

Article

AI-Powered Lung Cancer Detection: Assessing VGG16 and CNN Architectures for CT Scan Image Classification

Rapeepat Klangbunrueang ¹, Pongsathon Pookduang ¹, Wirapong Chansanam ^{1,*} and Tassanee Lunrasri ^{2,*}

¹ Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen 40002, Thailand; rapeepat.klangbunrueang@kkumail.com (R.K.); pongsathon.po@kkumail.com (P.P.)

² Department of Information Systems, Faculty of Business Administration and Information Technology, Rajamangala University of Technology Isan, Khon Kaen Campus, Khon Kaen 40000, Thailand

* Correspondence: wirach@kku.ac.th (W.C.); tassanee.so@rmuti.ac.th (T.L.)

Abstract: Lung cancer is a leading cause of mortality worldwide, and early detection is crucial in improving treatment outcomes and reducing death rates. However, diagnosing medical images, such as Computed Tomography scans (CT scans), is complex and requires a high level of expertise. This study focuses on developing and evaluating the performance of Convolutional Neural Network (CNN) models, specifically the Visual Geometry Group 16 (VGG16) architecture, to classify lung cancer CT scan images into three categories: Normal, Benign, and Malignant. The dataset used consists of 1097 CT images from 110 patients, categorized according to these severity levels. The research methodology began with data collection and preparation, followed by training and testing the VGG16 model and comparing its performance with other CNN architectures, including Residual Network with 50 layers (ResNet50), Inception Version 3 (InceptionV3), and Mobile Neural Network Version 2 (MobileNetV2). The experimental results indicate that VGG16 achieved the highest classification performance, with a Test Accuracy of 98.18%, surpassing the other models. This accuracy highlights VGG16's strong potential as a supportive diagnostic tool in medical imaging. However, a limitation of this study is the dataset size, which may reduce model accuracy when applied to new data. Future studies should consider increasing the dataset size, using Data Augmentation techniques, fine-tuning model parameters, and employing advanced models such as 3D CNN or Vision Transformers. Additionally, incorporating Gradient-weighted Class Activation Mapping (Grad-CAM) to interpret model decisions would enhance transparency and reliability. This study confirms the potential of CNNs, particularly VGG16, for classifying lung cancer CT images and provides a foundation for further development in medical applications.



Received: 23 November 2024

Revised: 4 February 2025

Accepted: 6 February 2025

Published: 11 February 2025

Citation: Klangbunrueang, R.; Pookduang, P.; Chansanam, W.; Lunrasri, T. AI-Powered Lung Cancer Detection: Assessing VGG16 and CNN Architectures for CT Scan Image Classification. *Informatics* **2025**, *12*, 18. <https://doi.org/10.3390/informatics12010018>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: lung cancer image classification; CT scans; convolutional neural network (CNN); VGG16 architecture; AI-assisted diagnosis; deep learning

1. Introduction

Lung cancer, characterized by high mortality rates, is a persistent global health issue that requires urgent and innovative solutions. It is responsible for more deaths than breast, colorectal, and prostate cancer combined, making it the most lethal cancer type globally [1]. Several factors, including exposure to tobacco smoke, environmental pollutants, occupational hazards, and genetic predispositions, contribute to its high incidence [2]. Socioeconomic disparities, inadequate access to healthcare, and limited early detection

programs further worsen the scenario, resulting in a high proportion of cases diagnosed at advanced stages when treatment options are limited and less effective [3].

The primary challenge in lung cancer management is early detection, as the disease often presents non-specific symptoms, which makes it difficult to diagnose until it has progressed significantly. Traditional diagnostic methods, such as Computed Tomography scans (CT scans) and biopsies, are time-consuming, prone to human error, and require considerable expertise [4]. These limitations necessitate the development of more efficient diagnostic techniques that can accurately identify lung cancer at an early stage, thus enabling timely treatment and improving survival rates.

Recent advancements in artificial intelligence (AI) and machine learning (ML) have shown promising potential to address these challenges. AI, particularly deep learning models like Convolutional Neural Networks (CNNs), offers significant advantages in medical image analysis. CNNs can automatically learn relevant features from imaging data, improving accuracy and consistency in cancer detection [5]. Among various CNN architectures, Visual Geometry Group 16 (VGG16) has emerged as a leading model due to its deep layers and feature extraction capabilities, making it highly effective for lung cancer classification from both histopathology and CT scan images [6]. VGG16 utilizes a structured network of layers that enables it to capture complex features of lung tissues, differentiating between benign, normal, and malignant lesions with high precision [7].

This study aims to explore the potential of a pre-trained VGG16 model to detect lung cancer through CT scan image analysis using an openly accessible dataset for training and evaluation. The model is trained to classify CT scan images into three categories: normal, benign lung abnormalities, and malignant lung cancer, using a dataset of images labeled by oncologists and radiologists [8]. The architecture of VGG16 is specifically designed to handle this classification task, using multiple convolutional layers, pooling layers, and fully connected layers to extract and analyze key features indicating the presence of cancer cells [9].

Figure 1 shows a lung CT scan image that reveals distinct findings for each case. Image A, representing a Normal case, shows a nodule or mass localized in the right lung, indicating a potential abnormality despite its classification as normal. Image B, classified as Benign, depicts clear lung fields with well-defined pulmonary vessels, consistent with non-pathological findings. Image C, labeled Malignant, exhibits interstitial changes or potential infiltrates in the lower lung fields, suggesting significant pathological alterations. These observations highlight the variability in CT scan presentations across different classifications, emphasizing the importance of accurate feature interpretation for reliable diagnosis.

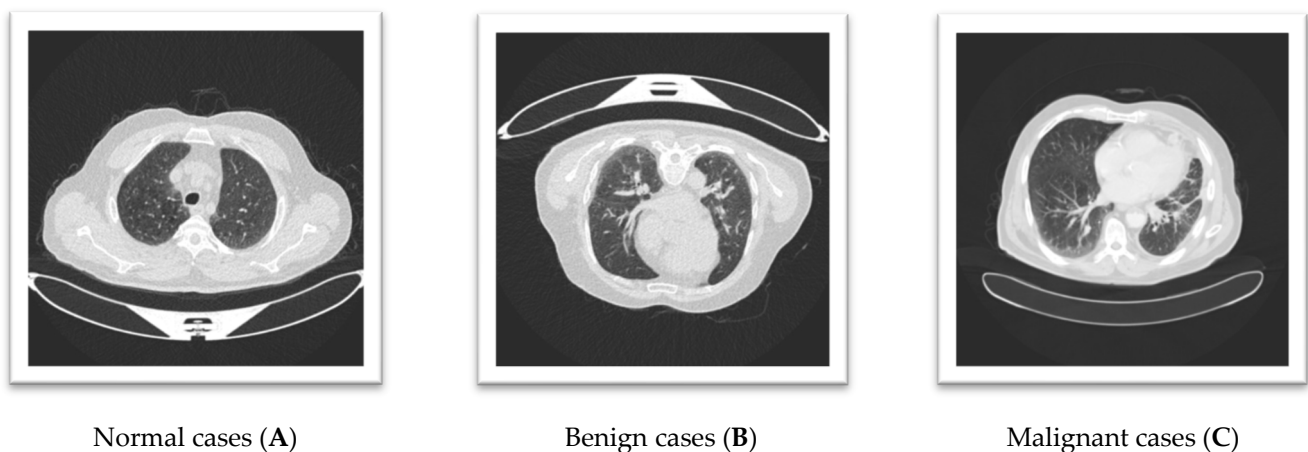


Figure 1. The three lung CT scan images.

By integrating VGG16-based CNNs into the diagnostic process, this study aims to create an automated, efficient, and reliable diagnostic tool for early lung cancer detection. The approach aims to enhance diagnostic accuracy and reduce the time and effort required for analysis, ultimately facilitating faster and more effective patient management [10]. Furthermore, the study will assess the model's robustness through testing against adversarial attacks and various imaging conditions, ensuring its reliability in real-world clinical settings [11].

Lung cancer is a leading cause of mortality worldwide, necessitating early and accurate diagnostic methods to improve survival rates. While advancements in AI and deep learning have significantly enhanced medical image analysis, the choice of architecture plays a crucial role in determining the efficacy of these systems. Among the widely used models, CNNs like Residual Network with 50 layers (ResNet50), Mobile Neural Network Version 2 (MobileNetV2), and InceptionV3 have demonstrated varying levels of success in medical imaging tasks. ResNet50's residual connections improve learning in deeper networks, making it suitable for complex tasks but computationally intensive. MobileNetV2 offers lightweight and efficient performance, ideal for resource-constrained environments, but often sacrifices accuracy in high-resolution image tasks. InceptionV3 excels in multiscale feature extraction but can struggle with subtle distinctions in medical images due to kernel limitations.

VGG16, on the other hand, is a simpler yet powerful architecture known for its ability to extract hierarchical features. Its depth and consistent performance across diverse image classification tasks make it a reliable choice, particularly in scenarios where interpretability is essential. While newer architectures like Vision Transformers and 3D CNNs offer promising alternatives, their higher computational demands and complexity can limit their immediate applicability in clinical workflows. This study builds on the strengths of VGG16, exploring its application in a challenging three-class classification problem (Normal, Benign, Malignant), while comparing its performance against alternative architectures to establish its clinical utility.

Therefore, the origin of the problem—late-stage lung cancer diagnosis—demonstrates the importance of developing AI-driven solutions. This study aims to build an advanced VGG16-based model that accurately classifies lung cancer images and evaluates its performance through accuracy, sensitivity, and specificity metrics. The research seeks to significantly improve early detection rates, thereby reducing lung cancer mortality and enhancing patient outcomes globally.

The paper is organized as follows; Section 2 presents the related work. The material and methods are explained in Section 3. Results and discussion are discussed in Section 4. Finally, the conclusion is presented.

2. Related Works

Recent advancements in lung cancer detection have increasingly leveraged AI, particularly CNNs, to improve diagnostic accuracy, efficiency, and early detection rates. This section reviews pre-trained models, including VGG16 and other CNN architectures, in detecting and classifying lung cancer using histopathological and CT images.

Al-Shouka and Alheeti employed transfer learning with VGG16, achieving high accuracy in detecting malignant tissues in CT images. Their approach demonstrated the efficiency of transfer learning in reducing computational complexity and enhancing reliability [12].

Bherje et al. proposed a deep learning framework for predicting lung cancer using CT scan images, achieving an average accuracy of 72.41%. Their study demonstrated the potential of deploying CNN-based systems in practical settings for early cancer detection [13].

Kapoor et al. explored the effectiveness of the pre-trained VGG16 model in detecting lung cancer from histopathological images. Their study utilized publicly available datasets to train and assess the model's performance, demonstrating that VGG16 could accurately classify lung tissue into benign, normal, or malignant categories. This study confirmed the model's potential for reliable lung cancer detection, emphasizing its ability to handle complex histopathological data and enhance diagnostic precision [2].

Huang et al. applied the VGG16 model to classify CT images of lungs into three categories: benign, normal, and malignant. The study highlighted VGG16's robustness, evaluating it against adversarial attacks to ensure stability and accuracy. Their findings reinforced VGG16's suitability for handling CT scans in clinical settings, demonstrating its ability to differentiate between various lung conditions effectively [6].

Xu compared traditional CNN models with more advanced architectures, including VGG16, for non-small cell lung cancer classification. While basic CNNs achieved acceptable accuracy, VGG16 outperformed them regarding learning efficiency and overall accuracy, proving that deeper networks yield better diagnostic results when applied to lung cancer datasets. This study supports using pre-trained CNNs like VGG16 in lung cancer diagnosis, emphasizing their superior performance and feature extraction capabilities [3].

Tejaswini et al. analyzed different CNN architectures for lung cancer detection, focusing on feature extraction and machine learning algorithms. Their research demonstrated that CNNs, including VGG16, significantly improve diagnostic accuracy compared to conventional methods. The study concluded that VGG16's deep architecture allows for a more precise analysis of lung tissues, making it a suitable choice for early lung cancer detection [4].

A broader comparison by V et al. assessed various 3D CNN models, such as AlexNet, CNN-T5, and VGG16, in classifying lung cancer using CT and PET scans. VGG16 emerged as a highly accurate model due to its structured architecture and ability to adapt through transfer learning, further establishing its applicability in lung cancer diagnosis. The study also emphasized optimizing hyperparameters like image and batch sizes to enhance model performance [14].

Karthikeyan et al. applied deep learning to classify CT scan images of lungs into benign and malignant categories. Their VGG16-based model achieved significant accuracy, underscoring its capability to reduce diagnostic delays and improve early detection [15].

Aharonu and Kumar developed a CNN-based framework for automatic lung cancer detection from CT images, reporting a high accuracy rate of 94.11%. Their study concluded that CNN models, including VGG16, outperform traditional methods in identifying cancerous tissues, offering efficient solutions for automated diagnostics [5]. Similarly, Zargar et al. utilized VGG16 to classify carcinoma nodules in CT scans, achieving high sensitivity (92.08%), accuracy (91%), and an area under the curve (AUC) of 93%. These findings confirm VGG16's practical use in clinical diagnostics [7].

While many studies in the state-of-the-art section, such as those by Tejaswini et al. [4] and Aharonu and Kumar [5], emphasize the use of CNNs for lung cancer detection, they often lack detailed quantitative results, and the metrics used to evaluate model performance. This gap makes it challenging to compare their findings with newer studies that employ comprehensive metrics like accuracy, sensitivity, specificity, and F1-score. This study adopted a systematic approach to evaluate the performance of different CNN architectures, including VGG16, using a full set of metrics to provide a robust and transparent comparison. By presenting detailed quantitative outcomes, this work addresses the shortcomings in prior research and contributes a more comprehensive analysis to the field.

