Comparative Evaluation of ResNet50, Xception, and EfficientNetB3 for Lung Cancer Classification and Explainable AI-based Tumor Localization

Abhiram Sharma[1[0009-0004-0822-0603]], R Srivats[1[0009-0001-5791-4255]], Krishna P B[1[0009-0000-1944-0514]], Suganya R[1[0000-0003-1874-6479]], Parvathi R[1[0000-0002-6633-9897]]

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Tamil Nadu, India
parvathi.r@vit.ac.in

Abstract. Accounting for 22% of global medical deaths, with a death rate of 57%, Lung Cancer stands being the primary cause of cancer-related mortality in the globe, as per the World Health Organization. Early detection and classification through imaging has proven to be crucial for improving survival rates. The study aims at training and validating existing state of the art models on our dataset of 1,190 CT Scan images, which includes well distributed cases spanning across normal, benign and malignant cases. Utilizing a novel approach of transfer of learning features between ResNet50 to the EfficientNetB3 architectures, the model was assessed using standard evaluating metrices including accuracy, precision, recall and F1 Score. The methodology has been able to enhance the metrics for the predictive model's overall performance with EfficientNetB3 emerging at the top with superior performance of accuracy 96%, precision 91%, recall 90% and an F1 Score of 92%. To further understand the decisions made by the proposed model, an Explainable AI technique SHAP (SHapely Additive exPlanations) heatmaps was applied to the best-performing model for tumor localization, offering a detailed understanding of decision making process followed by the model. The SHAP visualizations enabled us to identify the most influential features contributing to the classification, highlighting the clinical relevance of this technique. The research signifies the impact of both model selection and the application of Explainable AI for enhancing the interpretability and diagnostic reliability in medical imaging and automated diagnoses.

Keywords: Lung Cancer Classification, Tumor Localization, Deep Learning, SHAP (SHapley Additive Explanations, Medical Imaging Interpretability

1 Introduction

Lung cancer is among the leading causes of tumor related fatalities, contributing to a substantial portion of the overall cancer mortality rate. Despite advancements in medical treatments and technologies, the outlook for Lung cancer patients is frequently poor due to the prevalence of late-stage diagnoses, limiting the effectiveness of treatment options. Early diagnosis is crucial to improving patient outcomes and survival rates, with CT scans being one of the most reliable methods for lung cancer detection. However, the immense volume and complexity of CT scan data create significant challenges for manual interpretation by radiologists, sparking interest in automated analysis techniques for cancer detection and classification.

Over the past few decades, deep learning has grown in importance as a method for medical image analysis. CNNs have shown remarkable success due to their capacity for learning and feature extraction from complex datasets without manual feature engineering, improving the speed and accuracy of lung cancer detection. However, selecting the most appropriate model is crucial for achieving reliable clinical results. This study compares three deep learning models viz. ResNet50, Xception, and EfficientNetB3 on lung cancer classification using CT data, aiming to determine which performs best in terms of accuracy for clinical use.

The deep learning models utilized in the study namely ResNet50, Xception, and EfficientNetB3 is highly relevant for this study as each model offers unique advantages in handling the attributes of medical imaging. ResNet50's ability to extract fine grained features essential for accurate classification of lung cancer. Xception's use of depthwise separable convolutions, and EfficientNetB3's compound scaling ensures high precision with optimized resource.

This study highlights the significance of model interpretability across medical applications, as deep learning models are often seen as black boxes. To address this, Explainable AI techniques, specifically SHapley Additive exPlanations, are applied to EfficientNetB3 to highlight the CT regions contributing most to predictions, aiding radiologists in validating results.

By integrating XAI for enhanced interpretability, this study bridges the gap between model performance and clinical applicability. The findings contribute to the development of more accurate and interpretable models for lung cancer detection, supporting the amalgamation of AI, Artificial Intelligence and related technologies into medical procedures and improving patient care outcomes.

2 Literature Review

Wang et al. [1] investigates the ability of traditional CNN based on deep learning for early and timely detection with proper classification of lung tumor in medical photos. The main evaluation parameters were specificity, sensitivity, and accuracy. The study mentions limitations in generalization due to overfitting on small datasets and potential computational inefficiency.

