It is also known as F-measure and described as the harmonic mean of the precision and recall of a classification model. This accurately reflects a model's reliability because both precision and recall metrics make equal contributions to the score by the Eq. (6)

$$F1\text{-Score} = \frac{2 \times PR \times RE}{PR + RE} \quad (6)$$

• Confusion Matrix: A confusion matrix is an NxN matrix used to specify how well a classification system performs on a test data.

- True Positive (TP): The model correctly predicted a positive outcome (the actual outcome was positive).
- True Negative (TN): The model correctly predicted a negative outcome (the actual outcome was negative).
- False Positive (FP): The model incorrectly predicted a positive outcome (the actual outcome was negative). Also known as a Type I error.
- False Negative (FN): The model incorrectly predicted a negative outcome (the actual outcome was positive). Also known as a Type II error.

### 3.7 Explainable AI

There are several Explainable AI techniques from which we used the most recent and efficient technique SHapley Additive exPlanation (SHAP). It is an approach influenced by game theory that estimates significance values for each characteristic for individual predictions in an effort to improve interpretability [46]. In this study, LIME is interpreted for highlighting key image regions

influencing individual predictions, while SHAP offers global and local feature attributions across the dataset. The implementation of LIME and SHAP is discussed below:

### **3.7.1 LIME Implementation**

Local Interpretable Model-Agnostic Explanations (LIME) takes an image of CT scan defined as “malignant” while modifying some parts of the image to understand which part of the image is mostly responsible for the prediction.

By integrating local sampling, distance-based reweighting, and a streamlined approximation approach, LIME offers a targeted, instance-level explanation. LIME operates under the premise that even extremely non-linear models behave pretty simply in a very small region around a single input, rather than attempting to read the entire model, which is typically impractical for deep networks. It checks how the model’s prediction changes for each version. The steps for the LIME implementation are:

1. Load Pre-trained Model: The already trained Weighted Average Ensemble model is loaded.
2. Modifying Image data: LIME creates 1000 perturbed images for each model of CNN, VGG16, and ResNet50.
3. Model Prediction Interpretation: The trained ensemble model’s prediction function returns class probabilities then argmax gets the class with the higher predicted probability.
4. Generating LIME explainer: LIME divides the images into superpixels, feeds each modified image to the trained ensemble model. It measures how much the superpixels affects model’s prediction. This fits a local linear model to identify the most influential superpixel in predicting the class.
5. Visualizing final decision: Visualized the regions that drove the model’s decision.

### **3.7.2 SHAP Implementation**

The SHAP method works by computing the Shapley values for each feature in the input space. By calculating the difference in the model's output between the presence and absence of a feature, SHAP values indicate how important a feature is. The steps for the SHAP Implementation are described below:

1. Load Pre-trained Model: As we proposed Convolutional Neural Network (CNN) for lung cancer classification, so our previously trained CNN model has been loaded first.
2. Prepare Image Data: Then we images were pre-processed.
3. Define the Model Prediction Function: Create a function that preprocesses input data and returns model predictions.
4. Create a SHAP Explainer: Use SHAP to create an explainer object with the model and image masker.
5. Generate SHAP Values for Images: Compute SHAP values for a subset of images.
6. Visualize Feature Importance: Use `shap.image_plot` to visualize the SHAP values.

The masker masks out partitions of the image using a blurring method Telea. The Shapley value for feature  $i$ , denoted by  $\phi_i$ , is defined as the average contribution of the feature  $i$  across all possible coalitions of features. Mathematically, the Shapley value can be expressed as follows:

$$\phi_i(f, S) = \sum_{T \subseteq S \setminus \{i\}} \frac{|T|!(|S| - |T| - 1)!}{|S|!} (f(T \cup \{i\}) - f(T))$$

where  $T$  is the set of all input features,  $S$  is a subset of features that does not include feature  $i$ ,  $|S|$  is the size of the subset, and  $f(S \cup \{i\})$  is the output of the model when the features in  $S$  and  $i$  are present. The term  $f(S)$  is the output of the model when only the features in  $S$  are present. The Shapley value is the average marginal contribution of feature  $i$  over all potential coalitions.

SHAP uses a color-coded scheme to highlight the elements that influence a model's predictions:

- Red Pixels: These pixels represent characteristics that improve the model's prediction. In other words, they improve the possibility that a specific class is allocated.
- Blue Pixels: Conversely, blue pixels highlight features that negatively impact the prediction. They decrease the probability of a specific class.

## CHAPTER 4

# WEB APPLICATION DEVELOPMENT

This chapter outlines how the trained model was incorporated into a simple web platform that can analyze CT images and provide predictions in real time. It emphasizes the main design choices, system flow, and components that make the application usable in a practical medical setting.

### 4.1 Web Application Development Process

A web application has been developed leveraging a Python based web framework, Streamlit, in combination with Pyngrok to gain safe access to our local Streamlit server to the internet. To employ this, a few steps are followed as mentioned in Figure 15:

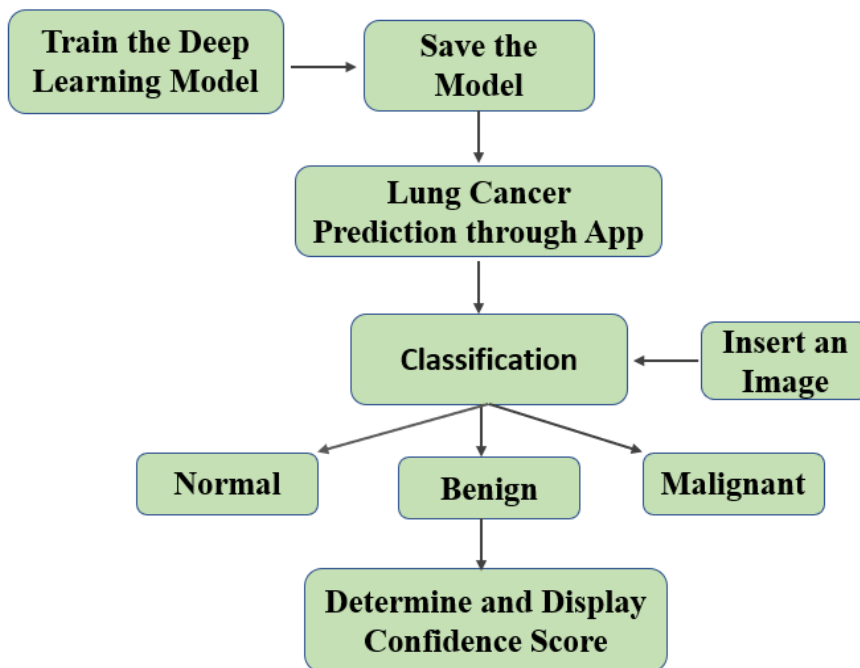


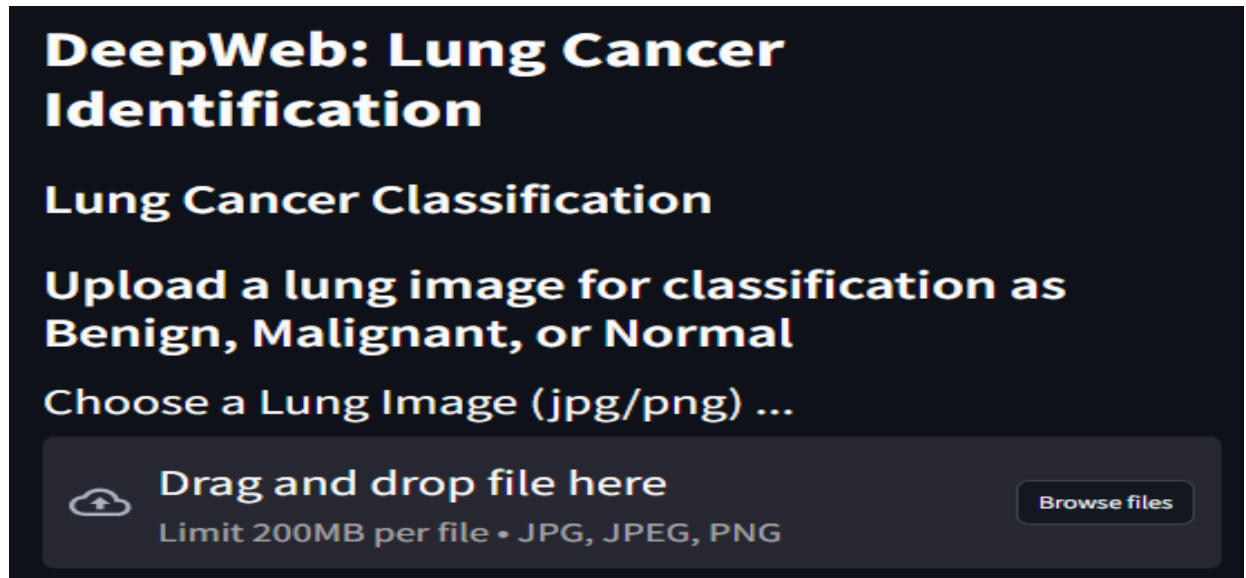
Figure 15. Workflow of our web application

**1) Save the Model:** We started the model saving procedure when we saw how well our ensemble model performed. A separate directory was created to contain the model, which would be the foundation for our upcoming web application development.