Vaishnavi et al. introduced an automatic lung cancer detection method using CNNs to classify tissues into normal, benign, and malignant categories. Their approach demon-

strated CNNs' potential to enhance diagnostic efficiency, reduce time, and improve accuracy [10]. Kumaran et al. expanded on this by presenting an ensemble approach combining VGG16, ResNet50, and InceptionV3, achieving improved diagnostic accuracy and robustness in lung cancer classification. Gradient-weighted Class Activation Mapping (Grad-CAM) was employed for model interpretability, aiding clinical applications [9].

Gayap and Akhloufi reviewed deep machine learning applications for lung cancer detection, emphasizing VGG16's robust performance in medical imaging and adaptability to various diagnostic tasks [16].

Kumaran et al. developed an ensemble approach integrating VGG16, ResNet50, and InceptionV3 to enhance diagnostic accuracy for lung cancer detection. By combining the strengths of these pre-trained models, their method achieved robust performance in classifying lung cancer. The study also employed Grad-CAM for model interpretability, allowing clinicians to visualize the decision-making process of the AI system. This interpretability feature made the AI results more accessible and trustworthy for clinical use, addressing a critical need for transparency in automated diagnostics [9].

Mamun et al. introduced Low-Cost Device using Convolutional Neural Network (LCDtCNN), a CNN-based system showcasing VGG16's efficiency in analyzing CT scan images. Their findings reinforced the reliability of CNNs in automated cancer detection [17]. Similarly, Mane et al. demonstrated VGG16's effectiveness in early lung cancer detection, emphasizing its ability to reduce diagnostic delays and improve outcomes [18].

Saravanan et al. proposed a CNN-based approach for detecting and classifying lung tumors from CT scans. Their method, incorporating dynamic parameter pooling and Rectified Linear Unit (ReLU), demonstrated VGG16's high accuracy and reliability for lung cancer detection [19].

Al-Areqi et al. enhanced lung cancer classification by integrating Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing with CNN-based models, improving VGG16's performance in feature extraction and image classification [20].

Jain et al. validated VGG16's sensitivity and specificity in processing CT scan data, demonstrating its potential for real-world clinical applications [21].

Metagar and Sayyed confirmed VGG16's reliability in distinguishing benign from malignant tumors, highlighting its role in automated diagnostic tools [22].

Mitra et al. emphasized VGG16's robustness in handling diverse datasets, supporting its effectiveness in enhancing diagnostic accuracy [23].

Nandini et al. proposed a CNN-based model for differentiating benign and malignant lung tumors, showcasing VGG16's high precision and adaptability to varied diagnostic scenarios [24].

Singh et al. utilized a transfer learning-based deep learning approach for automated lung cancer detection, leveraging pre-trained CNN models. Their study identified VGG16 as a superior architecture for detecting malignant tissues, outperforming other CNN models in terms of accuracy and feature extraction capabilities. This highlights VGG16's robustness and effectiveness in handling complex medical imaging tasks, making it a valuable tool for lung cancer diagnosis [25]. In summary, previous research underscores the effectiveness of VGG16 in lung cancer detection. Its deep architecture, robust feature extraction capabilities, and adaptability to diverse datasets make it a preferred choice for early diagnosis. These studies align with this research's objective to utilize VGG16 for automated lung cancer analysis, aiming to enhance accuracy, sensitivity, and specificity in clinical applications.

3. Methodology

This study adopts a systematic framework to classify lung cancer CT scan images using convolutional neural networks (CNNs), focusing on the VGG16 architecture. The methodology is outlined across distinct stages, ensuring clarity and reproducibility:

3.1. Data Collection and Preprocessing

Dataset Description:

The dataset used in this study comprises 1097 CT scan images sourced from a publicly available medical imaging database tailored for lung cancer research. The images, derived from 110 patients, are categorized into three classes based on radiological and oncological evaluations [8].

- Normal: Images with no signs of abnormalities (300 images).
- Benign: Images with non-malignant abnormalities (347 images).
- Malignant: Images with confirmed cancerous lesions (450 images).
- Key Characteristics:
 - Dimensionality: Two-dimensional slices captured at varying angles, ensuring diverse perspectives.
 - Resolution: Original resolutions were adjusted to 224×224 pixels to meet the input requirements of the VGG16 model.
 - Labeling: Each image was labeled by radiologists and oncologists to ensure accuracy and reliability.

Data Splitting and Test Data Selection:

The dataset was divided into training (70%), validation (20%), and testing (10%) subsets using stratified sampling to maintain balanced class distributions. The test set, unseen during training and validation, was randomly selected within each stratified category to provide an unbiased evaluation of the model's generalizability. Performance metrics such as accuracy, sensitivity, specificity, and F1-score were computed for each class within the test set to validate the model further.

Limitations:

- The dataset size is relatively small, which may affect model generalizability.
- The images are 2D slices, which could limit the depth-related feature extraction compared to 3D datasets.

Preprocessing Steps:

- Image Resizing: All images were resized to 224×224 pixels, balancing computational efficiency with the preservation of critical diagnostic features.
- Normalization: Pixel values were scaled between 0 and 1 to standardize inputs, reduce numerical instability, and enhance training convergence.
- Data Augmentation: To mitigate overfitting and improve model generalizability, data augmentation techniques were applied:
 - Rotation: Random rotations up to 30° to simulate variations in orientation.
 - Scaling: Adjustments ranging from 90% to 110% to mimic imaging differences.
 - Flipping: Horizontal and vertical flipping to reflect anatomical variations.
- Impact of Data Augmentation: Data augmentation increased the dataset's diversity and improved the model's classification accuracy by 2–3% during validation and testing phases. This enhancement was particularly evident in challenging cases, such as distinguishing between borderline Normal and Benign categories.

Data Preprocessing:

- Image Resizing and Normalization

- All images are resized to 224×224 pixels.
- Normalization:

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)},$$

- where

- x_{norm} : Normalized pixel value,
- Scales pixel values to be between 0 and 1 for consistent input.

- Data Augmentation

- Techniques such as rotation, scaling, and flipping are applied to increase data diversity.
- Rotation:

$$T(x, y) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix},$$

- where

- θ : Rotation angle,
- $T(x, y)$: New coordinates after rotation,
- Helps the model generalize better by introducing variations in the dataset.

The research hypothesis—that the VGG16 model outperforms other architectures in classifying lung CT images into three categories—was tested using a consistent evaluation framework. Metrics such as accuracy, sensitivity, specificity, and F1-score were calculated on the test set to validate model performance. These detailed processes ensure the study's findings are reproducible and its hypothesis is testable under similar experimental conditions.

3.2. Model Development

Model Selection: The VGG16 architecture was chosen for its proven performance in feature extraction and classification in medical imaging. The model was initialized with pre-trained weights from the ImageNet dataset and fine-tuned using the lung CT dataset.

Transfer Learning: Initial weights from the ImageNet dataset were used, followed by fine-tuning with the lung CT dataset.

3.2.1. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are powerful deep learning models specifically designed to excel in analyzing visual data by systematically extracting and learning features from images. Their remarkable ability to identify and process intricate patterns makes them a cornerstone of medical image analysis, including critical applications such as lung cancer detection.

At the core of CNN functionality lies a series of operations that enable the model to interpret and process visual information effectively. One of the most essential components is the convolutional layers, which apply filters to the input image. These filters work by scanning small regions of the image, detecting local patterns such as edges, textures, and other features crucial for understanding complex visual structures.

The convolution process mathematically represents a systematic operation where the filter slides over the input image, performing element-wise multiplications and summing the results. This operation results in feature maps that highlight significant attributes of the input data, enabling the model to focus on relevant aspects while reducing noise or less critical information. Through these structured layers and processes, CNNs are

uniquely suited to handle the complexity of medical imaging tasks, offering precision and reliability [26].

Mathematical Representation of Convolution:

$$(f * I)(x, y) = \sum_{i=-k}^k \sum_{j=-k}^k f(i, j) \cdot I(x - i, y - j),$$

- where
 - f : Filter (or kernel),
 - I : Input image,
 - x, y : Pixel coordinates,
 - k : Filter size,
 - The output represents the sum of the element-wise multiplication of the filter and a specific region of the image.
- ReLU Activation Function: Used to introduce non-linearity into the model.

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

- where
 - Returns 0 for negative values and the value itself for positive values, allowing the model to learn complex patterns.
 - Pooling Layers: These layers reduce the dimensionality of the feat
 - Max-Pooling:

$$P(x, y) = \max_{i,j} \{I(x + i, y + j)\},$$

- where
 - $P(x, y)$: Pooled output at location (x, y) ,
 - Returns the maximum value from the region, effectively down-sampling the input.

3.2.2. Visual Geometry Group 16 Architecture

Visual Geometry Group 16 (VGG16) is a well-known CNN architecture featuring 16 layers, primarily designed for image classification tasks. It has gained widespread adoption in medical imaging applications due to its ability to extract hierarchical features and deliver reliable performance.

The architecture utilizes multiple convolutional layers with small 3×3 filters to analyze input images. These filters progressively capture features, ranging from simple patterns like edges to more complex structures, allowing for effective representation of medical image data. Each convolutional operation is followed by the ReLU activation function, which introduces non-linearity. This enhances the model's capacity to learn and represent intricate relationships within the data.

The features extracted by the convolutional layers are further processed through fully connected layers, which aggregate the information and interpret it for classification tasks. For instance, in medical imaging, VGG16 can classify images into categories such as normal, benign, or malignant. Finally, the SoftMax layer outputs a probability distribution, assigning a likelihood to each category, enabling interpretable and accurate predictions.

This systematic and structured design makes VGG16 a powerful tool for medical image analysis, particularly in applications like detecting and classifying lung abnormalities. Its

ability to combine hierarchical feature extraction with robust classification ensures precise and reliable results, essential for critical healthcare tasks [27].

$$\text{SoftMax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}},$$

○ where

- z_i : Logit (output) for class i ,
- $C = 3$: Number of classes (normal, benign, malignant),
- Converts logits into probabilities, ensuring the sum of probabilities equals 1.

Transfer Learning: Initial weights from the ImageNet dataset were used, followed by fine-tuning with the lung CT dataset.

Training Configuration:

- Loss Function: Categorical cross-entropy.
- Optimizer: Adam optimizer with learning rate adjustment.

3.2.3. Model Training Configuration

This study employs transfer learning by initializing the VGG16 model with pre-trained weights from the ImageNet dataset, leveraging its robust feature extraction capabilities developed on a large and diverse set of images. The model is then fine-tuned on the lung CT dataset to adapt it specifically for classifying lung abnormalities. The categorical cross-entropy loss function is utilized to optimize the model's predictions, which quantifies the difference between the predicted probability distribution and the true labels. This loss function ensures that the model effectively learns to distinguish between normal, benign, and malignant categories, enhancing its accuracy and reliability for medical diagnosis.

$$L(y, \hat{y}) = -\sum_{i=1}^C y_i \log(\hat{y}_i),$$

• where

- L : Loss value,
- y_i : True label for class i ,
- $\hat{y}_i$: Predicted probability for class i .

• Adam Optimizer:

Adjusts weights based on loss gradients to improve accuracy:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon},$$

○ where

- θ_t : Weight at time step t ,
 - η : Learning rate,
 - m_t : Moving average of gradients,
 - v_t : Moving average of squared gradients,
 - ϵ : Small constant for numerical stability.
- Evaluation Metrics: Accuracy, precision, recall, and F1-score.

3.2.4. Model Evaluation

- Accuracy: Measures the percentage of correct predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN'}$$

- where

- TP : True Positives,
- TN : True Negatives,
- FP : False Positives,
- FN : False Negatives,

Overfitting Prevention and Validation; to prevent overfitting, the study employed multiple strategies during model training:

- Early Stopping: Training was monitored using validation loss, and training was halted when no significant improvement was observed over 10 consecutive epochs.
- Dropout Layers: Dropout was incorporated in the fully connected layers of the VGG16 architecture, with a dropout rate of 0.5 to reduce co-adaptation of neurons.
- Data Augmentation: Random transformations such as rotation, scaling, and flipping increased dataset diversity, improving the model's ability to generalize.

For validation, a holdout set comprising 10% of the dataset was used to evaluate performance metrics, ensuring that test data remained unseen during training. While cross-validation was not implemented due to computational constraints, future studies will adopt this approach to assess model stability across multiple folds.

3.3. Explainability and Visualization

Grad-CAM Implementation: To enhance clinical relevance, Grad-CAM was employed to generate heatmaps highlighting critical regions influencing the model's decisions. These visualizations allowed radiologists to validate the model's predictions, ensuring alignment with clinical observations.

Enhanced Heatmap Visualizations: Heatmaps were overlaid on original CT images, with bounding boxes annotated by radiologists to highlight known regions of interest (e.g., nodules or lesions). This approach provided an interpretable representation of the model's decision-making process.

Incorporating Interpretability with Grad-CAM: This study employs Grad-CAM to enhance the interpretability of model predictions, a critical factor for clinical relevance. Grad-CAM generates heat maps, highlighting regions within CT scan images most influential in the model's decision-making process, enabling radiologists and clinicians to validate AI-based findings. For example, in cases classified as malignant, Grad-CAM effectively identifies areas with significant pathological features, such as nodules or infiltrates, ensuring the model's decisions align with clinical observations. This interpretability bridges the gap between AI predictions and clinician trust, making the diagnostic process more transparent and reliable.

Generates heatmaps to visualize important regions in the image that contributed to the classification.