Lung Cancer Detection Using Transfer Learning Techniques [2] utilized transfer learning with pretrained models like ResNet and VGG to detect lung cancer in CT images. Accuracy, precision, recall, and F1-score were utilized as evaluating criteria. Limited availability of labeled data was cited as a constraint for achieving higher accuracy.

AI-Based Systems for Lung Cancer Classification: A Comparative Study [3] draws an analysis between various machine learning and deep learning approaches for lung cancer classification, focusing on support vector machines and neural networks.

Evaluation metrics included accuracy, F1-score, and area under the curve (AUC). Computational complexity and difficulty in tuning hyperparameters were mentioned as key challenges.

In Hybrid Deep Learning Model for Early Detection of Lung Cancer [4] the authors proposed a hybrid model combining CNN and recurrent neural networks to better the early classification of lung tumor. Model was evaluated using accuracy, sensitivity, and specificity. The hybrid faced limitations in processing large-scale datasets and showed sensitivity to noise in images.

Detterbeck et al. [5] deploys a custom convolutional neural network which was developed for accurately localizing tumors in lung cancer images. Accuracy, intersection over union, and precision metrics were reported. The model's performance degraded when tested on images with significant variations in tumor size and shape.

Ensemble Learning for Lung Cancer Classification [6] employed ensemble learning techniques combining various machine learning models, together with decision trees and gradient boost. Accuracy, precision, recall, and ROC-AUC were used for evaluation. The study noted computational inefficiencies and the challenge of balancing between different classifiers.

Rami-Porta et al. [7] uses a fully automated CNN-based approach for classifying lung cancer from medical images was proposed. The system was evaluated using accuracy, sensitivity, and specificity. The method showed susceptibility to overfitting, especially when trained on small datasets.

Gao et al. [8] talks about a deep learning framework based on U-Net architecture which was used for segmenting lung tumors in CT images. Dice similarity coefficient and accuracy were employed for evaluation. The model struggled with identifying small or irregularly shaped tumors, affecting overall segmentation accuracy.

Detterbeck et al. [9], this paper proposed a hybrid CNN-RNN model to capture both spatial and temporal features for lung cancer classification. The model was checked upon metrices like accuracy, precision, and recall. Integration of CNN in addition with RNN resulted in increased computational requirements, limiting its real-time application.

Zhu et al. [10] Artificial intelligence-assisted techniques, particularly CNNs, were used for both detecting and classifying lung tumors. Recall, precision, accuracy and F1-score was made to be utilized in order to measure the system’s capability and excellence. Paper highlights a lack of generalizability across diverse datasets due to the model's tendency to overfit.

Deep Learning for Multi-Class Lung Cancer Classification [11] adopted a multi-class deep learning approach was proposed for distinguishing between different types of lung cancer. Precision, accuracy, and confusion matrix were used for evaluation. Performance degraded when the dataset was imbalanced, making it less effective for underrepresented classes.

Prasad et al. [12] introduces a CNN-based approach for tumor localization using image segmentation techniques. Dice coefficient and IoU were used to evaluate the model's segmentation performance. The model had difficulty handling tumors with irregular shapes, leading to reduced segmentation accuracy.

3 Methodology

The dataset used in the study is publicly available on Kaggle comprising of 1,190 CT scanning images from more than 110 cases. The cases comprise from 40 malignant, 15 benign, and 55 normal cases collected from male and female patients across diverse age groups and demographics. The scanned images were originally in DICOM format using a Siemens SOMATOM scanner with 1mm slice thickness, the scans were taken at breath hold with full inspiration and split into a train dataset and test dataset in the ratio of 70-30. Institutional review board of the collaborating medical centers gave their approval to the study, with written consent waived by the oversight review board. Preprocessing steps included image normalization and augmentation to improve generalization and model robustness.

In this section we present the comparison oh three state-of-the-art deep learning models namely ResNet50, Xception, and EfficientNetB3 for the CT scan-based categorization of lung cancer. Each model brings unique strengths and architectural innovations, making them well-suited for medical image analysis.