**2) Preprocess the Image:** Image preprocessing is an essential stage in our approach. We transform RGB photos to grayscale to improve computing performance while keeping important characteristics. Each picture is then resized to 256 by 256 pixels to adhere to the input shape required by our model. The resized photos are then normalized, and the channel dimension is adjusted to an axis (0, 1) to meet the model's input criteria.

**3) Make Prediction with Confidence Score:** Using Pyngrok and Streamlit in the Google Colab environment, we succeeded in bringing our suggested model online. Following preprocessing, user-uploaded photos are classified by the model into one of three categories: benign, malignant, and normal. In addition, our system assigns a percentage-based confidence value to each forecast. This value is the chance that our system properly recognized the image's class.

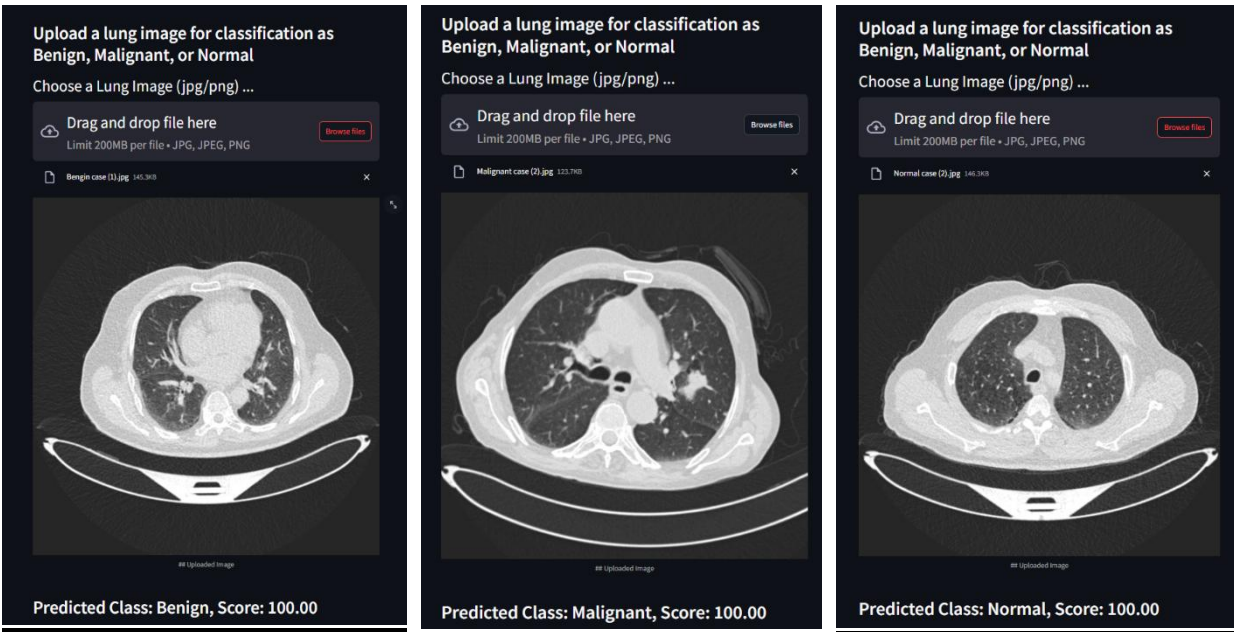
**4) Link to Web App:** The Ngrok command generates a public URL within the terminal. This URL provides easy access to our Streamlit application from any web browser. Furthermore, consumers may interact with our model remotely by accessing the application via the specified URL. Figure 16 shows the user interface of our online program for classifying lung cancer. Furthermore, users can categorize lung cancer by uploading a CT lung picture in jpg, jpeg, or png format from their device of up to 200 MB.



**Figure 16.** UI of our Web application.

## **4.2 Lung Cancer Prediction with the Web Application**

Our web application can accurately forecast the presence of lung cancer in three different categories: benign, malignant, and normal. It assigns a confidence score to each prediction, reflecting the degree of certainty associated with the categorization. Figure 17 illustrates the application's output. In this scenario, the confidence levels to forecast benign, normal, and malignant lung cancer are all 100%. The higher the confidence level, the more certain the user is that the forecast will meet his or her expectations. This high degree of confidence suggests that our proposed model can generate predictions that are compatible with user expectations.



Benign Case

Malignant Case

Normal Case

**Figure 17. Lung Cancer prediction with confidence score**

Streamlit functions as the project's unifying framework for frontend and backend development, enabling the production of a dynamic user experience. We use Pyngrok, an effective tool that rapidly gets over the constraints imposed by NAT or firewall installation, to guarantee that our local server has regular internet access.

Our web application boasts a range of compelling features:

- ✓ **User-Friendliness:** This application offers accessibility for a broad range of users, with minimum prior experience of web development required.
- ✓ **Versatility:** It can be readily applied in a range of conditions due to its clarity of design and flexibility.
- ✓ **Real-time Responsiveness:** It enables dynamic updates in real time that make any modifications according to the associated context.

- ✓ Enhanced Security: Secure end-to-end encrypted connections are delivered by Pyngrok, which enhances the security of this application.
- ✓ Controlled Access: To safeguard application access, various authentication methods have been established.

The ERD diagram in Figure 18 illustrates the web application's working flow. The entities of the diagram are:

- User: Indicates a user with permission to upload photos.

Attributes:

- UserID: This is the Primary Key which is a unique ID for each user.
- Username
- PasswordHash
- Email
- CreatedAt

- Image: Reflects photos that the user has submitted. Multiple images can be uploaded by a single person.

Attributes:

- ImageID: Primary key
- UserID: Foreign key that references User.
- Filename
- FilePath
- Format
- FileSize
- UploadTimeStamp

- Model illustrates the ensemble model which employs a picture to predict lung cancer.

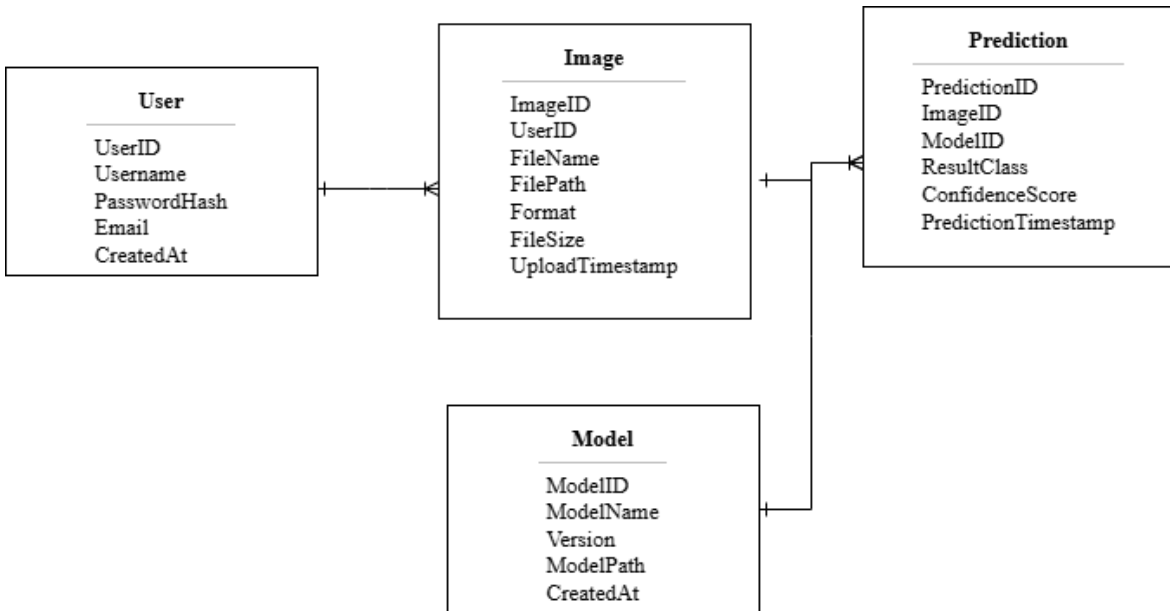
Attributes:

- ModelID: Primary key
- ModelName
- Version
- ModelPath

- CreatedAt
- Prediction: Shows the result after analyzing an image.

Attributes:

- PredictionID: Primary key
- ImageID: Foreign key that references Image
- ModelID: Foreign key that references Model
- ResultClass: Predicted class
- ConfidenceScore: The confidence score of the model's prediction
- PredictionTimeStamp



**Figure 18. ERD diagram for web application**

## **CHAPTER 5**

### **EXPERIMENT & RESULT ANALYSIS**

To correctly examine the efficiency of each deep learning model deployed in our research, we conducted a comprehensive evaluation that comprised a number of vital criteria. These metrics included an F1 score, recall, accuracy, and precision. For every model, we also produced a confusion matrix that shows its effectiveness as it performs in classification. The accuracy of a model is determined by how well its predictions appear.