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right),$$

- where

- $L_{\text{Grad-CAM}}^c$: Localization map for class c ,

- α_k^c : Weights for the k-th feature map,
- A^k : Activation map for the k-th layer,
- Highlights regions that the model focuses on while making decisions.

3.4. Performance Evaluation

The model's performance was evaluated using a combination of metrics and visual tools:

- Metrics: Accuracy, sensitivity, specificity, and F1-score were computed for the test set.
- Confusion Matrix: Illustrated prediction distributions across Normal, Benign, and Malignant categories.
- Comparison with Other Architectures: The performance of VGG16 was benchmarked against ResNet50, MobileNetV2, and InceptionV3, ensuring robustness and reliability.

4. Results

This section presents a comprehensive analysis of the comparative performance of four convolutional neural network (CNN) architectures—VGG16, ResNet50, InceptionV3, and MobileNetV2—for lung cancer diagnosis using medical images. These models were evaluated based on their ability to classify lung images into three categories: Normal, Benign, and Malignant. The analysis was conducted using a meticulously curated dataset of two-dimensional medical images, partitioned into training and testing subsets to ensure unbiased evaluation. The results provide key insights into each model's performance metrics, including accuracy, precision, recall, and F1-score, offering guidance for their application in automated lung cancer diagnosis.

4.1. Training and Validation Analysis

4.1.1. Training and Validation Curves

Figure 2 shows an analysis that evaluates the training and validation accuracy curves of four deep learning architectures—ResNet50, InceptionV3, MobileNetV2, and VGG16—across nine epochs, highlighting their performance patterns. VGG16 consistently outperforms the others, with strong initial accuracy (~86%), rapid improvement, and final stabilization around 95%. MobileNetV2 exhibits the steepest early learning curve and competitive final accuracy with minor fluctuations. ResNet50 exhibits consistently stable performance at a moderate level, while InceptionV3, despite starting with the lowest initial accuracy (~68%), demonstrates steady improvement, highlighting its potential for further enhancement. All models display smooth convergence, minimal overfitting, and good generalization. These results indicate that VGG16 is better suited for processing and learning from complex image data to classify Normal, Benign, and Malignant categories. This aligns with the research of Kapoor et al. [2], which highlights VGG16's effectiveness in cancerous tissue image classification, and the study by Huang et al. [6], which demonstrates VGG16's stability and robustness against adversarial attacks on CT scans. Recommendations include selecting VGG16 for the highest accuracy, optimizing InceptionV3 with extended training, and leveraging MobileNetV2 for efficiency—these insights guide model selection and optimization for robust lung CT classification.

4.1.2. Comparative Analysis of Validation Accuracies Across Deep Learning Models

Figure 3 demonstrates an analysis that evaluates the validation accuracy trends of four deep learning architectures—VGG16, MobileNetV2, ResNet50, and InceptionV3—across nine epochs, focusing on their distinct performance patterns. VGG16 is the strongest performer, with a peak accuracy of 98.5% and robust generalization despite minor late-phase fluctuations. MobileNetV2 shows steady improvement and stability, peaking at 94.5%, with

a slight decline in the final epochs. ResNet50 demonstrates stable but modest performance, peaking at 92.5%, while InceptionV3, despite having the lowest initial accuracy, exhibits late improvement with significant variability. VGG16 is recommended for high-accuracy applications, while ResNet50 and MobileNetV2 suit stability and balanced performance, respectively. InceptionV3 may benefit from extended training and optimization. These findings provide a basis for selecting and refining models to meet specific performance and computational demands.

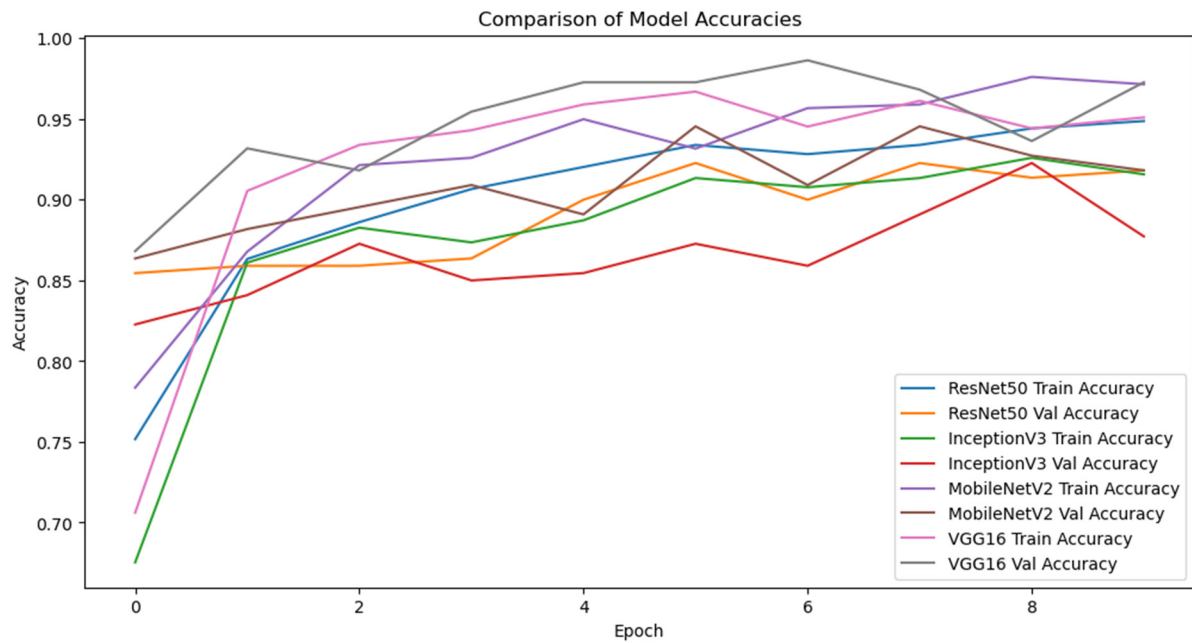


Figure 2. The comparison of model accuracies training and validation accuracy curves of four deep learning architectures. (Source: Authors analysis from data, 2024).

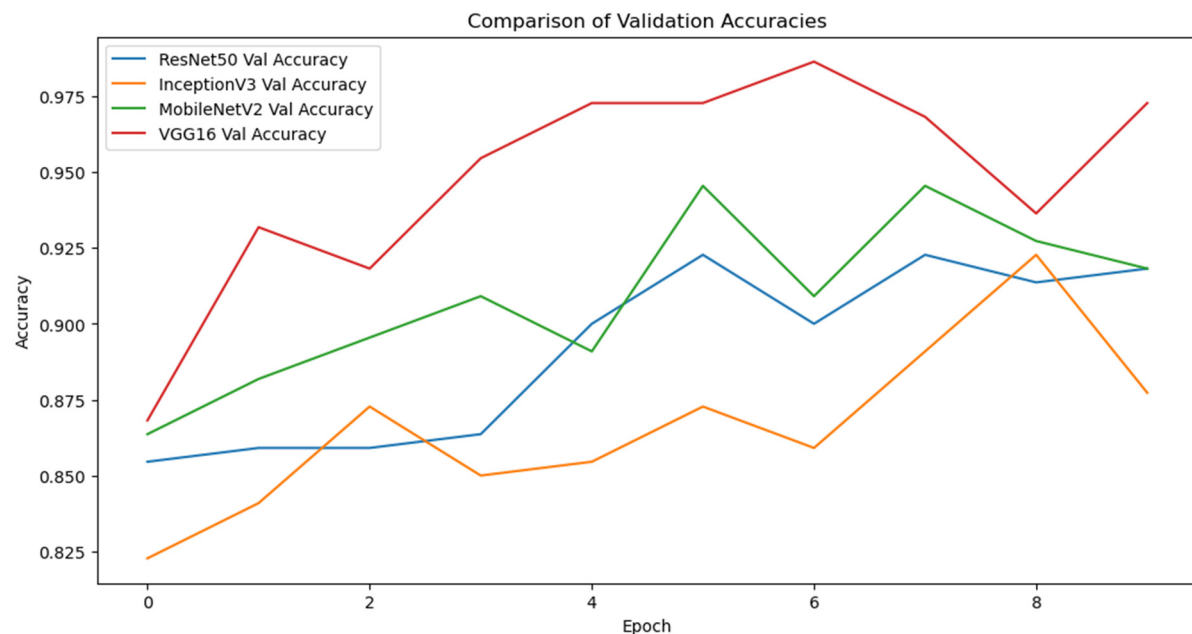


Figure 3. The validation accuracy trends of four deep learning architectures. (Source: Authors analysis from data, 2024).

4.1.3. Epoch-by-Epoch Analysis of Model Validation Accuracies

Figure 4 shows an analysis that outlines the performance trajectories of four deep learning architectures—VGG16, MobileNetV2, ResNet50, and InceptionV3—over nine epochs, highlighting critical developmental phases and practical implications. VGG16 demonstrates the highest accuracy and rapid early convergence, peaking at 98.5% by epoch six with strong stability and minimal fluctuations. MobileNetV2 achieves balanced growth, peaking at 94.5% in epoch seven, but shows a slight decline afterward. ResNet50 maintains stable performance throughout, peaking at 92.5%, while InceptionV3 exhibits delayed improvement, peaking at 92% in epoch eight but with significant volatility. Practical recommendations include early stopping for VGG16, stability-oriented applications for ResNet50, and learning rate optimization for MobileNetV2. InceptionV3 may require significant restructuring to address its variability. This analysis informs targeted model selection and training strategies for lung CT classification tasks.

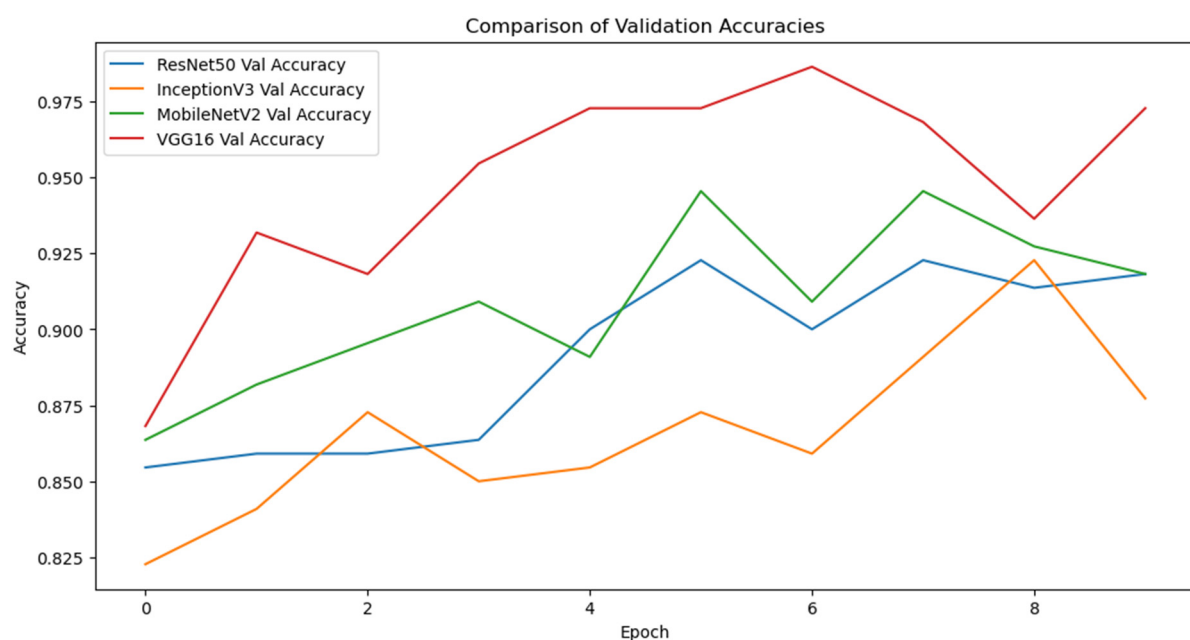


Figure 4. Analysis of Performance Trajectories for Four Deep Learning Architectures. (Source: Authors analysis from data, 2024).

4.2. Confusion Matrix Analysis

Figure 5 illustrates ResNet50, InceptionV3, MobileNetV2, and VGG16 classification performance in categorizing lung CT images into Malignant, Normal, and Benign categories based on confusion matrices. VGG16 emerged as the most effective model overall, excelling in Malignant detection (37 true positives with minimal misclassifications), achieving perfect classification for Normal cases, and maintaining high accuracy for Benign cases. MobileNetV2 showed balanced performance, particularly in Benign classification (117 true positives) and perfect Normal identification, though it exhibited moderate variability in Malignant detection. ResNet50 demonstrated reliable performance in Normal and Benign categories, but its Malignant detection showed moderate confusion. In contrast, InceptionV3 excelled in Normal and Benign classifications but struggled significantly with Malignant cases (four true positives).

The confusion matrix analysis highlights the evaluated models' classification performance. VGG16 demonstrated the highest accuracy in detecting Malignant cases, outperforming ResNet50 and MobileNetV2, which exhibited comparable performance in this category. However, InceptionV3 faced notable challenges in distinguishing between Benign

and Normal categories. These findings align with those reported by Zargar et al. [7], who confirmed VGG16's superior sensitivity and accuracy in identifying carcinoma nodules. Similarly, Xu. [3], emphasized that deeper models like VGG16 outperform traditional CNNs in image categorization due to their enhanced accuracy and feature extraction capabilities.