ResNet50 is an advanced convolutional neural network identified for introducing residual learning, using skip connections to bypass certain layers during training. This helps tackle the vanishing gradient problem, which often affects deep networks, allowing ResNet50 to train effectively without performance loss. With 50 layers, it has proven highly successful in image classification tasks, making it a strong contender for lung cancer classification, where complex medical images require advanced feature extraction.

Xception improves upon the Inception architecture by using depth-wise separable convolutions. This splits the convolution process into depth-wise filtering and pointwise combination, making the model more efficient in capturing spatial hierarchies. Xception excels at detecting tumors that vary in size and location in lung CT scans, offering both fine detail and computational efficiency for large-scale medical imaging tasks.

EfficientNetB3 belongs to the EfficientNet family, which introduces compound scaling to optimize depth, width, and resolution of the network. This method enhances accuracy while maintaining computational efficiency, making it ideal for complex tasks like lung cancer detection. EfficientNetB3's balanced approach allows it to perform well on large datasets, offering high accuracy without excessive resource use, a key factor for medical diagnostics.

The architecture used in this study, as shown in Figure 1, integrates ResNet50 and EfficientNetB3 for lung cancer classification, with custom layers tailored to the task. The process begins with image pre-processing and augmentation to improve model robustness. ResNet50, serving as the initial feature extractor, captures general image features using residual connections for efficient learning. The output is then flattened into a vector of single dimension and fed into EfficientNetB3, which extracts more intricate patterns specific to lung cancer. A copiously connected dense layer having 128 ReLU-activated units learns complex feature interactions, trailed by a secondary dense layer of 3 units to reduce feature dimensions. The final classification layer uses softmax activation for categorizing images. SHAP (SHapley Additive exPlanations) is employed for model interpretability, highlighting important features to explain predictions. This architecture combines strong feature extraction, efficient processing, and accurate classification.

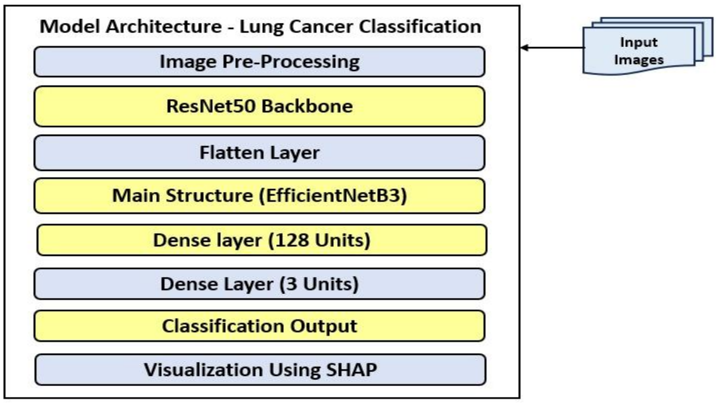

Fig. 1. Model Architecture – Lung Cancer Classification

Metrics like recall, precision, F1 Score and accuracy were utilized to evaluate the performance of the model. These measurements offer a thorough understanding of classification abilities, particularly critical in medical imaging to minimize false positives and negatives. The use of these metrics ensures that the model selected for clinical deployment achieves high accuracy while minimizing diagnostic errors, enhancing the safety and reliability of AI incorporated lung cancer detection.

4 Model Analysis and Results

This study, assessed how well a number of deep learning models performed when classifying lung cancer from CT scan pictures, including ResNet50, Xception and EfficientNetB3. The performance of each model was measured by applying the key metrices namely, recall, precision, accuracy and F1-Score as described below under Table 1. We can see a pictorial analysis of the same in Fig.2.