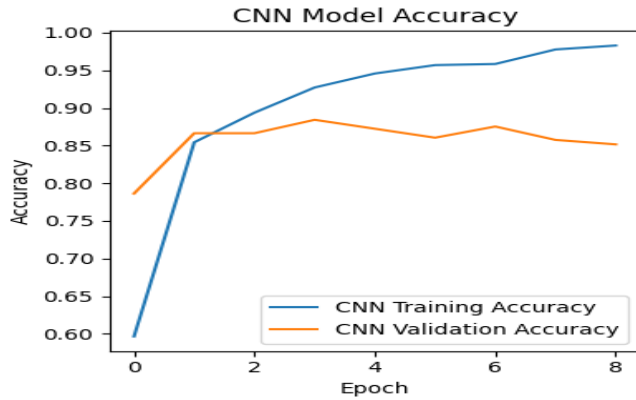
#### **5.1 Performance Analysis of Different Deep Learning Approaches**

Three deep neural networks are combined in a weighted structure to establish the basis of the proposed system. The ultimate decision will be impacted by each model's forecast based on the weight assigned to it.

Firstly, the CNN model that was built performed exceptionally well in the specified condition. To thoroughly investigate its potential, we applied a contrasting evaluation technique, comparing its performance to the well-known ResNet50 and VGG16 architectures.

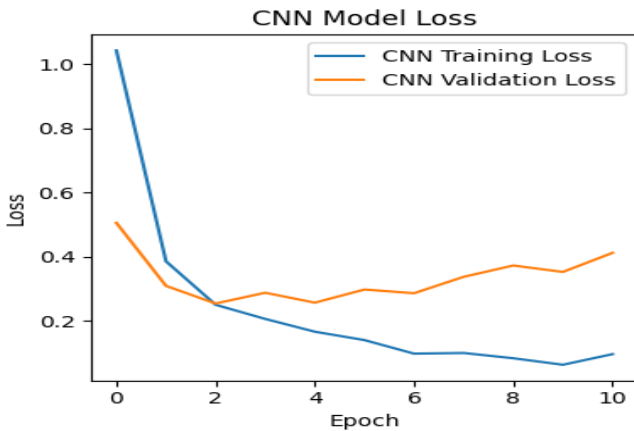
##### **5.1.1 Convolutional Neural Network**

During training, the custom CNN model showed a distinct and steady learning tendency. With each epoch, the accuracy increased while the loss reduced, implying that the model was adequate to the underlying patterns in the dataset. To avoid overfitting, an early-stopping mechanism was utilized, enabling the model to preserve the set of weights that performed best on the validation data rather than training perpetually. On the validation set, the CNN achieved an accuracy of 0.88, as seen in Figure 19 which implies the identification between the various lung CT scan classes with a high degree of precision. The dense layers assisted in transforming the learnt features into relevant class outputs, while the convolutional and pooling layers were vital in identifying the significant visual structures in the images. Overall, The CNN performed consistently contributing significantly to the ensemble model.



**Figure 19. Training accuracy vs validation accuracy curve of CNN**

When it comes to diagnosing illnesses, false positive results could have negative implications. Improper predictions may result in redundant and potentially hazardous therapy, and subject patients to unnecessary dangers and undesired impacts. That’s why, limiting false positives is vital. Figure 20 demonstrates the training and validation loss for the model CNN.



**Figure 20. Training loss vs validation loss curve of CNN**

To systematically evaluate the CNN model's performance, we performed a complete performance study using important measures such as the F1-score, recall, and accuracy. These metrics give a

holistic assessment of the model's capacity to correctly determine true positives, reduce false positives, and detect all cases of each illness class.

We compared the results of the Convolutional Neural Network (CNN) model with two popular deep learning algorithms, VGG16 and ResNet50, to evaluate the efficiency and other potential benefits.

### 5.1.2 VGG-16

Among the individual models evaluated in this work, VGG16 delivered the most convincing results. Since the convolutional layers were kept unchanged, the model could take advantage of the detailed feature representations already learned from large-scale image collections. Only the classifier section was retrained, which helped the network adjust these features to the characteristics of lung CT scans without overfitting. Throughout training, the model showed stable learning behavior and gradually improved its predictive ability. On the validation dataset, VGG16 reached an accuracy of 0.93, represented in Figure 21. It demonstrates that its strong capability in separating the different classes. The deeper structure of the network enabled it to capture fine-grained variations in tissue appearance and other subtle cues found in medical images. As a result, VGG16 stood out as the most reliable and informative component within the overall ensemble setup.

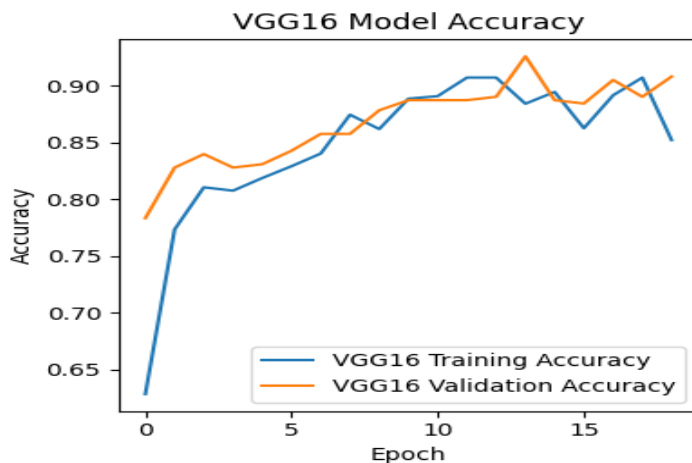
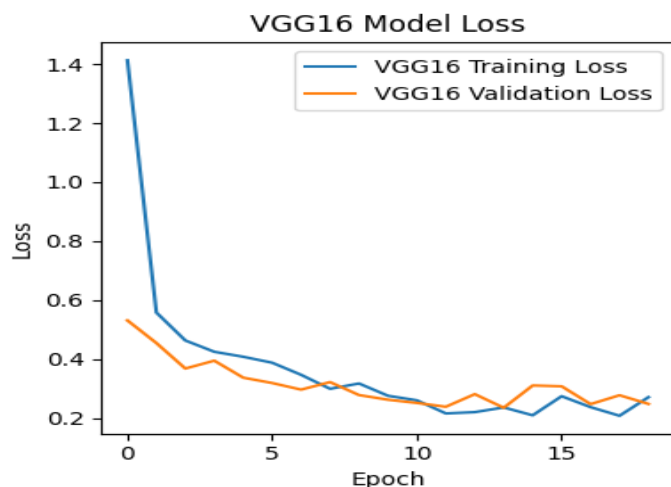


Figure 21. Training accuracy vs validation accuracy curve of VGG16

Figure 22 reflects the loss curve having the training loss and validation loss of the VGG16 model.

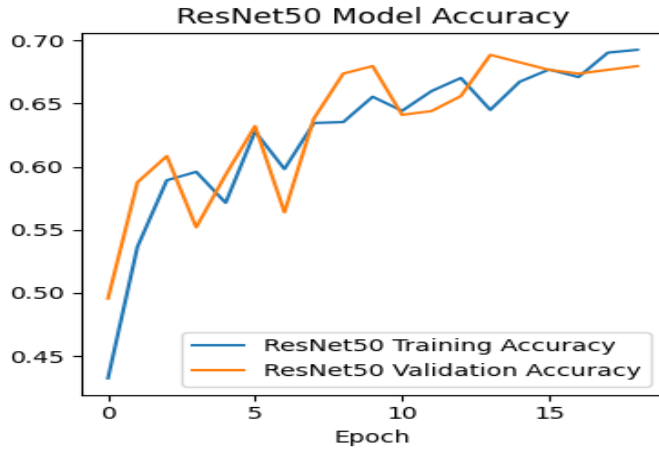


**Figure 22. Training loss vs validation loss curve of VGG16**

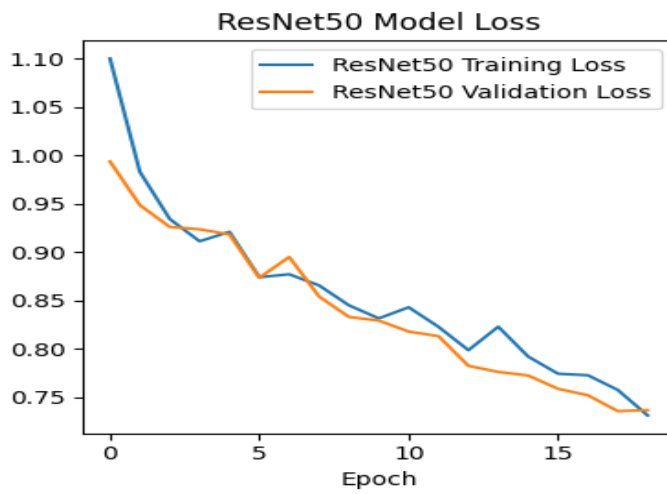
### 5.1.3 ResNet50

The performance of the ResNet50 model showed a moderate level of effectiveness when applied to the lung CT scan dataset. Because the pre-trained layers were kept fixed, the model mainly relied on the residual blocks to extract high-level visual information, while only the final layers were adjusted during training. As a result, the model's accuracy improved gradually but did not match the rate of improvement seen in the simpler architectures. When evaluated on the validation set, ResNet50 achieved an accuracy of 0.69, indicating that it could recognize some useful patterns but was less consistent than the other models used in this study.