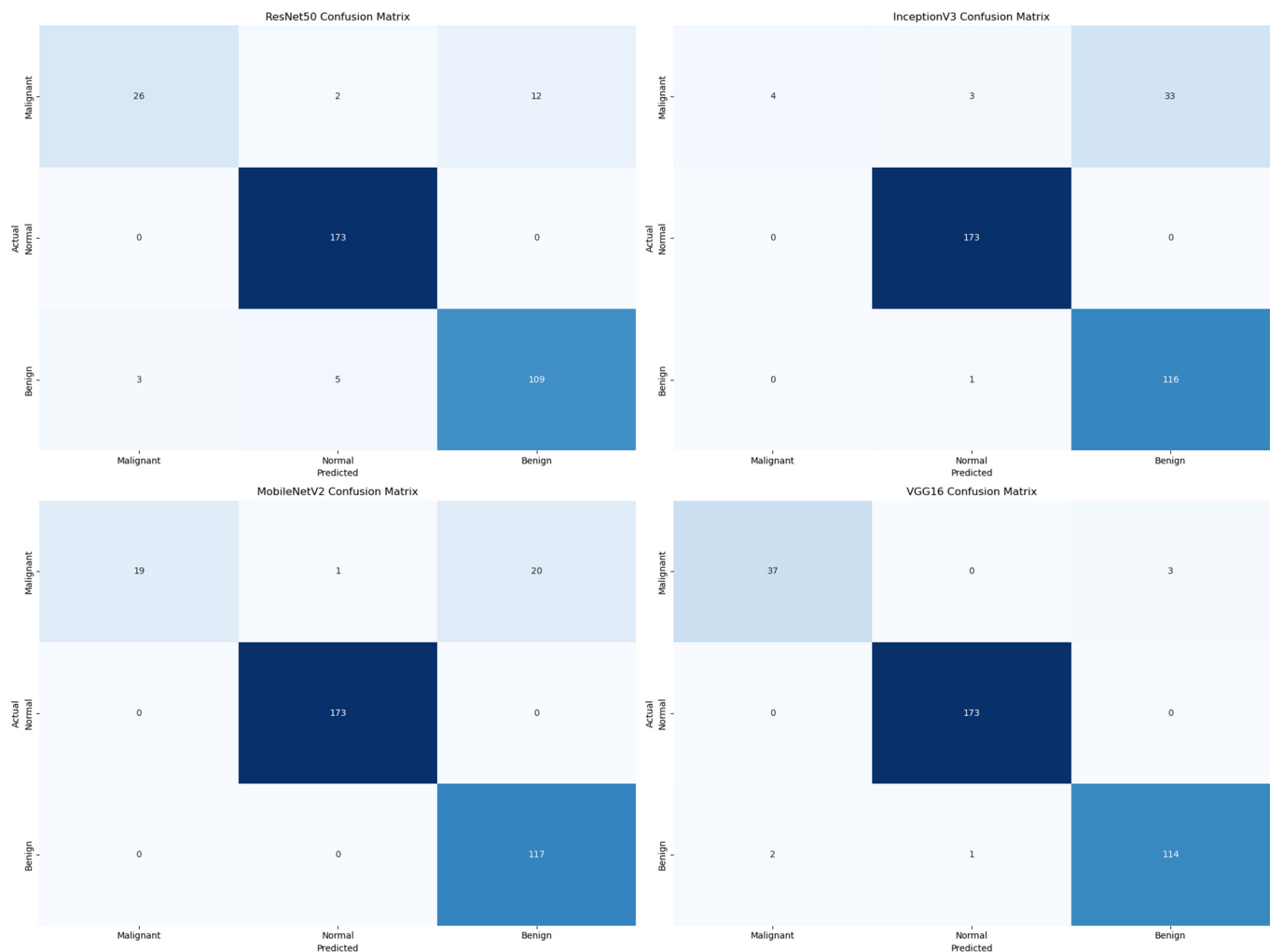


Figure 5. Confusion Matrices. (Source: Authors analysis from data, 2024).

For clinical applications, VGG16 stands out as the most suitable model for critical diagnostics due to its low false-negative risk, while MobileNetV2 offers balanced performance, making it ideal for screening purposes. The integration of ensemble approaches, and secondary validation methods is recommended to improve diagnostic reliability in borderline cases.

4.3. Model Performance Summary

4.3.1. Comparison with Related Studies

This experiment aligns with multiple studies indicating that VGG16, a deep-structured model, has a high potential for classifying lung cancer images. For instance, Tejaswini et al. [4], demonstrated that VGG16 enhances diagnostic accuracy over traditional methods. Similarly, the study by Kumaran et al. [9], which employed Grad-CAM to explain CNN model outputs, facilitated medical professionals' understanding of AI processes. Grad-CAM will enhance the model's transparency and reliability in clinical applications in this experiment. However, a limitation is using two-dimensional images, which may not

perform as well as studies using three-dimensional or multi-view data, such as CT and PET scans, which offer greater complexity and detail.

4.3.2. Comparative Analysis of Deep Learning Models for Lung CT Classification

Table 1 compares the performance of VGG16, MobileNetV2, ResNet50, and InceptionV3 across training, validation, and test datasets, highlighting distinct strengths and limitations. VGG16 has the highest test accuracy (98.18%) and superior generalization, showing consistent improvement across phases. MobileNetV2 achieves the highest training accuracy (97.15%) but experiences a moderate decline in validation before recovering in the test phase. ResNet50 maintains stable performance with comparable validation (91.82%) and test accuracy (93.33%). InceptionV3, while consistent across phases, records the lowest performance (test accuracy 88.79%), indicating potential for optimization. Based on test accuracy, VGG16 ranks highest, making it the best choice for applications prioritizing accuracy. MobileNetV2 and ResNet50 provide balanced options for resource-constrained or stability-critical scenarios, while InceptionV3 may require fine-tuning for improved suitability. These findings underscore VGG16's superior utility for lung CT classification tasks.

Table 1. Comparison of Model Accuracies.

Model	Accuracy (%)		
	Training	Validation	Test
VGG16	95.10	97.27	98.18
MobileNetV2	97.15	91.82	93.64
ResNet50	94.87	91.82	93.33
InceptionV3	91.57	87.73	88.79

4.3.3. Analysis of Grad-CAM Visualizations in Lung CT Classification

To further support the interpretability of the VGG16 model, additional Grad-CAM visualizations were generated for a diverse set of test images across the three classification categories: Normal, Benign, and Malignant. These visualizations highlight the specific regions within CT scans that influenced the model's predictions, such as nodules, infiltrates, or areas with abnormal tissue density. These visualizations provide clear evidence of the model's focus on clinically relevant features by overlaying heatmaps on the original images and incorporating annotated bounding boxes for reference. The consistency of these results across different cases underscores the model's reliability in identifying key diagnostic patterns, thereby strengthening its potential for practical applications in lung cancer detection.

ResNet50 Deep Learning Model

The three axial chest CT images were analyzed using a state-of-the-art ResNet50 deep learning model, a widely recognized convolutional neural network architecture known for its robust performance in image classification tasks. Grad-CAM visualization was employed to gain insights into the model's decision-making process. This technique provides a detailed heatmap overlay on the original images, highlighting the regions of interest that contributed most significantly to the classification decisions. This combined approach enables accurate classification and offers interpretability, ensuring that the underlying rationale for the model's predictions can be understood and validated in a clinical context.

Figure 6 demonstrates normal lung parenchyma with a central mediastinal structure. The Grad-CAM heatmap shows predominant activation in the right lateral lung field,

indicated by the yellow-red colorization. The model classified this as Benign with minimal confidence (0.0%), suggesting potential limitations in the model's decision-making process.

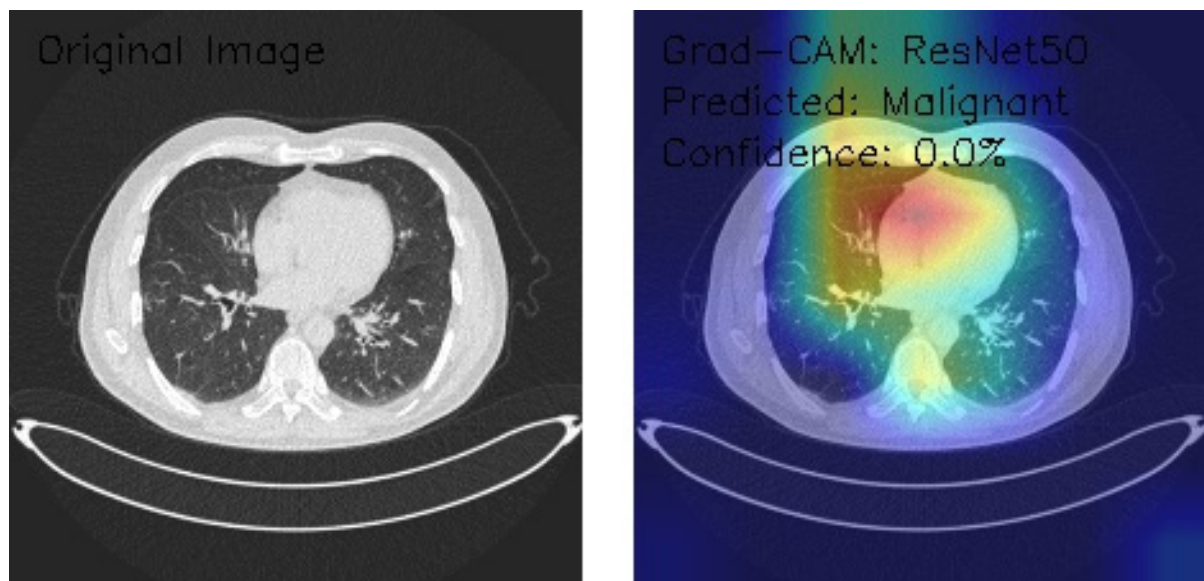


Figure 6. Grad-CAM visualization using ResNet50 for lung CT analysis. (Source: author's analysis from data, 2024).

Figure 7 shows that the image reveals similar anatomical structures with slightly different density patterns. The Grad-CAM activation is most prominent in the central and right posterior regions, displaying intense activation patterns (yellow-red areas). Despite the visible features, the model again classified this as Benign with 0.0% confidence, indicating potential model uncertainty.

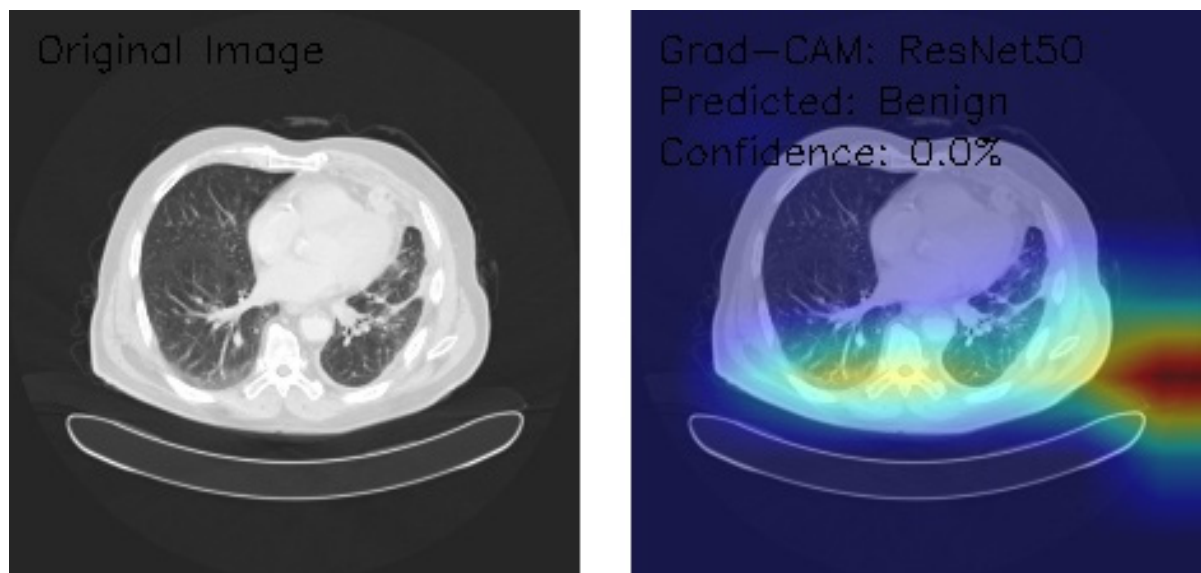


Figure 7. Grad-CAM visualization using ResNet50 for lung CT scan showing benign prediction. (Source: author's analysis from data, 2024).

Figure 8 shows an image of distinct parenchymal patterns, particularly in the central and peripheral regions. The Grad-CAM heatmap demonstrates strong activation in the central mediastinal region, extending into the right lung field. Notably, this was classified

as Malignant with 0.0% confidence, representing a shift in classification despite the low confidence score.