Model	Accuracy	Precision	Recall	F1-Score
EfficientNetB3	0.96	0.91	0.90	0.92
Xception	0.91	0.89	0.85	0.88
ResNet50	0.87	0.83	0.88	0.85

Table 1. Comparison of Model Performance

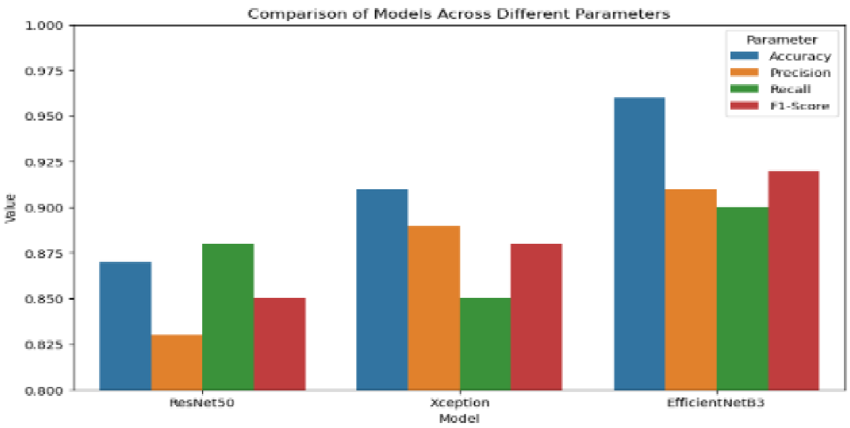

Fig. 2. Comparison of Models across Different Parameters

EfficientNetB3 emerged as the best-performing model with an accuracy of 96%, significantly outperforming the other models in terms of precision of 91%, with recall of 90% and F1-Score of 92%. It is more evident in the model comparison heatmaps and stick plot for all the models trained in Figure 3 and 4.

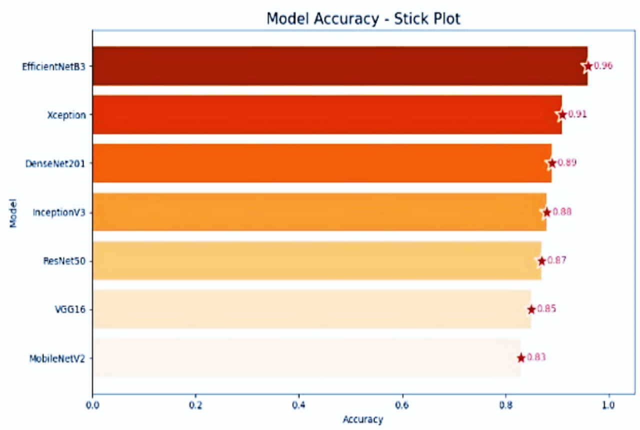

Fig. 3. Model Accuracy – Stick Plot

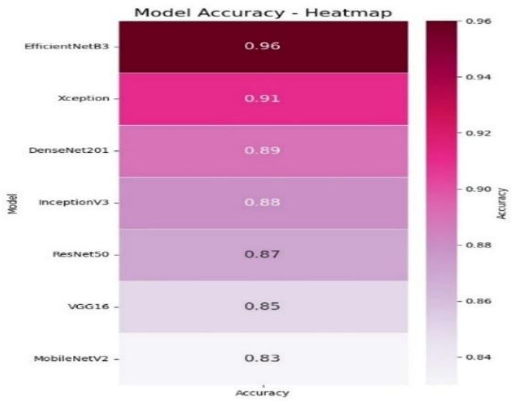

Fig. 4. Model Accuracy – Heatmap

Figure 5 and 6 below illustrates the variation in training accuracy and observed loss curve of the model, showcasing how EfficientNetB3 consistently outperforms others in terms of convergence and stability. The training curve for EfficientNetB3 shows a steady rise with minimal overfitting, further solidifying its position as the best model for this task.

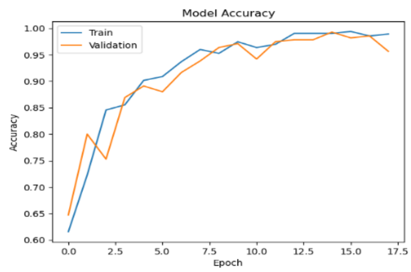

Fig. 5. EfficientNetB3's Training and Validation Accuracy Curves

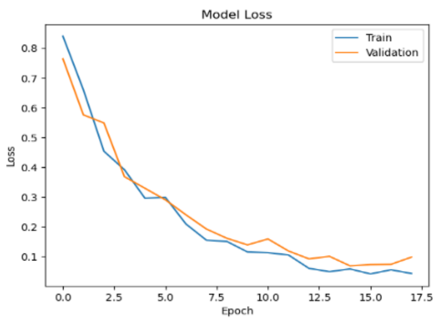

Fig. 6. EfficientNetB3’s Training and Validation Loss Curves

The confusion matrix, depicted in Fig. 7. further supports the dominance of EfficientNetB3 in this study. It shows that the model achieved near-perfect classification in the majority of cases, with very few instances of misclassification. The most common error occurred in distinguishing between the early and moderate stages of lung cancer, where the differences in tissue characteristics may be more subtle.