To acquire additional information regarding the accuracy of the model, Figure 23 and 24 shows the accuracy and loss curve of ResNet50 respectively. This visual representation provides a detailed breakdown of the model's predictions, highlighting instances of correct classifications, as well as instances where the model misclassified samples into different classes.



**Figure 23. Training accuracy vs validation accuracy curve of ResNet50**

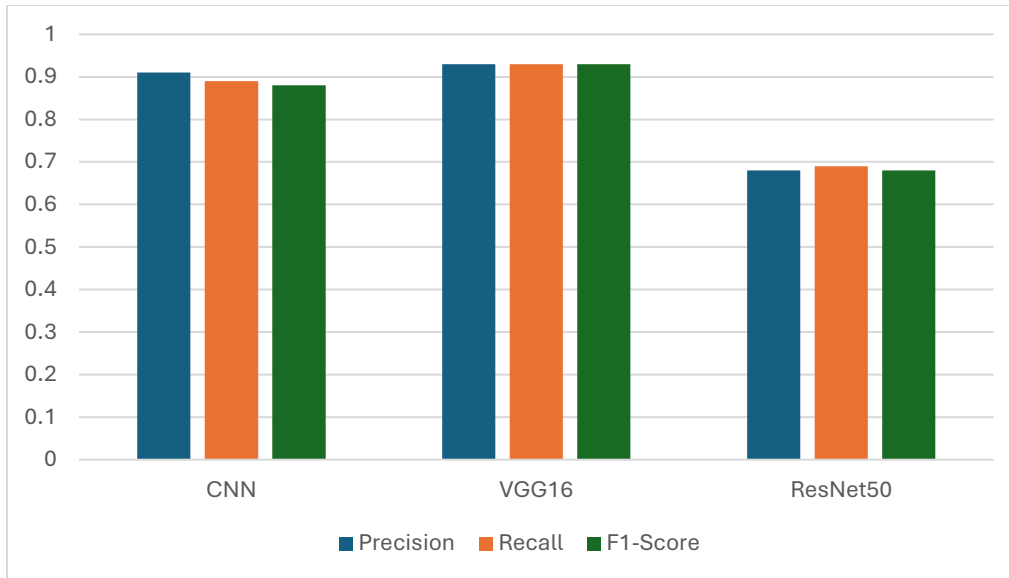


**Figure 24. Training loss vs validation loss curve of ResNet50**

Table 4 and Figure 25 reflect the performance evaluation among the three deep learning models we have used.

Algorithm Used	Precision	Recall	F1-Score
CNN	0.92	0.92	0.92
VGG16	0.93	0.93	0.93
ResNet50	0.68	0.69	0.68

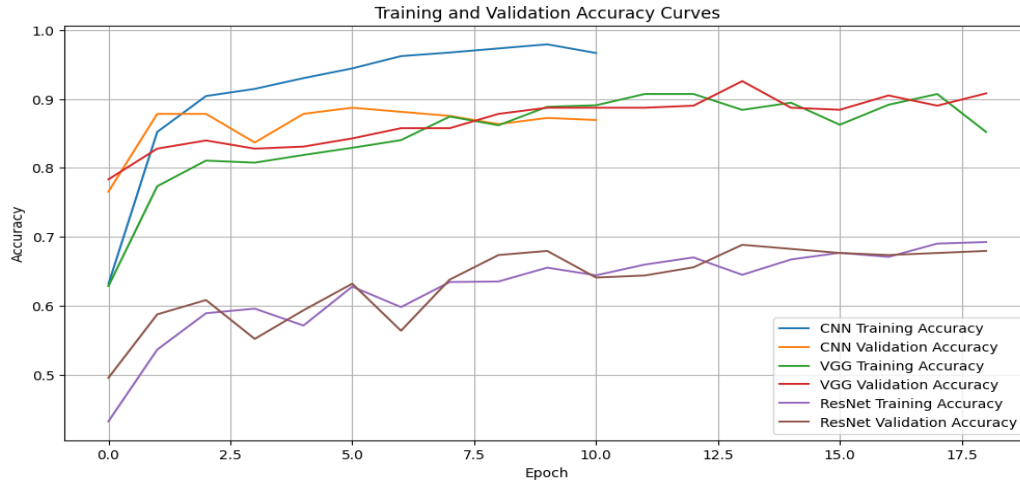
**Table 4. Performance evaluation of the three deep learning models**



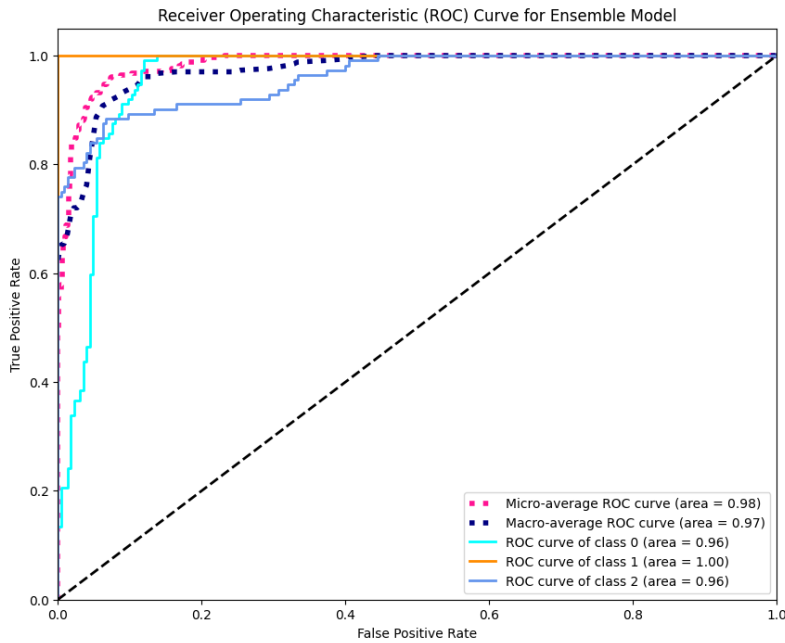
**Figure 25. Performance comparison of deep learning models**

## 5.2 Performance Analysis of Weighted Average Ensemble Model

The performance of the proposed weighted average ensemble model is evaluated on IQ-OTH/NCCD dataset. By combining the softmax probability outputs of CNN, VGG16, and ResNet50, a final prediction is made achieving an accuracy of 94%. This ensemble model also reflects 0.94, 0.94 and 0.3 of precision, recall and f1-score respectively. The macro-averaged precision, recall, and f-1 score shows consistent improvement indicating that the ensemble model effectively captured complementary features from each base model. Figure 26 and Figure 27 show the training vs validation accuracy curve and Receiver Operating Characteristic (ROC) curve for ensemble model integrating all three models respectively.



**Figure 26. Training accuracy vs validation accuracy curve for ensemble model**

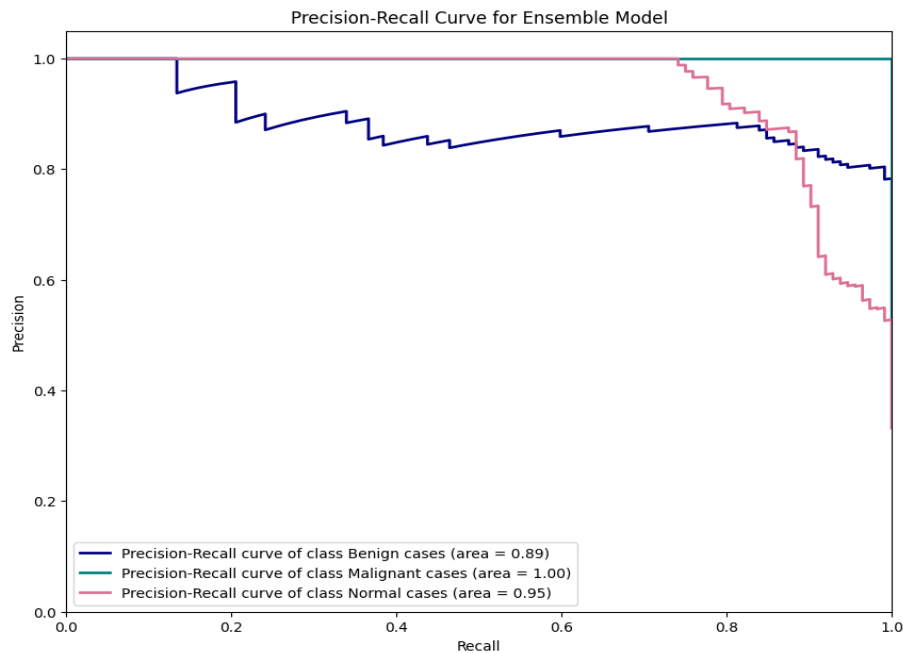


**Figure 27. Receiver Operating Characteristic (ROC) curve for ensemble model**

The ensemble model's Receiver Operating Characteristic (ROC) analysis demonstrates higher Area Under the Curve (AUC) compared to individual deep learning models indicating strong discriminative capability across benign, malignant, and normal cases.