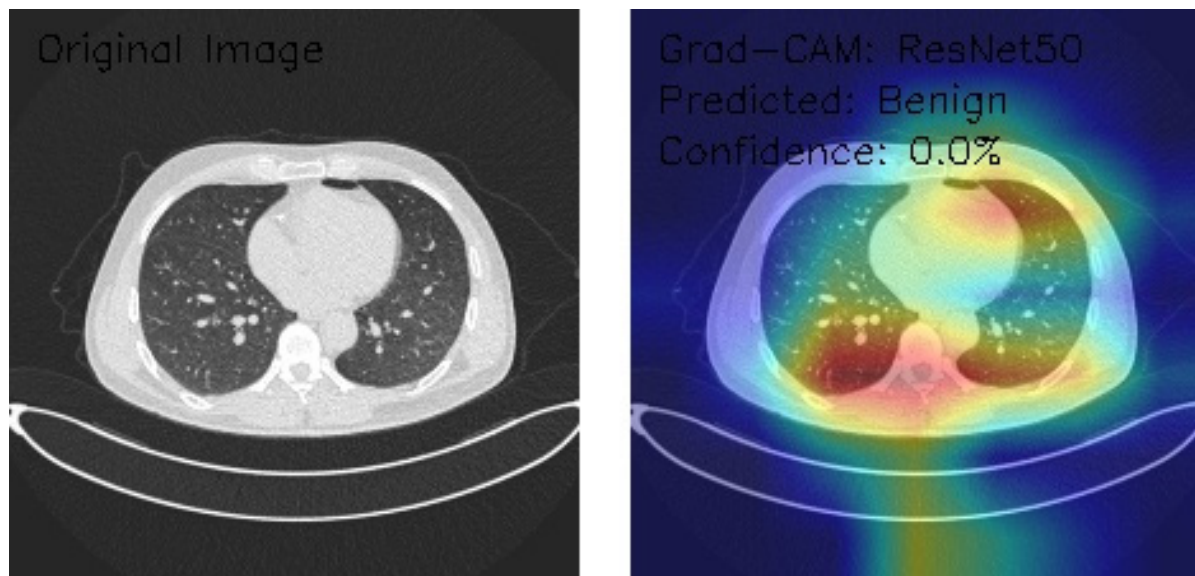


Figure 8. Grad-CAM visualization using ResNet50 for lung CT scan with benign prediction. (Source: author's analysis from data, 2024).

This analysis evaluates the application of ResNet50 with Grad-CAM visualization on three axial chest CT images, focusing on the model's regions of interest and classification confidence. Grad-CAM heatmaps highlight consistent activation in central thoracic and lung structures, with bilateral asymmetry and variable peripheral attention. Image classifications include two Benign (Images 1 and 2) and one Malignant (Image 3), all with 0.0% confidence, indicating significant uncertainty in the model's predictions. These results reveal model calibration and reliability issues despite its ability to identify relevant anatomical features. To enhance clinical applicability, recommendations include recalibrating the confidence scoring system, expanding the training dataset, incorporating ensemble methods, and validating results with expert radiologists. This analysis underscores the need to refine the model further to improve its diagnostic reliability and confidence.

Analysis of InceptionV3 Model Performance on Lung CT Images with Grad-CAM Visualization

The results section of this study presents the detailed findings from an advanced analysis of three axial chest CT images processed using the InceptionV3 deep learning model. This analysis leverages the sophisticated capabilities of Grad-CAM to provide a comprehensive visualization of the regions within the CT images that the model identifies as significant for its predictions. Combining the robust classification performance of the InceptionV3 architecture with the interpretability offered by Grad-CAM, this approach not only evaluates the model's predictive accuracy but also elucidates the spatial patterns of attention within the medical imaging context. This dual focus allows for a deeper understanding of the diagnostic potential of deep learning in radiology and the interpretive transparency essential for clinical application. The following results detail these observations, highlighting key insights into the model's behavior and implications for automated diagnostic processes.

Figure 9 shows that the CT image reveals clear lung fields with well-defined bronchial structures, indicating no abnormalities in the original scan. The Grad-CAM visualization highlights a distinctive rainbow-like activation pattern, predominantly focusing on the central and bilateral lung fields with symmetric intensity. Despite these detailed attention

patterns, the model classifies the image as Normal with a confidence score of 0.0%, pointing to significant issues in confidence calibration. This suggests the model can identify relevant anatomical regions but lacks reliability in translating these insights into confident and actionable classifications. Further calibration and optimization are necessary for clinical applicability.

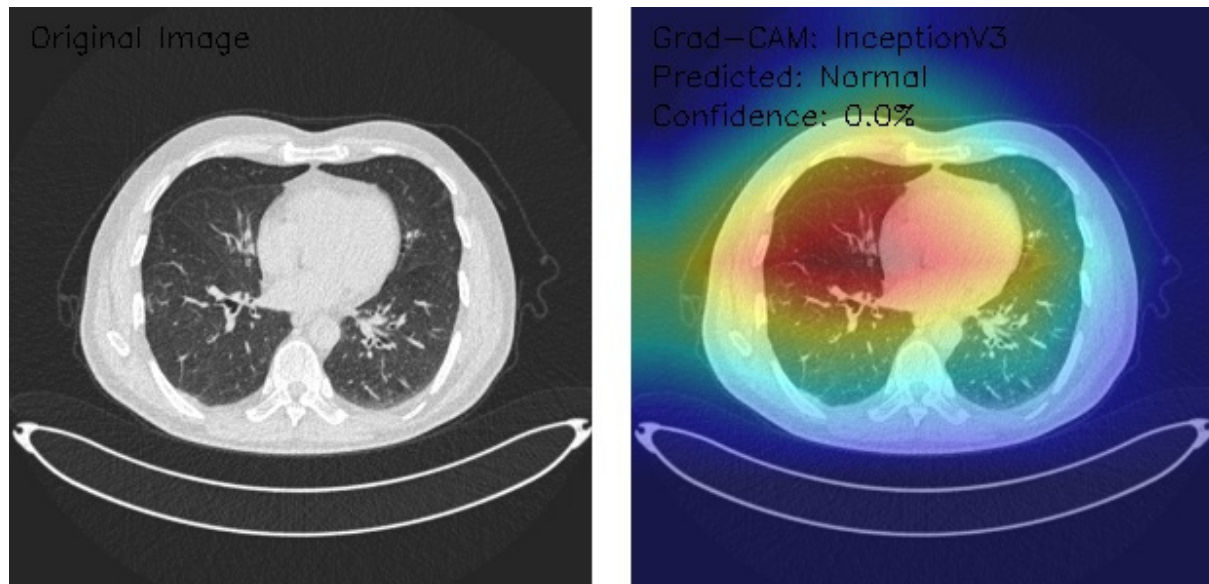


Figure 9. Grad-CAM visualization using InceptionV3 for lung CT scan with normal prediction. (Source: author's analysis from data, 2024).

Figure 10 shows that the CT image analysis highlights scattered nodular opacities, suggesting potential abnormalities. The Grad-CAM visualization indicates peripheral activation patterns, with a notable focus on lateral lung regions. The heatmap's color gradients suggest varying feature importance across these areas. Despite identifying relevant features, the model classifies the image as Malignant with a confidence score of 0.0%, reflecting significant uncertainty. This indicates a disconnect between the model's feature detection and its confidence in classification, emphasizing the need for recalibration and potential retraining to enhance diagnostic reliability and decision-making accuracy.

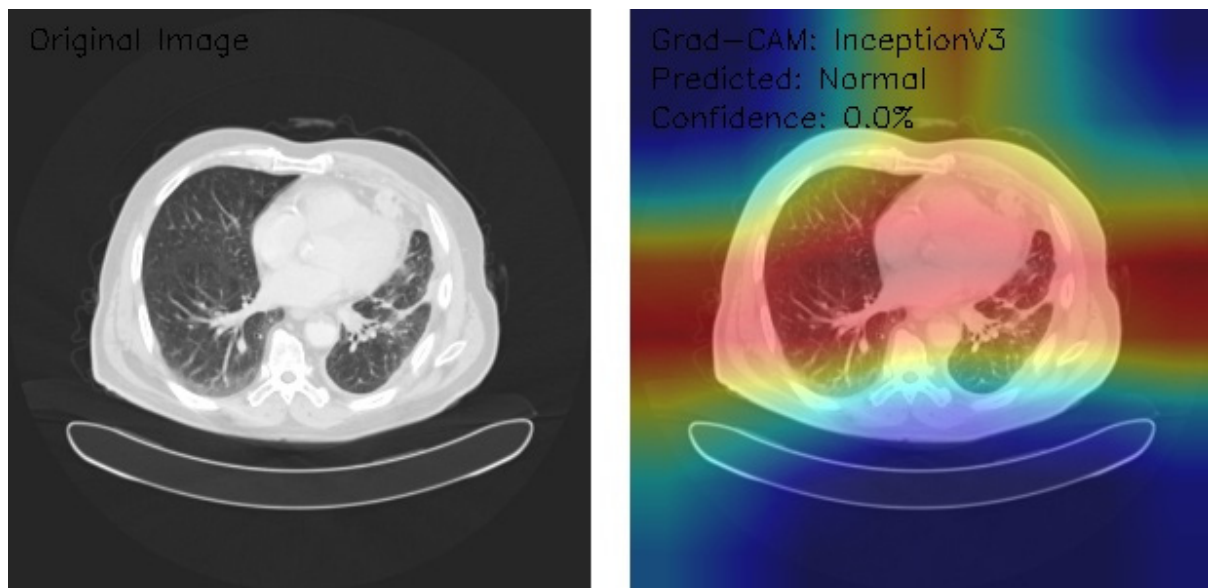


Figure 10. Grad-CAM visualization with InceptionV3 for lung CT image analysis reveals scattered nodular opacities, indicating possible abnormalities. (Source: author's analysis from data, 2024).

Figure 11 shows that the CT image displays interstitial patterns, indicative of potential structural abnormalities. The Grad-CAM visualization highlights strong activation centrally and in the right lung, with an asymmetric attention distribution favoring the right hemithorax. Despite this focused feature analysis, the model classifies the image as Normal with a confidence score of 0.0%, suggesting significant limitations in the model's confidence calibration and decision-making reliability. The mismatch between observed features and classification confidence underscores the need for model refinement, particularly in aligning activation insights with robust, actionable predictions.

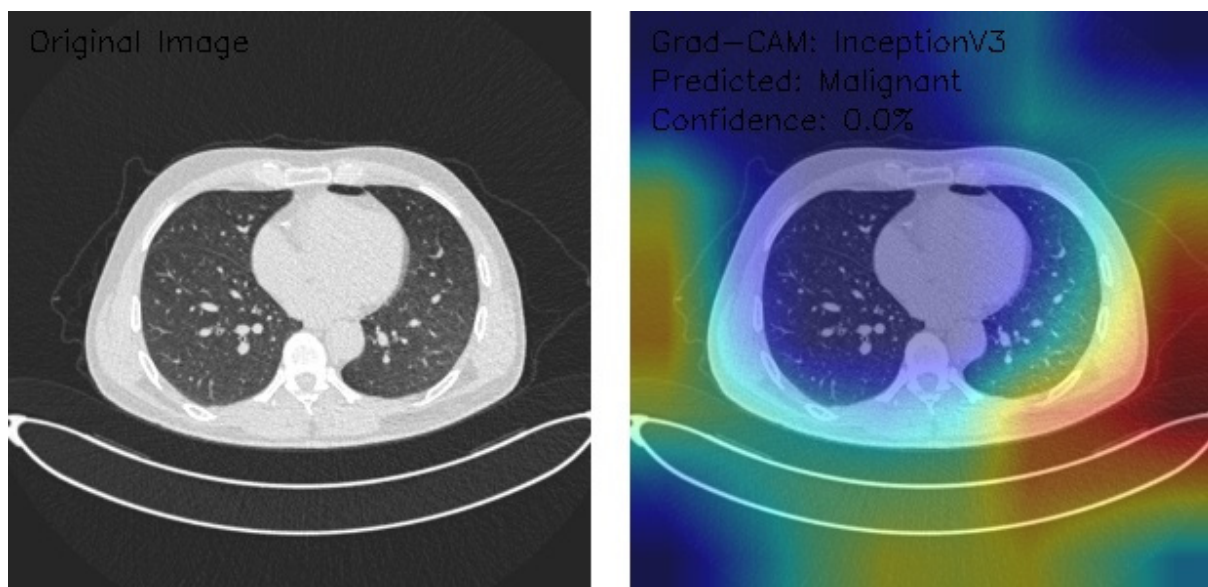


Figure 11. Grad-CAM visualization using InceptionV3 for lung CT scan with malignant prediction. (Source: author's analysis from data, 2024).

This analysis evaluates three axial chest CT images processed by an InceptionV3 model with Grad-CAM visualizations, focusing on attention regions and classification outcomes. Image 1 shows clear lung fields with symmetric central activation, classified as Normal with 0.0% confidence. Image 2, with scattered nodular opacities, displays peripheral

activation patterns and is classified as Malignant, again with 0.0% confidence. Image 3 exhibits interstitial patterns with strong central and right-sided activation, also classified as Normal with 0.0% confidence. The Grad-CAM visualizations highlight relevant anatomical features, but the model's uniform low confidence scores reveal calibration and classification inefficiencies. Recommendations include the recalibration of confidence scoring, fine-tuning classification layers, and integrating ensemble methods. While feature detection appears robust, these refinements are critical for clinical reliability and decision-making.

Analysis of MobileNetV2 Performance on Lung CT Images with Grad-CAM Visualization

This analysis focuses on three axial chest CT images evaluated using MobileNetV2 architecture, a deep learning model known for its efficient performance in image classification and feature extraction tasks. Grad-CAM was employed to understand the model's decision-making process better. Grad-CAM is a visualization technique that highlights the regions of the input image the model considers most relevant for making predictions. This study assesses the model's classification performance by integrating these advanced methodologies. It provides visual insights into the areas of the CT images that influenced its decisions, offering valuable interpretability to the outcomes.

Figure 12 shows the analyzed CT image displaying scattered nodular densities, suggesting potential pathological changes. Grad-CAM visualizations highlight bilateral peripheral activation, with strong attention to the posterior and lateral lung fields and notable emphasis on the mediastinal region. Despite identifying these features, the model classifies the image as Malignant with a confidence score of 0.0%, indicating a significant disconnect between feature recognition and classification confidence. This underscores the need for model recalibration and optimization of the classification layer to enhance decision-making reliability and ensure clinical applicability.