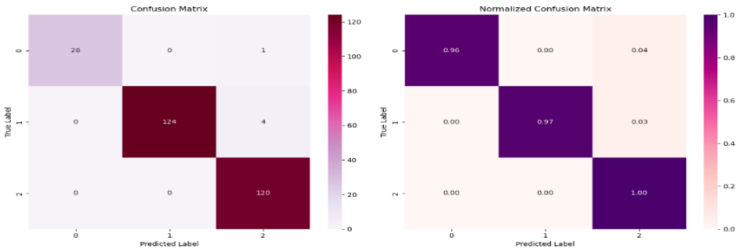

Fig. 7. Confusion Matrix for EfficientNetB3

The Radial Graph Analysis in Figure 8 for the different models provided a comprehensive view of the compromise among sensitivity and specificity for each CNN model. EfficientNetB3 had the largest polygonal area, further solidifying its ability to distinguish between different lung cancer stages. The other models, while still performing well, all fell short of this benchmark.

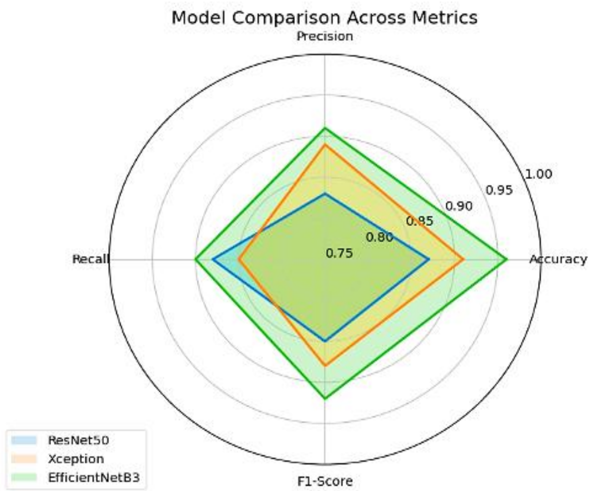

Fig. 8. Radial Graph comparison of the Models across evaluation Metrics

The outcomes obtained during this study illustrate the prospective of deep learning, particularly EfficientNetB3, for use in clinical settings to assist detecting lung cancer. Given the higher accuracy and recall, this model could play a critical role in supporting radiologists by providing a second opinion, especially in cases where early-stage cancer detection is paramount. The precision of the model also reduces the likelihood of false positives, ensuring that unnecessary treatments or invasive procedures are minimized.

However, it is important to note that while EfficientNetB3 shows great promise, further validation on more vast and varied collections of datasets is pivotal to ensure its robustness over different demographical regions and medical conditions. Additionally, the integration of Explainable AI methods could enhance the model's interpretability, providing clinicians with clear explanations for its class determination is essential for increasing the rate of trust growth and adoption in real-world medical practice.

5 Explainable AI Technique Application

SHAP technique has been used to enhance the interpretability of deep learning models. These techniques give information about the model's most contributing decision-making parameters allowing a deeper understanding of model behavior. This is essential in pivotal environments like healthcare.