The Precision-Recall (PR) curve was also used to evaluate the model's overall performance. It demonstrates the classifier handles positive instances under various threshold conditions effectively which is significant when class frequencies are unevenly distributed.

Figure 28 depicts the PR curve, which graphs accuracy on the vertical axis and recall on the horizontal. Precision indicates the percentage of correct positive predictions, whereas recall shows the model's capacity to recover real positive cases. A well-performing model is often shown by a curve that increases toward the plot's upper-right corner, showing that precision and recall stay high at the same time. The area under this curve, known as Average Precision (AP), serves as a single, interpretable score that summarizes the model's trade-off between these two measurements.



**Figure 28. Precision-Recall (PR) curve for ensemble model**

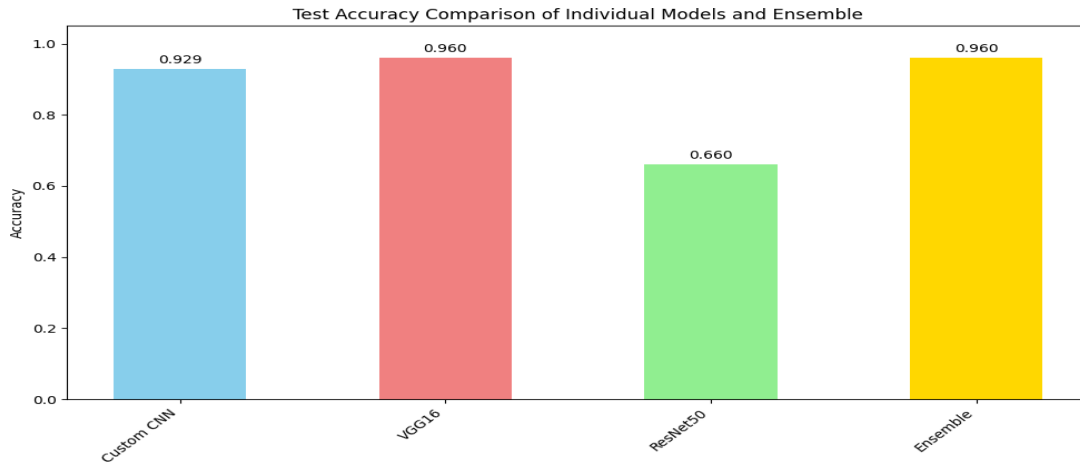
From the above figure, the class-wise curves have been demonstrated as below:

- i. Class 0 (Benign): The obtained AP value of 0.89 shows that the model maintains a reasonable balance between identifying benign samples correctly and reducing false positives.

- ii. Class 1 (Malignant): This class achieved an AP of 1.00, suggesting that the model is extremely reliable in recognizing malignant cases, with virtually no loss in either precision or recall.
- iii. Class 2 (Normal): With an AP of 0.95, the classifier demonstrates strong capability in distinguishing normal cases as well.

A micro-average AP of 0.97, which considers all individual predictions regardless of class, further confirms the stability of the ensemble model across the entire dataset.

After completing the training and validation phases, the final ensemble model was evaluated on an independent test dataset that was not used at any stage of model optimization. The proposed weighted average ensemble achieved a test accuracy of 0.96, demonstrating a strong ability to generalize to unseen lung CT images. This result indicates that the integration of multiple deep learning models effectively enhanced prediction robustness and reduced individual model biases. Figure 29 represents the test accuracy comparison of the individual models and ensemble model.



**Figure 29. Test accuracy comparison of the individual models and ensemble model**

### 5.3 Performance Analysis of our Proposed Model with Existing Works

Moreover, the effectiveness of our proposed approach has been contrasted with earlier studies that made use of the identical IQ-OTH/NCCD dataset displayed in Table 5. It indicates that authors used VGG16 to reach 98.18% accuracy in the paper [23], while GoogleNet was used to obtain 94.38% accuracy in another study [24]. Once again, using VGG16 and DenseNet, researchers in papers [25] and [26] achieved an accuracy of 98.83% and 53%, respectively. Overall, our

suggested approach has yielded a higher accuracy rate of 92.28% compared with other earlier relevant works. Additionally, a web application has been created that was not noticed in any other study. Also, our proposed weighted average ensemble model can detect lung cancer precisely which balances the strengths of multiple architectures while compensating for their individual limitations.

<b>Sl.</b>	<b>Dataset</b>	<b>Pre-processing Steps</b>	<b>Model Performance</b>	<b>Web Application</b>
[23]	IQ-OTH/NCCD	Yes	VGG16: Acc.: 98.18%	No
[24]	IQ-OTH/NCCD	Yes	GoogleNet: Acc.: 94.38%	No
[25]	IQ-OTH/NCCD	Yes	VGG16: Acc.: 98.83%	No
[26]	IQ-OTH/NCCD	Yes	DenseNet: Acc.: 53%	No
<b>Proposed Method</b>	<b>IQ-OTH/NCCD</b>	<b>Yes</b>	<b>Weighted Average Ensemble Methods:</b> <b>Acc.: 94%</b>	<b>Yes</b>

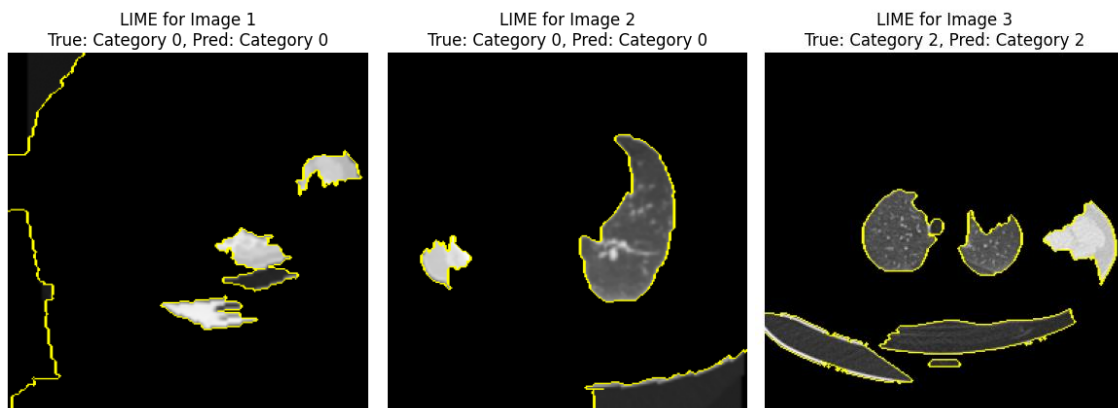
**Table 5. Performance Comparison of our proposed method and previous research on same Dataset**

## 5.4 Explainable Artificial Intelligence

It's significant to understand how a deep learning system makes decisions, particularly in applications involving medical imaging where diagnosis accuracy is vital. Following the ensemble model's training, two well-known explainability techniques LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) were used to learn more about the model's analysis of lung CT images and the regions that affect its predictions.

### 5.4.1 LIME Technique

The LIME of Figure 30 highlights the superpixels that most strongly influence the model's decision. The yellow-outlined areas in each image represent the contributing pixels in the image that increased model's confidence for the predicted class. The intensity of white regions shows higher LIME indicator proving that those regions have stronger impact on classification.



**Figure 30. Images interpreted through LIME technique**

#### **Benign Cases:**

- In Benign images, LIME mostly focuses on localized, smooth, and less dense tissue regions.
- The model relies on edges and soft nodular areas indicating awareness of benign lesion texture.

## Malignant Cases

- In Malignant prediction, LIME focused on high density, heterogenous, and irregular areas in the CT scan.
- The bright activated clusters represent rough textures which indicates malignancy.
- It finds that tissues with uneven gray intensity are a strong indicator of the development of cancer.

## Normal Cases:

- LIME activations were sparse and minimal for normal cases.
- Only minor regions have been highlighted for lung borders or noise concerns.

### ❖ Feature Attribution Across Class:

- **Benign:** LIME characteristics show low-risk configurations of symmetric, smooth, nodular, and circular forms.
- **Malignant:** Fractured patches and dense, uneven textures are highlighted when interpreting active tissue patterns and high-risk patterns.
- **Normal:** There are smooth areas that indicate the absence of lesion-related traits.

The feature attribution reveals that the ensemble model adequately differentiates among classes based on textual diversity and spatial irregularity. Each model contributes complimentary feature extraction capabilities as CNN captures fine-grained textures, VGG16 identifies spatial features, and ResNet50 encodes deep contextual representations. This synergy compliments the ensemble model showing perfect accuracy to predict lung cancer from CT scan images.