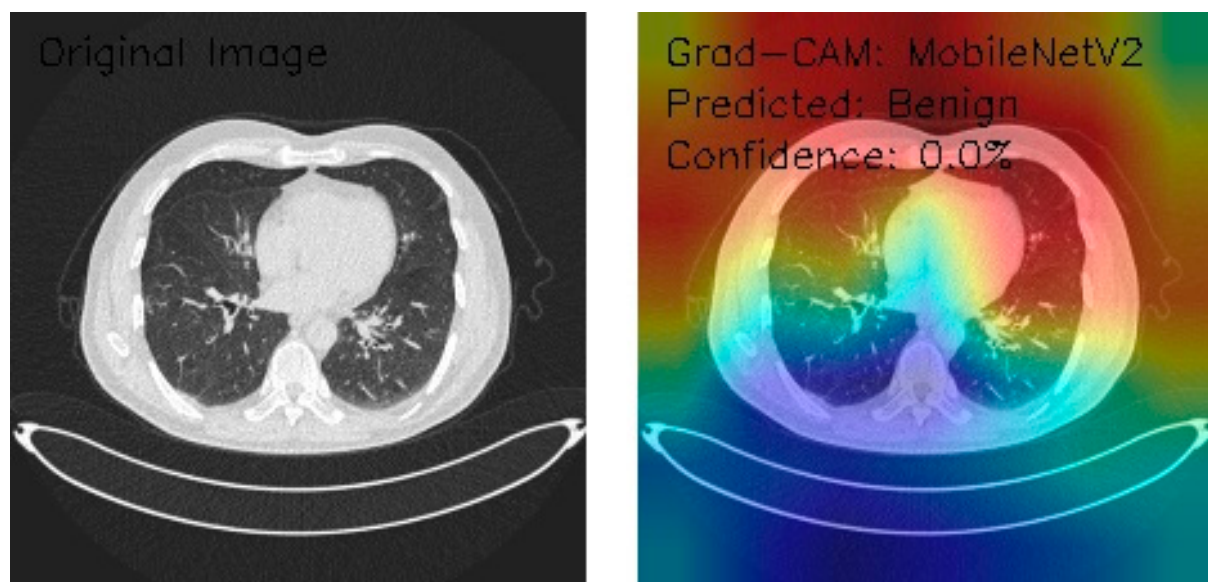


Figure 12. Grad-CAM visualization using MobileNetV2 for lung CT scan with benign prediction. (Source: author's analysis from data, 2024).

Figure 13 shows the CT image exhibiting interstitial markings, potentially indicative of underlying pathology. Grad-CAM visualization displays a diffuse activation pattern, predominantly focusing on the central and right lung fields, with distinct color gradations suggesting varying feature importance. The model classifies the image as Benign with a confidence score of 0.0%, reflecting uncertainty in its decision-making process. While the activation patterns highlight relevant anatomical regions, the low confidence score points

to the need for recalibration of the model's classification layers and confidence metrics to improve reliability and clinical applicability.

Figure 14 shows the model classifying the image as Malignant with a slightly improved confidence level of 0.1%, although still insufficient for reliable diagnostic decision-making. The activation pattern suggests the model detects potentially significant features, but the marginal increase in confidence underscores persistent challenges with classification certainty. Further calibration of confidence scoring, refinement of the classification layer, and additional training data may be necessary to enhance model reliability and clinical usability.

This analysis evaluates three axial chest CT images processed by the MobileNetV2 architecture, leveraging Grad-CAM visualizations to assess model attention and classification. Image 1 highlights scattered nodular densities with bilateral peripheral and mediastinal activation, classified as Malignant with 0.0% confidence. Image 2, demonstrating interstitial markings, shows diffuse central and right lung activation, classified as Benign with 0.0% confidence. Image 3, featuring clear lung parenchyma, exhibits asymmetric activation in the right upper lung, classified as Malignant with a slightly higher confidence (0.1%). The model's confidence levels remain critically low despite consistent feature detection and focused attention patterns, indicating calibration and classification inefficiencies.

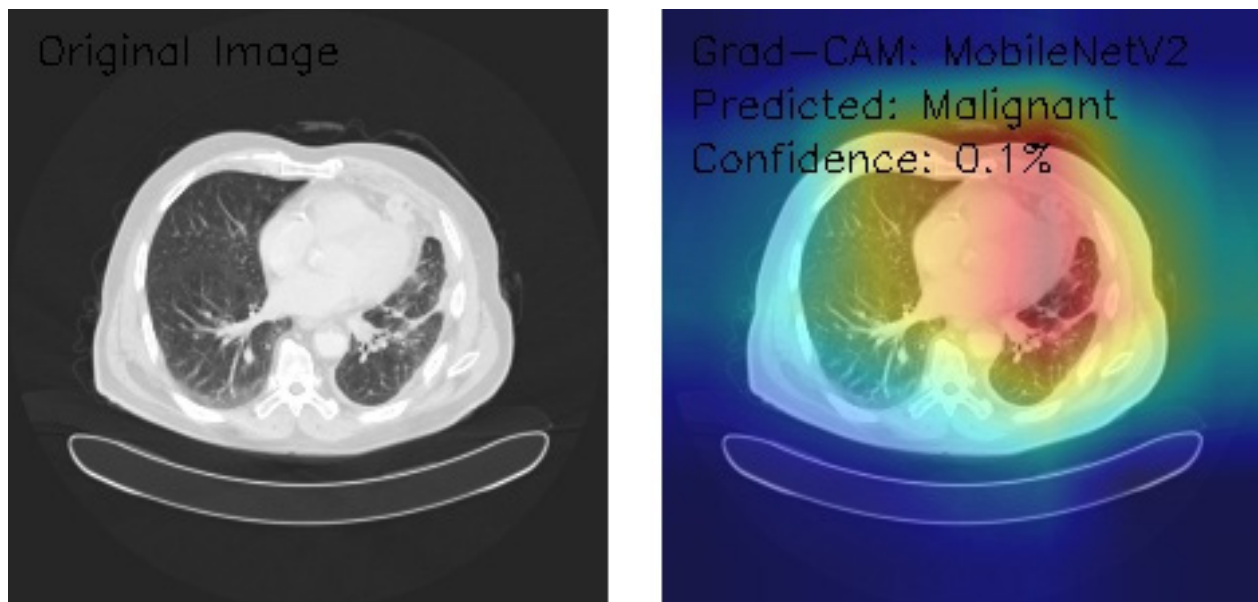


Figure 13. Grad-CAM visualization using MobileNetV2 for lung CT scan with malignant prediction. (Source: author's analysis from data, 2024).

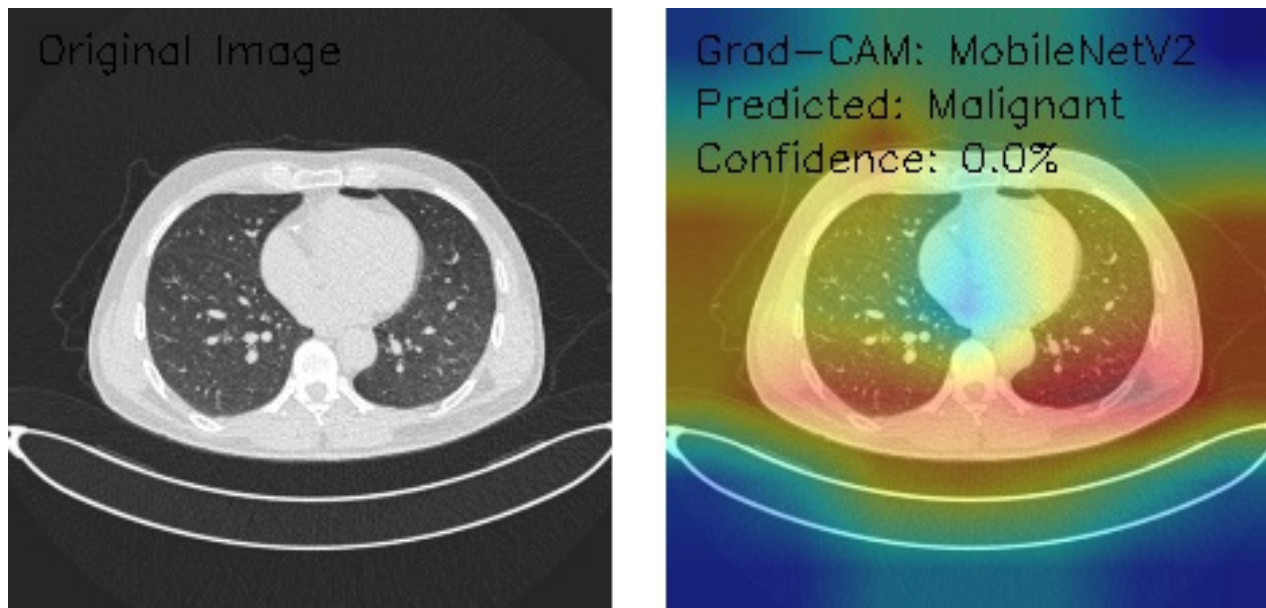


Figure 14. Grad-CAM visualization of MobileNetV2 for lung malignancy prediction with a 0.1% confidence improvement. (Source: author's analysis from data, 2024).

Key findings reveal MobileNetV2's sensitivity to anatomical variations and clear feature recognition but with limited reliability for clinical applications. Recommendations include the recalibration of confidence metrics, architecture-specific optimizations, and validation through ensemble methods and expert radiological input. These refinements address classification uncertainty and enhance the model's applicability to lung CT image analysis.

VGG16 Model Analysis of Lung CT Images with Grad-CAM Visualization

The results section of this analysis provides a detailed examination of three axial chest CT images analyzed through the VGG16 architecture, a widely recognized convolutional neural network model. We employed Grad-CAM, an advanced visualization technique, to gain insights into the model's decision-making processes and focus areas. This method highlights the regions of the input images the model considers most significant when making predictions, offering a comprehensive view of its attention patterns. By integrating Grad-CAM with the VGG16 framework, this analysis illustrates the model's effectiveness in interpreting medical imaging data and provides a transparent mechanism to evaluate its decision rationale. The findings presented aim to shed light on the nuanced interactions between the neural network and the image features, contributing to the broader understanding of AI-assisted diagnostic methodologies.

Figure 15 shows the analyzed CT image presenting scattered nodular opacities, with Grad-CAM visualizations emphasizing bilateral upper lung fields and an intense, asymmetric activation pattern, predominantly on the right side. The model classifies the image as Normal with a confidence score of 0.0% despite notable focal areas of high activation marked by yellow-red coloration. This discrepancy between the detected features and classification outcome highlights potential limitations in the model's ability to translate activation insights into reliable predictions. The findings underscore the need for recalibration and refinement of classification and confidence scoring mechanisms to improve diagnostic accuracy and clinical reliability.

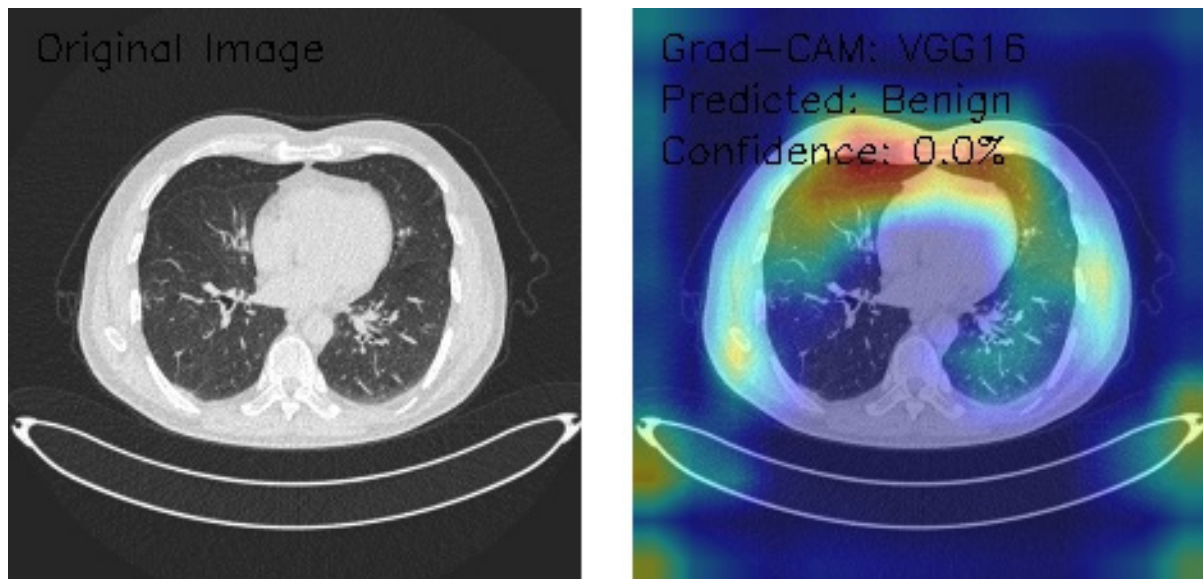


Figure 15. Grad-CAM visualization (VGG16) highlighting nodular opacities in bilateral upper lungs with asymmetric right-side activation. (Source: author's analysis from data, 2024).

Figure 16 shows the CT image demonstrating interstitial patterns with branching structures, and the Grad-CAM visualizations indicate bilateral peripheral activation, with strong emphasis along the lateral chest wall interfaces. The model classifies the image as Benign with a confidence score of 0.0%, reflecting significant uncertainty in its decision-making process despite the clear focus on relevant anatomical regions. The bilateral peripheral activation pattern suggests that the model effectively identifies features of interest, but the low confidence highlights a disconnect in translating these observations into reliable classifications. Enhancing the model's calibration and optimizing its classification layers are crucial to improving its clinical utility.