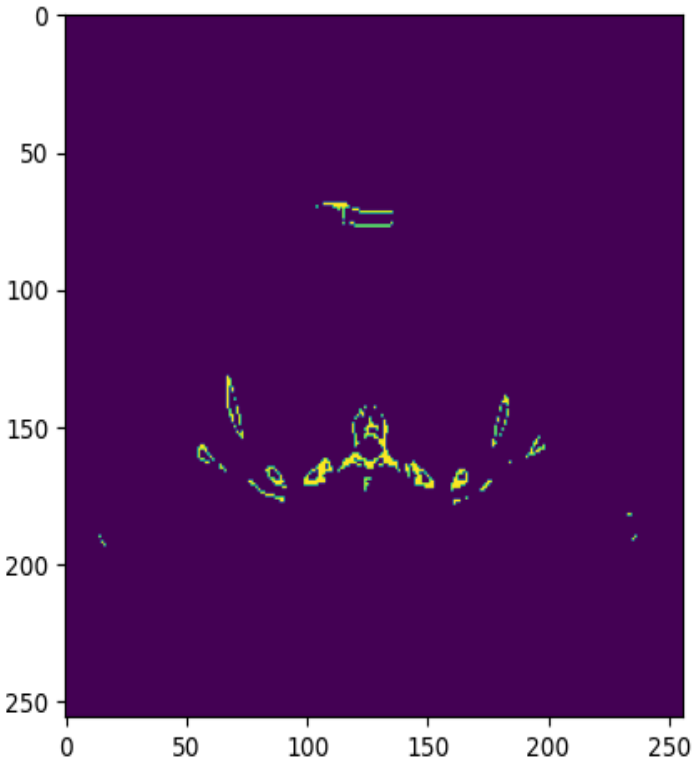
## 5.4.2 SHAP Technique

A suitable XAI method, SHAP helps to elucidate the decision-making process of the CNN model, providing insights into the specific image features that contribute most significantly to the classification predictions. For this, images have been scaled to visualize and interpret the learned features within our proposed model.

### ❖ Feature before Scaling

The SHAP values here reflect the direct influence of the raw, unprocessed feature values. Features with inherently larger magnitudes (e.g., pixel intensities or measurements in different units)

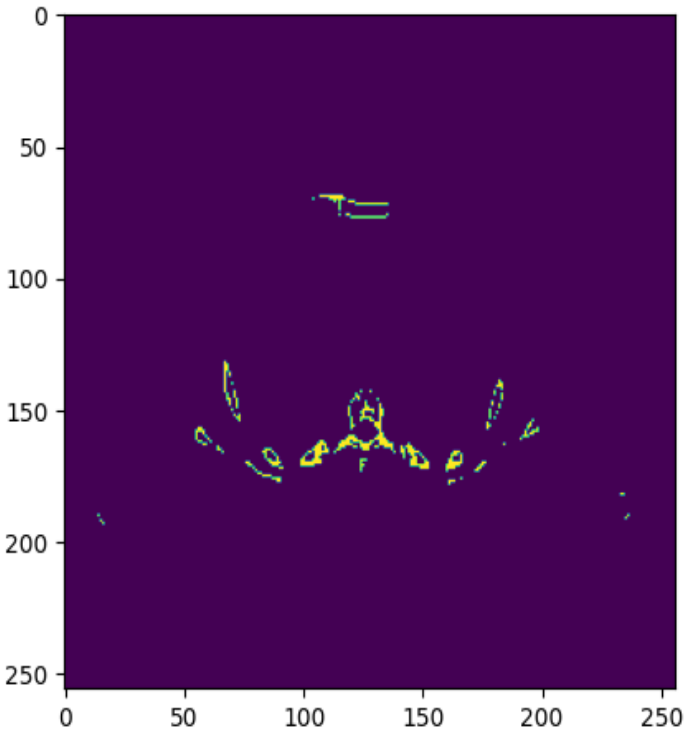
dominate the model's focus shown in Figure 31. This can lead to biased feature importance since the model may over-prioritize features with larger absolute values rather than their actual predictive power.



**Figure 31. Feature before scaling**

#### ❖ Feature after Scaling

After scaling, all features are brought to a comparable range (e.g., standardization to zero mean and unit variance). The phenomenon guarantees that each feature's SHAP value represents its true contribution to the model's decision, regardless of its original magnitude.

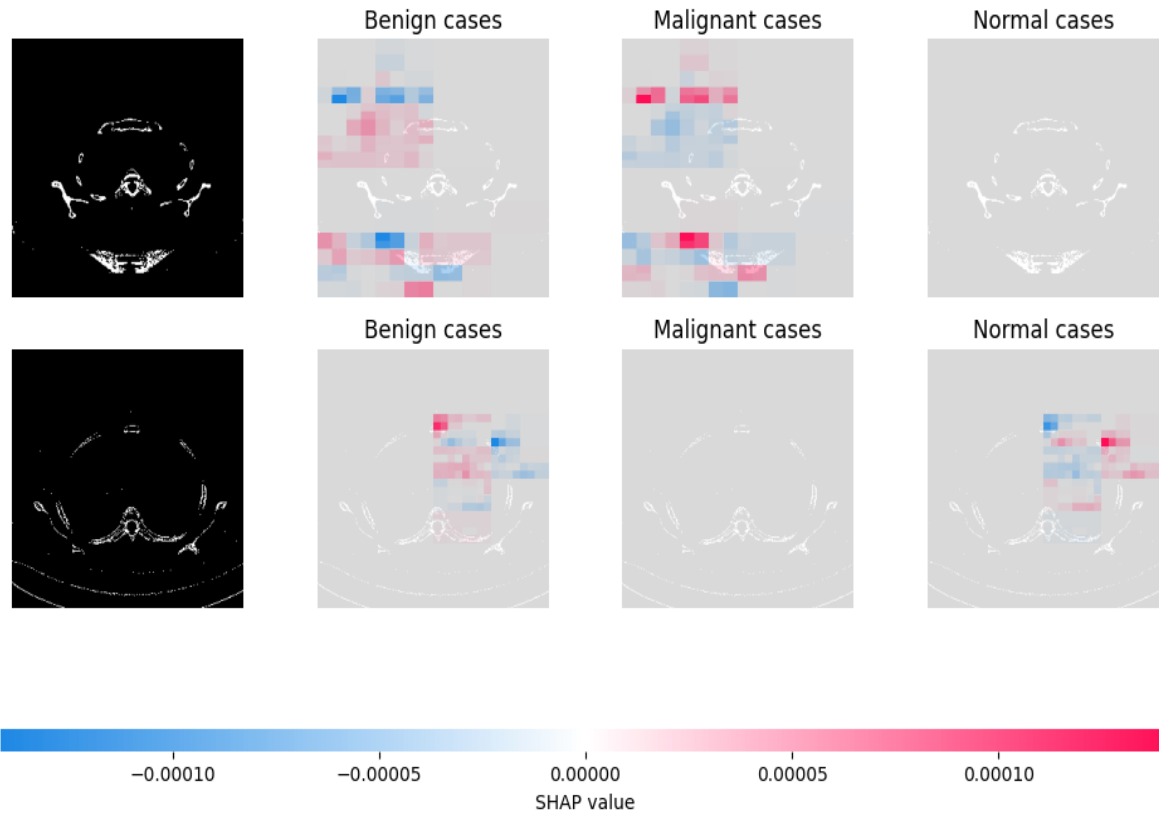


**Figure 32. Feature after scaling**

The highlighted regions in Figure 32 now better reflect the features the model actually relies on for its decisions. Because the model is no longer skewed toward characteristics with bigger sizes by nature, its significance is more precise and interpretable. This ensures that features such as specific patterns, textures, or pixel intensities crucial for lung cancer detection are equally weighted and visible in the SHAP analysis.

#### ❖ SHAP Value Analysis

SHAP values indicate the marginal contribution of each pixel to the prediction of a certain class. Positive SHAP values (red areas) indicate that the corresponding pixel intensities increase the likelihood of a specific prediction (e.g., benign, malignant, or normal). As well as negative SHAP values (blue areas) indicate that the pixel intensities decrease the likelihood of a specific prediction. The sum of SHAP values across all features (pixels) for a given input corresponds to the difference between the model's output and the average baseline prediction.



**Figure 33. Output with SHAP value**

The SHAP maps in Figure 33 reflect how specific regions of the input images contributed to the prediction for each class of lung cancer:

**Benign Cases:**

- Red zones indicate structures or traits linked to benign abnormalities. (e.g., localized growths, non-invasive patterns).
- Blue zones may be irrelevant to diagnosing benign tumors or have traits that contradict the prediction.

### **Malignant Cases:**

- Red spots indicate doubtful patterns, such as uneven growth, greater density regions, or irregular tissue structures linked to cancer.
- Blue areas indicate traits that lower confidence in predicting malignancy.

### **Normal Cases:**

- Red zones suggest healthy anatomical characteristics, supporting the prognosis of "normal."
- Blue spots may indicate noise or small abnormalities that contradict normality but were not considered important.

### **❖ Magnitude of SHAP Values:**

- The model's output rates are used to modify SHAP values, which signify how much each pixel contributes to increasing or lowering the predicted probability for a class.
- Positive SHAP values (red) around 0.00010 are associated with significant positive contributions to a class, while negative SHAP values (blue) near -0.00010 have strong negative consequences.
- This region exhibits fine-grained differences in model confidence due to local pixel-level effects.

### **❖ Feature Attribution Across Classes**

The varied patterns of SHAP values among classes indicate how the model may separate and focus on relevant data as follows:

- The model recognizes patterns of benign tissue development, such as organized, restricted growth void of invasive behavior.

- The malignant model targets areas with significant abnormalities such as uneven forms, size variations, or density changes suggest malignancy.

SHAP prohibits the model from generating predictions based on defects or irrelevant regions of the input images. If the SHAP values emphasized non-relevant locations, it would imply that the model was overfitted or biased.

## **5.5 Computational Complexity Analysis of the Proposed System**

The training complexity of the system is dominated by the deep learning models. The custom CNN involves multiple convolution and pooling layers, resulting in a computational cost proportional to the number of training samples, image dimensions, and training epochs. In contrast, VGG16 and ResNet50 leverage transfer learning, where the convolutional base layers are frozen. As a result, only the classifier layers are trained, significantly reducing training overhead. During inference, each image is independently processed by the CNN, VGG16, and ResNet50 models. The total inference cost is the sum of the individual model prediction costs. The ensemble stage combines the output probabilities using weighted averaging, which introduces negligible computational overhead compared to deep model inference. Explainability methods such as LIME and SHAP increase computational cost due to repeated model evaluations on perturbed inputs. However, these techniques are applied selectively after training and do not affect real-time system performance. Overall, the system achieves a balance between computational efficiency and predictive accuracy, making it suitable for practical medical decision-support applications.

The system was implemented and evaluated using GPU-accelerated computing to handle the high computational demand of deep neural networks. Training the ensemble models requires moderate to high GPU memory due to the large number of parameters, particularly in VGG16 and ResNet50. However, freezing the pretrained layers significantly reduces memory consumption and training time. The ensemble aggregation step consumes minimal memory and processing power. From a deployment perspective, the trained ensemble model can be integrated into a web-based application and executed on cloud or hospital-grade systems without requiring specialized hardware, ensuring scalability and accessibility.

## **5.6 Cost Analysis of the Proposed System**

In this study, Model training and evaluation were performed on cloud-based computational resources (Google Colab with GPU support), which eliminates the need for dedicated high-end hardware during development. The weighted average ensemble approach does not introduce significant additional cost, as it involves simple mathematical operations on model outputs. Explainability tools such as LIME and SHAP increase computational time but do not incur extra financial cost, since they are executed offline and on demand. Overall, the system offers a cost-effective solution for early lung cancer prediction, combining high diagnostic performance with minimal infrastructure and maintenance expenses.

## CHAPTER 6

### CONCLUSION & FUTURE WORKS

This chapter summarizes the main findings of the study and evaluates how effectively the suggested approach achieved its objectives. It highlights potential topics for further research, including suggestions for enhancing the model and expanding its application in actual clinical settings.

#### 6.1 Conclusion

One of the leading causes of death worldwide, lung cancer poses a serious threat to public health because of its relatively hidden character and rapid progression. Higher survival rates and improved patient outcomes rely on early detection and precise diagnosis. The purpose of this work was to create and evaluate an ensemble-based deep learning system for using CT scan pictures to categorize lung cancer into three categories: benign, malignant, and normal. The system integrated two pretrained architectures, VGG16 and ResNet50, with a bespoke CNN to take use of each model's complementary qualities, resulting in a more stable and reliable prediction framework. Several preprocessing and augmentation techniques were employed to improve data quality and reduce the potential of model bias while image sources diverged. The CNN demonstrated consistently stable learning behavior, while VGG16 had the best validation performance among the various models. By merging the probability outputs of all three networks, the weighted ensemble elevated predicted confidence. LIME and SHAP were used to interpret the ensemble's decision-making processes in order to guarantee transparency. The reliability of the model's predictions was supported by these explainability tools, which verified that the model concentrated on medically significant image regions during categorization. This visualization revealed that the model concentrates on key anatomical components, with unique patterns of pixel contributions found for benign, malignant, and normal instances. This level of interpretability significantly enhances the clinical utility of the model, fostering trust and confidence in its predictions among healthcare professionals.

Finally, to enhance the efficacy of our system, we developed a user-friendly online application capable of predicting lung cancer with a high degree of confidence in its early stages. This application facilitates rapid and convenient access to the predictive model for healthcare professionals. By providing a convenient and intuitive platform for utilizing the ensemble model, we aim to streamline the lung cancer prediction process and improve the efficiency of clinical workflows.

## **6.2 Future Scope**

Future research endeavors should prioritize enhancing the system's accessibility and practical applicability. A crucial area of focus lies in optimizing system implementation within resource-constrained environments, thereby expanding its reach and facilitating broader adoption. Despite the positive outcomes of the current effort, there are still a number of areas that need more investigation. Adding pictures from more medical facilities to the dataset is a crucial step that would improve the model's generalizability and lessen domain shifts. Additionally, using 3D CT volumes instead of discrete 2D slices may enhance lesion characterization and offer greater spatial information. To improve the interaction between component models, future research may try more sophisticated ensemble techniques as attention-based fusion or stacking with meta-learners. Fine-tuning the pretrained models rather than deploying them in a frozen state is another interesting approach, particularly when larger and more varied datasets become available. Integrating the system into a decision-support workflow, followed by clinical validation with radiologists, would mark a significant step toward practical adoption.

Furthermore, utilization of the state-of-the-art deep learning architectures, including generative adversarial networks (GANs) or transformer-based models, might result in notable improvements in the prediction of lung cancer. These advanced models have the potential to learn more complex representations from medical images, leading to improved generalization and enhanced diagnostic performance. Finally, optimizing the system's sustainability and processing efficiency is crucial for its successful integration into real-world healthcare settings.

## References

- [1] D. Singh and A. Taneja, "Assessment of risk factors of lung cancer in non-smokers in india," in 2022 7th International Conference on Intelligent Informatics and Biomedical Science (ICIIBMS), vol. 7. IEEE, 2022, pp. 7–12.
- [2] N. Horeweg, J. van Rosmalen, M.A. Heuvelmans, C.M. Aalst, R. Vliegenthart, E.Th Scholten, et al., Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening, *Lancet Oncol.* 15 (2014) 1332–1341, [https://doi.org/S1470-2045\(14\)70389-4](https://doi.org/S1470-2045(14)70389-4) [pii]
- [3] B.C. Bade, C.S.D. Cruz, Lung cancer 2020: epidemiology, etiology, and prevention, *Clin. Chest Med.* 41 (1) (2020) 1–24.
- [4] A.S. Ahmad, A.M. Mayya, A new tool to predict lung cancer based on risk factors, *Heliyon* 6 (2) (2020) e03402.
- [5] S. Kummer, J. Waller, M. Ruparel, J. Cass, S. M. Janes, S. L. Quaipe, Mapping the spectrum of psychological and behavioural responses to low-dose CT lung cancer screening offered within a Lung Health Check. *Health Expect* Apr. 2020;23(2):433–41. <https://doi.org/10.1111/HEX.13030>.
- [6] K. Punithavathy, M. Ramya, and S. Poobal, "Analysis of statistical texture features for automatic lung cancer detection in pet/ct images," in 2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE). IEEE, 2015, pp. 1–5.
- [7] L. Evangelista, M. Sepulcri, and G. Pasello, "Pet/ct and the response to immunotherapy in lung cancer," *Current Radiopharmaceuticals*, vol. 13, no. 3, pp. 177–184, 2020.
- [8] A. R. Neravetla, S. Samreen, A. S. Mohammed, N. Jiwani, and J. Logeshwaran, "An improved lung cancer prediction algorithm using generative adversarial network in modern healthcare," in 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS). IEEE, 2024, pp. 1–6.
- [9] M. Mamun, S. B. Shawkat, M. S. Ahammed, M. M. Uddin, M. I. Mahmud, and A. M. Islam, "Deep learning based model for alzheimer's disease detection using brain mri images," in 2022

IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2022, pp. 0510–0516.

[10] Z. Ren, Y. Zhang, S. Wang, LCDAE: data augmented ensemble framework for lung cancer classification. *Technol Cancer Res Treat.* 2022;21: doi:10.1177/15330338221124372.

[11] Á. D. Tárnoki, D. L. Tárnoki, M. Dąbrowska, M. Knetki-Wróblewska, A. Frille, H. Stubbs, K. G. Blyth, A. D. Juul. New developments in the imaging of lung cancer. *Breathe (Sheff).* 2024 Mar;20(1):230176. doi: 10.1183/20734735.0176-2023. Epub 2024 Apr 9. PMID: 38595936; PMCID: PMC11003524.