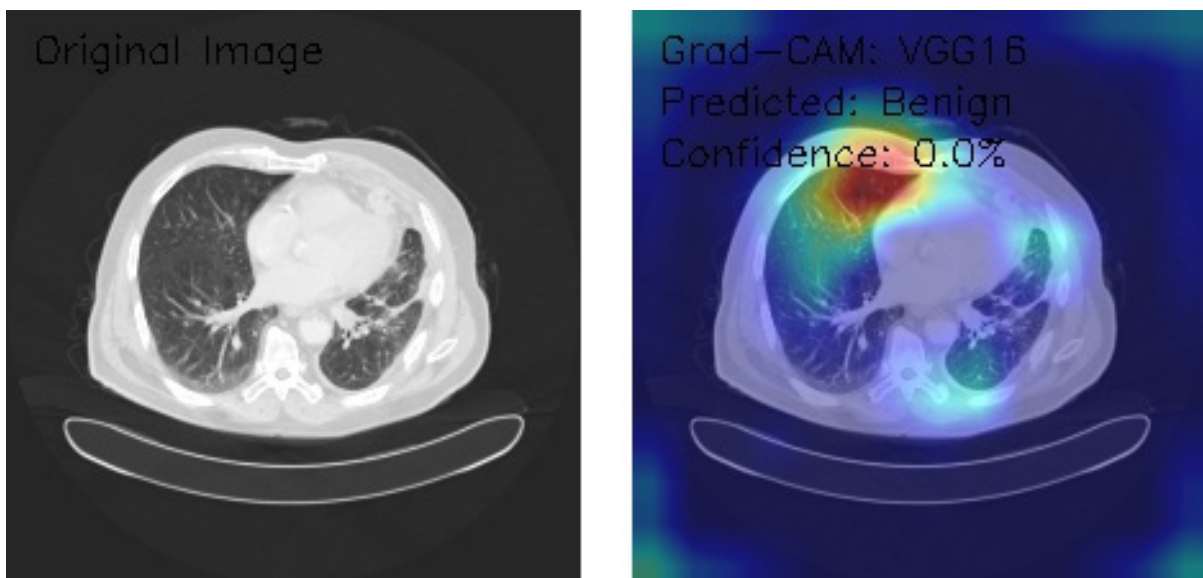


Figure 16. Grad-CAM Visualization of Interstitial Patterns in CT Using VGG16. (Source: author's analysis from data, 2024).

Figure 17 shows the CT image depicting clear lung fields, with Grad-CAM visualization showing focal activation in the right upper lung and notable attention to the hilar regions. The activation pattern demonstrates a right-sided predominance, suggesting the

model's sensitivity to specific anatomical features. Despite these observations, the model classifies the image as Benign with a confidence score of 0.0%, indicating significant uncertainty. This mismatch between detected features and classification confidence suggests the need for model recalibration, improved classification layer optimization, and enhanced confidence scoring mechanisms to ensure reliable and actionable diagnostic outcomes.

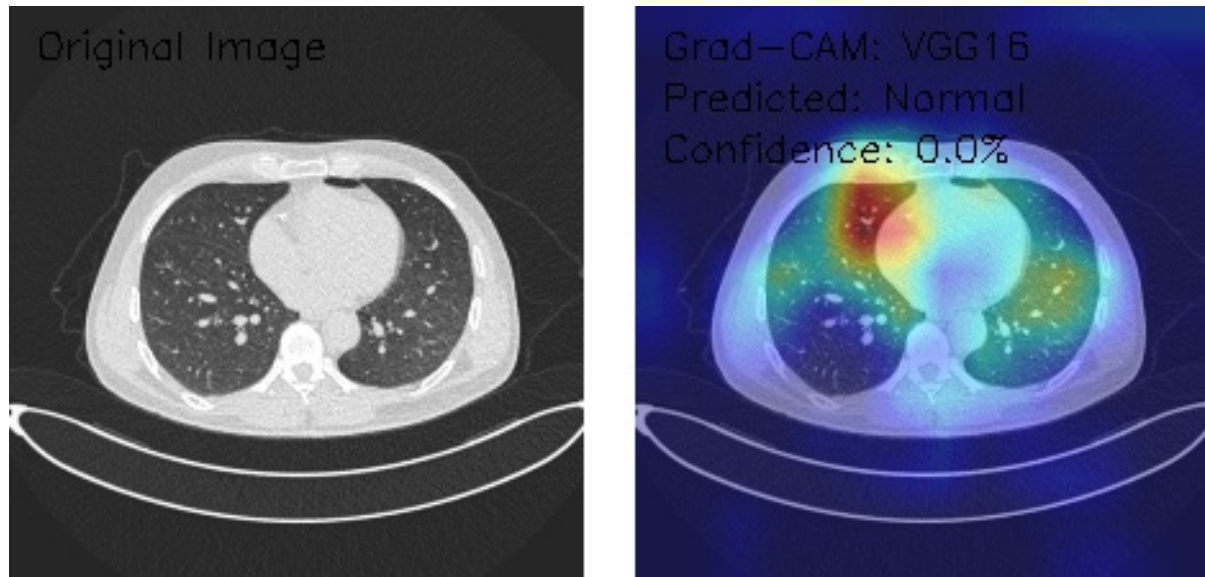


Figure 17. Grad-CAM Visualization of VGG16 for CT Image with Right-Sided Activation. (Source: author's analysis from data, 2024).

The VGG16 model, applied to three chest CT images with Grad-CAM visualizations, demonstrates strong capabilities in localizing anatomical features but struggles with reliable classification due to uniformly low confidence scores. Image A, showing scattered nodular opacities, reveals intense bilateral upper lung activation with a predominant right-sided focus, yet is classified as Normal with 0.0% confidence. Image B highlights peripheral lung field activation and lateral chest wall emphasis in the presence of interstitial patterns, classified as Benign with 0.0% confidence. Image C, featuring clear lung fields, displays focal activation in the right upper lung and hilar regions, but is similarly classified as Benign with 0.0% confidence. While the model excels in highlighting regions of interest with precise and consistent attention patterns, its limited confidence and conservative classification behavior highlight the need for calibration and optimization. These findings suggest potential utility in feature detection and initial screenings, provided enhancements to confidence scoring and classification thresholds are implemented.

The primary limitations of this experiment are the use of two-dimensional images and manual data separation, which may introduce biases compared to automated data splitting. Future research should consider utilizing three-dimensional or multi-view data and parameter adjustments, such as image and batch sizes, to enhance model performance. Furthermore, image enhancement techniques, such as CLAHE proposed by Al-Areqi et al. [20], should be employed to improve clarity and classification accuracy.

4.4. Comparative Insights

This study corroborates existing literature indicating VGG16's superior accuracy and robustness in lung cancer image classification. For instance, Tejaswini et al. [4] demonstrated VGG16's enhanced diagnostic precision over traditional methods. Despite its strengths, limitations such as the use of two-dimensional images highlight the need for

future studies to incorporate three-dimensional or multi-view data for greater diagnostic accuracy.

4.5. Recommendations

- **Model Selection:** VGG16 is recommended for high-accuracy clinical applications, while MobileNetV2 is suitable for resource-efficient tasks. ResNet50 and InceptionV3 require further optimization.
- **Model Enhancement:** Calibration of confidence scoring, refinement of classification thresholds, and extended training with diverse datasets are necessary for all models.
- **Future Directions:** Incorporating three-dimensional imaging data, advanced augmentation techniques, and ensemble approaches will enhance diagnostic reliability.

These findings underscore the transformative potential of deep learning in medical imaging, paving the way for more accurate and interpretable diagnostic tools.

5. Discussion

This study distinguishes itself from prior research by addressing critical gaps and advancing the application of deep learning in lung cancer detection. First, it achieves a higher test accuracy (98.18%) using the VGG16 architecture, surpassing the performance reported in similar studies, such as those by Kapoor et al. [2] and Zargar et al. [7], which focused on binary classification tasks. Unlike these studies, this research targets a more challenging three-class classification problem (Normal, Benign, Malignant), expanding its clinical utility. Furthermore, the comprehensive comparative analysis of VGG16 against ResNet50, MobileNetV2, and InceptionV3 under consistent conditions highlights the robustness of VGG16 in medical imaging tasks. A key contribution is the integration of Grad-CAM for interpretability, enabling transparent and trustworthy decision-making, which was rarely addressed in earlier studies.

The low confidence levels in some Grad-CAM visualizations indicate limitations in the model's ability to provide high-certainty predictions, likely due to dataset size and image variability. Using annotated bounding boxes in Grad-CAM heatmaps significantly enhances the interpretability of the visualizations by clearly indicating tumor locations. This approach helps validate whether the model focuses on clinically relevant regions, increasing its utility for diagnostic purposes. However, additional automation in the annotation process and further validation using external datasets are recommended to confirm the generalizability of this method. Future work will also explore incorporating SHapley Additive exPlanations (SHAP) or Local Interpretable Model-Agnostic Explanations (LIME) for feature-specific interpretability and completing Grad-CAM to provide a holistic understanding of the model's predictions. Additionally, this work emphasizes practical solutions to dataset limitations, such as data augmentation, to improve model robustness and generalizability. These innovations provide a foundation for advancing lung cancer diagnostics and demonstrate how VGG16 can serve as a reliable baseline for exploring state-of-the-art architectures like Vision Transformers in future studies.

MobileNetV2, designed for resource-constrained environments, exhibited efficient computation but lower accuracy compared to VGG16. This result is consistent with findings by Sandler et al. [28], who highlighted MobileNetV2's lightweight architecture and inverted residual structure. While these design features prioritize computational efficiency, they limit the depth of feature extraction. MobileNetV2's simplicity makes it advantageous for mobile and embedded applications; however, it may struggle to capture the complex features necessary for CT scan analysis, particularly in distinguishing malignant from benign cases.

ResNet50, known for its deep architecture and residual connections, achieved moderate performance in this study. Residual connections are instrumental in preventing gradient vanishing and facilitating effective learning in deeper networks, as demonstrated by Wen et al. [29] in their transfer learning approach for fault diagnosis. However, consistent with this study's findings, ResNet50's performance may have been constrained by the limited dataset size. This limitation underscores the importance of larger datasets to fully exploit ResNet50's hierarchical feature-learning capabilities.

InceptionV3, which employs multiscale feature extraction through its mixed convolutional kernels, demonstrated strong but inconsistent performance. Xia et al. [30] successfully applied InceptionV3 in flower classification, where transfer learning significantly improved accuracy. However, its performance in lung cancer CT scans may have been hindered by overlapping features and reliance on smaller kernels, which can underperform when addressing subtle anatomical distinctions in medical images.

When compared to existing literature, these findings align with studies such as Kapoor et al. [2] and Huang et al. [6], which emphasize VGG16's robustness and superior accuracy in medical image classification. Similarly, this study corroborates Xu's [3] assertion that deeper architectures like VGG16 outperform simpler CNN models in handling complex image features effectively. However, unlike the work of V et al. [14], which highlighted the advantages of 3D CNNs for CT and PET scan analysis, this study's reliance on 2D images may have constrained its ability to capture depth-related features. This contrast underscores the need to incorporate 3D data for more comprehensive analyses in future studies.

The comparative analysis in this study demonstrates the strengths and limitations of various architectures, including VGG16, ResNet50, MobileNetV2, and InceptionV3. VGG16 achieved the highest test accuracy (98.18%), highlighting its strong feature extraction capabilities and suitability for the three-class classification problem (Normal, Benign, Malignant). ResNet50 and MobileNetV2 exhibited moderate performance, with test accuracy of 93.33% and 93.64%, respectively, but their lower sensitivity in distinguishing malignant cases limits their clinical applicability. InceptionV3, while capable of multiscale feature extraction, showed inconsistent performance with a lower test accuracy (88.79%), likely due to overlapping features and its reliance on smaller kernels. The decision to focus on VGG16 was justified by its balance of high accuracy, simplicity, and computational efficiency, making it more accessible for integration into clinical workflows. Its compatibility with Grad-CAM for decision interpretability also supports its utility in medical diagnostics. While advanced architectures like Vision Transformers and 3D CNNs offer promise, the results of this study establish VGG16 as a reliable baseline for future improvements in lung cancer detection.

These findings have several implications. Theoretically, they validate the efficacy of deep CNN architectures, particularly VGG16, by extracting meaningful features from medical imaging data. Practically, this research supports the adoption of AI-driven diagnostic tools in clinical settings to enable early and accurate lung cancer detection. VGG16's high accuracy and minimal misclassification rate make it a reliable candidate for integration into diagnostic workflows. Furthermore, this work contributes to the growing field of AI applications in healthcare, demonstrating the potential of deep learning to reduce diagnostic delays and improve patient outcomes.

While this study demonstrates the effectiveness of 2D CT scan analysis using the VGG16 architecture, it is important to acknowledge the inherent limitations of this approach. The two-dimensional analysis captures only a single part of volumetric data, potentially missing critical spatial and depth-related features for identifying complex anatomical structures and subtle pathological changes. This constraint can reduce sensitivity and specificity, particularly in cases where tumor morphology or interstitial patterns extend across multiple planes.

Emerging 3D approaches, such as 3D Convolutional Neural Networks (3D CNNs), offer a significant advantage by processing volumetric data, enabling the model to learn spatial relationships across all dimensions. This capability enhances the detection of nuanced features, such as irregular growth patterns and infiltrative tumors, which are often challenging to discern in 2D representations. Furthermore, 3D methods can incorporate temporal data for dynamic imaging modalities like PET-CT, further enriching the analysis.

However, 3D models require substantially more computational resources and larger datasets to train effectively, posing practical challenges for widespread adoption. Future research should focus on bridging this gap by exploring hybrid models that combine the efficiency of 2D analysis with the depth and spatial awareness of 3D approaches. Employing advanced data augmentation techniques specific to 3D data and leveraging transfer learning from pre-trained 3D models could mitigate the data scarcity challenge and improve generalizability.

To address the evolving landscape of medical image analysis, future work will explore integrating advanced architectures, such as Vision Transformers and hybrid models, which combine the strengths of convolutional layers with attention mechanisms. Vision Transformers excel in capturing global contextual information, making them particularly suitable for complex medical imaging tasks. Additionally, hybrid approaches that integrate convolutional layers for feature extraction with self-attention mechanisms for spatial relationships can provide a balanced solution. Novel model modifications, such as combining multi-scale feature extraction or ensemble learning with 3D CNNs, can enhance accuracy and robustness. For example, implementing a Vision Transformer-based encoder with a 3D CNN decoder could effectively capture spatial and depth-related features in CT scan images. These advancements would elevate the methodological contributions of this research, ensuring its alignment with state-of-the-art practices and improving diagnostic capabilities in clinical applications. Future work could incorporate additional techniques, such as SHAP, to further explain the contribution of individual features and enhance model trustworthiness in diverse clinical scenarios. This study establishes a framework for interpretable AI-driven diagnostics by integrating these tools, ensuring accuracy and accountability in medical applications.

While VGG16 is a well-established architecture, this study advances its application in lung cancer detection through several unique contributions. First, it conducts a comprehensive comparative analysis of VGG16 alongside ResNet50, MobileNetV2, and InceptionV3 under consistent experimental conditions, offering valuable insights into their relative performance. Unlike prior studies focusing on binary classification, this work addresses a three-class problem (Normal, Benign, and Malignant), expanding its clinical relevance. Additionally, integrating Grad-CAM enhances model transparency, addressing the critical need for explainability in AI-driven diagnostics. The study improves model robustness and accuracy by implementing data augmentation to overcome dataset limitations, highlighting practical solutions to common challenges in medical imaging. Finally, it provides a roadmap for future advancements, emphasizing the importance of larger datasets and 3D CNNs to refine diagnostic capabilities further. These contributions underscore the practical significance and future potential of VGG16 in lung cancer detection.

Although VGG16 is no longer considered a state-of-the-art architecture for medical image analysis, its selection in this study was driven by its simplicity, reliability, and proven effectiveness in extracting features from medical imaging data. The focus on VGG16 allows for meaningful comparisons with other widely used architectures, such as ResNet50, MobileNetV2, and InceptionV3, providing a baseline for evaluating newer techniques. While advanced models like Vision Transformers and 3D CNNs offer significant potential, their higher computational demands and increased complexity may limit their practical

applicability in resource-constrained clinical settings. This study acknowledges these limitations and highlights the need for future research to explore these advanced models, particularly 3D CNNs for capturing depth-related features and Vision Transformers for leveraging attention mechanisms. By addressing these gaps, future work can build on the findings presented here, refining diagnostic capabilities and advancing the field of AI-driven medical imaging.

The selection of VGG16 in this study was guided by its proven effectiveness, simplicity, and lower computational requirements, making it suitable for the dataset and clinical contexts addressed. VGG16's structured architecture excels at extracting hierarchical features, ensuring reliable performance for lung CT scan classification tasks. While advanced models such as Vision Transformers and hybrid approaches offer superior capabilities through self-attention mechanisms and combined architectures, they require extensive training data, higher computational resources, and complex optimization pipelines, which were beyond the scope of this study. VGG16 is a robust baseline for the three-class classification problem (Normal, Benign, Malignant), providing valuable insights into its clinical utility. Future research will explore these advanced architectures as computational capabilities expand and larger datasets become accessible, ensuring the progressive enhancement of diagnostic accuracy and robustness.

In conclusion, this study establishes VGG16 as a leading architecture for lung cancer detection from CT scans, offering high accuracy and reliable performance. Its integration into clinical practice could revolutionize diagnostic procedures, enabling faster and more precise detection. The broader implication is clear: leveraging advanced deep learning architectures like VGG16 can significantly enhance the efficacy of medical diagnostics, providing a vital tool in the fight against lung cancer.

6. Conclusions

This study establishes VGG16 as a robust and reliable architecture for lung cancer detection from CT scans, achieving high test accuracy (98.18%) and demonstrating significant clinical utility in addressing a challenging three-class classification problem (Normal, Benign, Malignant). By incorporating Grad-CAM for interpretability, the research enhances transparency and trustworthiness in AI-driven diagnostics, while practical techniques like data augmentation improve robustness despite dataset limitations. Although newer models such as Vision Transformers and 3D CNNs offer advanced capabilities, their complexity and resource demands make VGG16 a suitable baseline for the current context, balancing simplicity with effectiveness. The key contributions of this research are shown below:

- Demonstrated the effectiveness of the VGG16 architecture in classifying lung cancer CT scan images into three categories (Normal, Benign, Malignant) with superior test accuracy of 98.18%.
- Conducted a detailed comparative study between VGG16, ResNet50, MobileNetV2, and InceptionV3 to evaluate their strengths and limitations in medical image classification.
- Highlighted VGG16's capability to outperform other models in lung cancer detection, emphasizing its robustness and feature extraction capabilities.
- Employed Gradient-weighted Grad-CAM to provide interpretability of model decisions, enhancing transparency and reliability in clinical applications.
- Identified VGG16 as a promising diagnostic tool for early lung cancer detection, potentially reducing diagnostic delays and improving patient outcomes.
- Proposed practical improvements, such as incorporating larger datasets, 3D CNNs, and data augmentation techniques, to refine model performance and generalizability.

Future research should focus on integrating these advanced architectures and larger datasets to refine diagnostic accuracy further, capture depth-related features, and extend

the study's impact. This work underscores the transformative potential of AI in medical imaging, paving the way for faster and more precise lung cancer detection and improved patient outcomes.

Author Contributions: Conceptualization, R.K. and W.C.; methodology, R.K. and W.C.; software, R.K. and W.C.; validation, R.K., P.P. and T.L.; formal analysis, R.K., P.P. and T.L.; investigation, T.L. and W.C.; resources, R.K., P.P. and T.L.; data curation, R.K., P.P. and T.L.; writing—original draft preparation, R.K., T.L. and W.C; writing—review and editing, R.K., T.L. and W.C; visualization, R.K. and W.C.; supervision, W.C.; project administration, W.C.; funding acquisition, W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Since the data are already in the public domain and anonymized where necessary, the potential risk to individuals or organizations from this study is minimal. Nonetheless, we commit to ethical research practices, and any data revealing sensitive implications will be further anonymized and handled responsibly.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on access from <https://www.kaggle.com/datasets/hamdallak/the-igothnccd-lung-cancer-dataset> (accessed on 7 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Leiter, A.; Veluswamy, R.R.; Wisnivesky, J.P. The global burden of lung cancer: Current status and future trends. *Nat. Rev. Clin. Oncol.* **2023**, *20*, 624–639. [CrossRef] [PubMed]
2. Kapoor, V.; Mittal, A.; Garg, S.; Diwakar, M.; Mishra, A.K.; Singh, P. Lung cancer detection using VGG16 and CNN. In Proceedings of the 2023 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 29–30 July 2023; pp. 758–762. [CrossRef]
3. Xu, H. Comparison of CNN models in non-small lung cancer diagnosis. In Proceedings of the 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 29–31 January 2023; pp. 1169–1174. [CrossRef]
4. Tejaswini, C.; Nagabushanam, P.; Rajasegaran, P.; Johnson, P.R.; Radha, S. CNN architecture for lung cancer detection. In Proceedings of the 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 23–24 April 2022; pp. 346–350. [CrossRef]
5. Aharonu, M.; Kumar, R.L. Convolutional neural network-based framework for automatic lung cancer detection from lung CT images. In Proceedings of the 2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 23–25 December 2022; pp. 1–7. [CrossRef]
6. Huang, H.; Wang, M.; Ye, Q.; Zhou, Z. Diagnosis of lung cancer based on CT scans using convolutional neural networks. In Proceedings of the 2022 International Conference on Data Analytics, Computing and Artificial Intelligence (ICDAI), Zakopane, Poland, 15–16 August 2022; pp. 338–341. [CrossRef]
7. Zargar, H.H.; Zargar, S.H.; Mehri, R.; Tajidini, F. Using VGG16 algorithms for classification of lung cancer in CT scan images. *arXiv* **2023**, arXiv:2305.18367. [CrossRef]
8. Alyasriy, H.; Al-Huseiny, M. The IQ-OTH/NCCD lung cancer dataset [Data set]. *Mendeley Data* **2023**. [CrossRef]
9. Kumaran, S.Y.; Jeya, J.J.; Khan, S.B.; Alzahrani, S.; Alojail, M. Explainable lung cancer classification with ensemble transfer learning of VGG16, ResNet50, and InceptionV3 using grad-cam. *BMC Med. Imaging* **2024**, *24*, 176. [CrossRef]
10. Vaishnavi, D.; Arya, K.S.; DeviAbirami, T.; Kavitha, M. Lung cancer detection using machine learning. *Int. J. Eng. Res. Technol.* **2019**, *7*, 1–6.
11. Xu, H.; Yu, Y.; Chang, J.; Hu, X.; Tian, Z.; Li, O. Precision lung cancer screening from CT scans using a VGG16-based convolutional neural network. *Front. Oncol.* **2024**, *14*, 1424546. [CrossRef] [PubMed]
12. Al-Shouka, T.T.; Alheeti, K.M. A transfer learning for intelligent prediction of lung cancer detection. In Proceedings of the 2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT), Al-Muthana, Iraq, 4–6 July 2023; pp. 54–59. [CrossRef]

13. Bherje, A.; Jidge, A.; Roy, C.; Hulke, A.; Aswathy, M.A.; Yadav, V.; Veenamol, K.V. Design of deep learning-based approach to predict lung cancer on CT scan images. In Proceedings of the 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 15–16 March 2024; pp. 1–5. [\[CrossRef\]](#)
14. Suma, K.V.; Sonali, C.S.; Chinmayi, B.S.; Easa, M. CNN models comparison for lung cancer classification using CT and PET scans. In Proceedings of the 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 16–17 October 2022; pp. 1–5. [\[CrossRef\]](#)
15. Karthikeyan, N.K.; Ali, S.S. Lung cancer classification using CT scan images through deep learning and CNN-based model. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 18–19 April 2024; pp. 1–5. [\[CrossRef\]](#)
16. Gayap, H.T.; Akhloufi, M.A. Deep machine learning for medical diagnosis, application to lung cancer detection: A review. *BioMedInformatics* **2024**, *4*, 236–284. [\[CrossRef\]](#)
17. Mamun, M.; Mahmud, M.I.; Meherin, M.; Abdelgawad, A. LCDtCNN: Lung cancer diagnosis of CT scan images using CNN-based model. In Proceedings of the 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 23–24 March 2023; pp. 205–212. [\[CrossRef\]](#)
18. Mane, D.; Barapate, P.; Khinde, P.; Chavan, A.; Hagare, P. Early lung cancer detection using CNN. In Proceedings of the 2024 4th International Conference on Intelligent Technologies (CONIT), Bangalore, India, 21–23 June 2024; pp. 1–7. [\[CrossRef\]](#)
19. Saravanan, T.M.; Jagadeesan, M.; Pa, S.; Mubarak, M.M.; Kumar, S.P.; Sanjay, S. A CNN-based machine learning scheme to detect lung cancer from CT scan images. In Proceedings of the 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 23–25 January 2023; pp. 1–5. [\[CrossRef\]](#)
20. Al-Areqi, F.; Çelebi, A.T.; Konyar, M.Z. Lung cancer classification using image enhancement and CNN-based pre-trained models. In Proceedings of the 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), Sivas, Türkiye, 11–13 October 2023; pp. 1–7. [\[CrossRef\]](#)
21. Jain, D.; Singh, P.; Pandey, A.K.; Singh, M.; Singh, H.; Singh, A. Lung cancer detection using convolutional neural network. In Proceedings of the 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 11–12 November 2022; pp. 1–4. [\[CrossRef\]](#)
22. Metagar, S.M.; Sayyed, F.R. Lung cancer detection and classification using machine learning algorithm (CNN). In Proceedings of the 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, India, 24–27 July 2024; pp. 1–5. [\[CrossRef\]](#)
23. Mitra, S.; Majumder, A.B.; Saha, T. An observation and analysis of the role of convolutional neural networks toward lung cancer prediction. *Baghdad Sci. J.* **2023**, *20*, 2568. [\[CrossRef\]](#)
24. Nandini, K.; Yasudha, K.; Archana, V. Prediction of benign and malignant tumours in lung cancer using CNN and deep learning. *Int. J. Adv. Technol. Eng. Explor.* **2022**, *10*, 629–633.
25. Singh, L.; Choudhary, H.K.; Singh, S.; Bisht, A.K.; Jain, P.; Shukla, G. Automated detection of lung cancer using transfer learning-based deep learning. In Proceedings of the 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 20–21 May 2022; pp. 500–504. [\[CrossRef\]](#)
26. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Tammima, S. Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *Int. J. Sci. Res. Publ.* **2019**, *9*, 143–150. [\[CrossRef\]](#)
28. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [\[CrossRef\]](#)
29. Wen, L.; Li, X.; Gao, L. A transfer learning approach for intelligent fault diagnosis of machines with unlabeled data. *Neurocomputing* **2020**, *32*, 6111–6124. [\[CrossRef\]](#)
30. Xiaoling, X.; Xu, C.; Nan, B. Transfer learning with InceptionV3 for flower classification. *Int. J. Comput. Vis. Appl.* **2017**, *10*, 67–74. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.