[12] M. Bicakci, O. Ayyildiz, Z. Aydin et al (2020) Metabolic imaging based sub-classification of lung cancer. *IEEE Access* 8:218470–218476. <https://doi.org/10.1109/ACCESS.2020.3040155>

[13] Wang X, Chen H, Gan C et al (2020) Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE TransCybern* 50(9):3950–3962. <https://doi.org/10.1109/TCYB.2019.2935141>

[14] S. Hussain, I. Mubeen, N. Ullah, S. S. U.D. Shah, B. A. Khan, M. Zahoor, R. Ullah, F. A. Khan, M. A. Sultan, Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review. *Biomed Res Int.* 2022 Jun 6;2022:5164970. doi: 10.1155/2022/5164970. PMID: 35707373; PMCID: PMC9192206.

[15] R. Javed, T. Abbas, A. H. Khan A. Daud, A. Bukhari, R. Alharbey, 2024, Deep learning for lungs cancer detection: a review. *Artif Intell Rev* 57, 197. <https://doi.org/10.1007/s10462-024-10807-1>

[16] W. D. Wever, J. Coolen, J. A. Verschakelen, Imaging techniques in lung cancer. *Breathe* 2011; 7: 338–346. doi: 10.1183/20734735.022110

[17] J. H. Ong, K. M. Goh, L. L. Lim, Comparative Analysis of Explainable Artificial Intelligence for COVID-19 Diagnosis on CXR Image. In: 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA); 2021; Kuala Terengganu, Malaysia. p. 185–90. doi:10.1109/ICSIPA52582.2021.9576766.

- [18] C. Kaur, U. Garg, Artificial intelligence techniques for cancer detection in medical image processing: A review, *Materials Today: Proceedings*, Volume 81, Part 2, 2023, Pages 806-809, ISSN 2214-7853, <https://doi.org/10.1016/j.matpr.2021.04.241>.
- [19] S. Ali, T. Abuhmed, S. E. Sappagh, K. Muhammad, J. M. A. Moral, R. Confalonieri, R. Guidotti et. al, Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence, *Information Fusion*, Volume 99, 2023, 101805, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2023.101805>.
- [20] A. Singhal, K. K. Agrawal, A. Quezada, A. R. Aguiñaga, S. Jiménez, S. P. Yadav, Explainable Artificial Intelligence (XAI) Model for Cancer Image Classification, *Computer Modeling in Engineering & Sciences*, Volume 141, 2024, 1, ISSN 1526-1506, pg 401- 441.
- [21] J.H.-w. Hsiao, H.H.T. Ngai, L. Qiu, Y. Yang, C.C. Cao, Roadmap of designing cognitive metrics for Explainable Artificial Intelligence (XAI), 2021, arXiv preprint arXiv:2108.01737.
- [22] T. P. Hanna, W. D. King, S. Thibodeau, M. Jalink, G. A. Paulin, E. Harvey-Jones, D. E. O'Sullivan, C. M. Booth, R. Sullivan, A. Aggarwal, Mortality due to cancer treatment delay: systematic review and meta-analysis. *BMJ*. 2020 Nov 4;371:m4087. doi: 10.1136/bmj.m4087. PMID: 33148535; PMCID: PMC7610021.
- [23] Y. Kumaran S, J. J. Jeya, S. B. Khan, S. Alzahrani, and M. Alojail, "Explainable lung cancer classification with ensemble transfer learning of vgg16, resnet50 and inceptionv3 using grad-cam," *BMC medical imaging*, vol. 24, no. 1, p. 176, 2024
- [24] M. Humayun, R. Sujatha, S. N. Almuayqil, and N. Jhanjhi, "A transfer learning approach with a convolutional neural network for the classification of lung carcinoma," in *Healthcare*, vol. 10, no. 6. MDPI, 2022, p. 1058.
- [25] M. S. Al-Huseiny and A. S. Sajit, "Transfer learning with googlenet for detection of lung cancer," *Indonesian Journal of Electrical Engineering and computer science*, vol. 22, no. 2, pp. 1078–1086, 2021.
- [26] A. Mohite, "Lung cancer diagnosis using transfer learning," *Int. J. Sci. Res. Manag*, vol. 9, no. 11, pp. 621–634, 2021.

- [27] N. Gautam, A. Basu, R. Sarkar, Lung cancer detection from thoracic CT scans using an ensemble of deep learning models. *Neural Comput & Applic* 36, 2459–2477 (2024). <https://doi.org/10.1007/s00521-023-09130-7>
- [28] B. Madan, A. Panchal, D. Chavan, Lung Cancer Detection Using Deep Learning. In *Proceedings of the 2nd International Conference on Advances in Science and Technology (ICAST-2019)*, Makassar, Indonesian, 5–6 November 2019; p. 3.
- [29] M. Mamun, M. I. Mahmud, M. Meherin, and A. Abdelgawad, “Lcdctcnn: Lung cancer diagnosis of ct scan images using cnn based model,” in *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2023, pp. 205–212
- [30] M. Sohaib, M. Adewunm. Artificial intelligence based prediction on lung cancer risk factors using deep learning. *Int J Inf Commun Technol Apr.* 2023;12(2):188–94. <https://doi.org/10.11591/ijict.v12i2.pp188-194>
- [31] Q. Zhang, X. Kong, 2020 Design of Automatic Lung Nodule Detection System Based on Multi-Scene Deep Learning Framework. *IEEE Access*
- [32] M. Toğaçar, B. Ergen, Z. Cömert, 2020, Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks *Biocybernetics and Biomedical Engineering* 40(1) 23-39
- [33] A. Masood, P. Yang, B. Sheng et al (2020) Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT. *IEEE J Transl Eng Health Med* 8:1–13. <https://doi.org/10.1109/JTEHM.2019.2955458>
- [34] A. Asuntha, A. Srinivasan, Deep learning for lung Cancer detection and classification. *Multimed Tool Appl* Mar. 2020;79(11–12):7731–62.

- [35] S. Pang, Y. Zhang, M. Ding (2020) A deep model for lung cancer type identification by densely connected convolutional networks and adaptive boosting. *IEEE Access* 8:4799–4805. <https://doi.org/10.1109/ACCESS.2019.2962862>
- [36] S. K. Lakshmanaprabu, S. N. Mohant, K. Shankar et al (2019) Optimal deep learning model for classification of lung cancer on CT images. *Future Gener Comput Syst* 92:374–382. <https://doi.org/10.1016/j.future.2018.10.009>
- [37] M. R. Hassan, M. F. Islam, M. Z. Uddin, G. Ghoshal, M. M. Hassan, S. Huda et al. Prostate cancer classification from ultrasound and MRI images using deep learning based explainable artificial intelligence. *Future Gener Comput Syst.* 2022;127(1):462–72. doi: 10.1016/j.future.2021.09.030.
- [38] C. Dindorf, J. Konradi, C. Wolf, B. Taetz, G. Bleser, J. Huthwelker et al. Classification and automated interpretation of spinal posture data using a pathology-independent classifier and explainable artificial intelligence (XAI). *Sensors.* 2021;21(18):6323. doi:10.3390/s21186323.
- [39] M. Veerappa, M. Anneken, N. Burkart, M. F. Huber. Validation of XAI explanations for multivariate time series classification in the maritime domain, *journal of computational. Science* 2022;58:101539. <http://dx.doi.org/10.1016/j.jocs.2021.101539>.
- [40] M. S. Ahmed, K. N. Iqbal and M. G. R. Alam, "Interpretable Lung Cancer Detection using Explainable AI Methods," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-6, doi: 10.1109/ICONAT57137.2023.10080480.
- [41] C. Meske, E. Bunde, Transparency and trust in human-AI-interaction: The role of model-agnostic explanations in computer vision-based decision support, in: *International Conference on Human-Computer Interaction*, Springer, 2020, pp. 54–69.
- [42] C. Chen, O. Li, A. Barnett, J.K. Su, C. Rudin, This looks like that: deep learning for interpretable image recognition, in: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019, pp. 1–12.
- [43] V. Rajasekar, M.P. Vaishnave, S. Premkumar, V. Sarveshwaran, V. Rangaraaj, Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning

techniques, Results in Engineering, Volume 18, 2023, 101111, ISSN 2590-1230, <https://doi.org/10.1016/j.rineng.2023.101111>.

[44] R. I. Bendjillali, M. Beladgham, K. Merit, A. T. Ahmed. Illumination-robust face recognition based on deep convolutional neural networks architectures, Indonesian Journal of Electrical Engineering and Computer Science Vol.18, No.2, May 2020, pp. 1015~1027 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i2.pp1015-1027

[45] M. S. Hasan, G. Rabbi, R. Islam, M. H. I. Bijoy, M. A. Hakim, 2022, Bangla Font Recognition using Transfer Learning Method. 57-62. 10.1109/ICICT54344.2022.9850765.

[46] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, 2021, Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 23(1), 18. <https://doi.org/10.3390/e23010018>