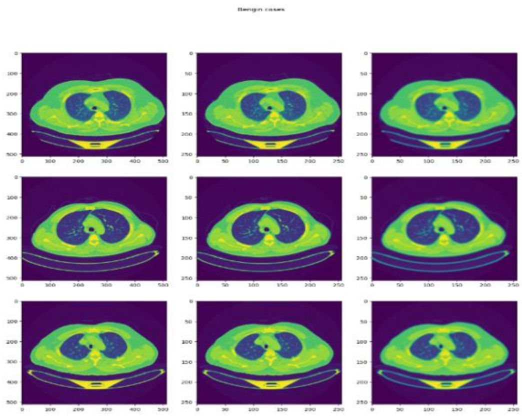

Fig. 9. Benign cases

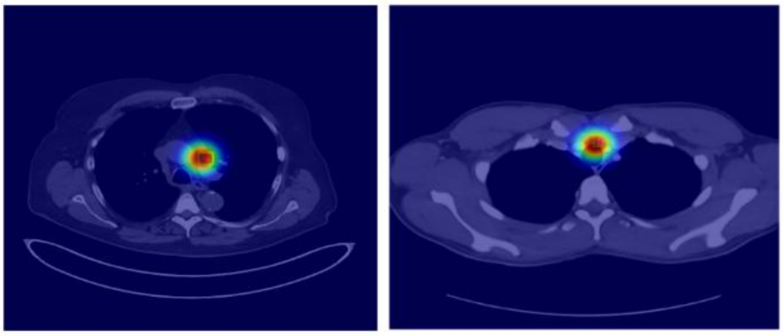

Fig. 10. a. SHAP-benign **b.** SHAP-benign

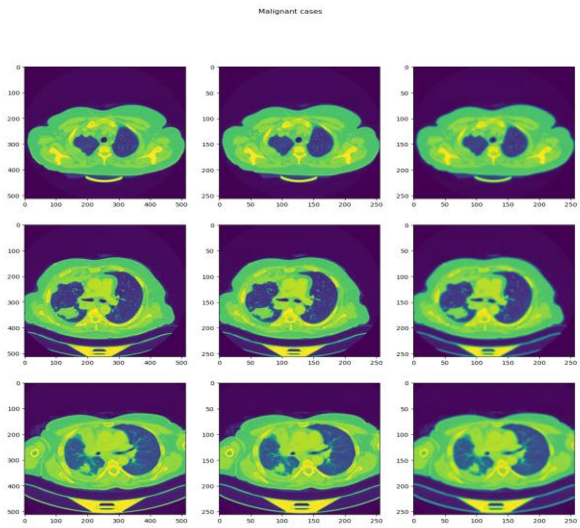

Fig. 11. Malignant cases

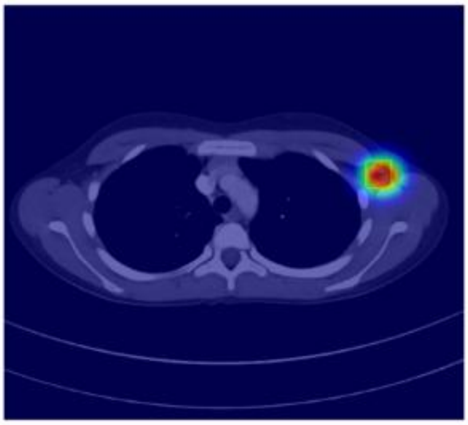

Fig. 12. SHAP-Malignant

We used three types of lung CT scans, as seen in the confusion matrix: healthy lungs and two varieties of infected lungs. Figures 9 and 11 show 3x3 grid visualizations of benign and malignant lung cases for cancer classification. In the benign scans, the lung tissues appear uniform, with smooth color transitions representing normal tissue density and healthy lung aeration. The green and yellow shades indicate well-functioning lung tissue without abnormal growth, with no signs of tumors.

In contrast, the malignant scans display complex patterns with patches indicating abnormal tissue density due to tumor growth. These irregular shapes and color variations highlight disruptions in the lung tissue, making them critical for the model to detect. The stark differences in intensity and structure serve as key indicators of malignancy, essential for tumor localization and classification.

The distinction between benign and malignant cases lies in the color variation and structural anomalies. These datasets are crucial for training the model, with benign cases teaching the appearance of healthy lungs and malignant cases helping the model detect and identify tumors, improving cancer prediction and localization.
We selected SHAP for its ability to provide both local and global explanations, offering deeper insights into overall model behavior and versatility across different models compared to other XAI algorithms. The provided images are SHAP generated heatmaps overlaid on the CT scan images, illustrating the regions that are most influential to the model in lung cancer classification. SHAP improves interpretability by highlighting scan areas with the highest impact on prediction of model, with colors ranging from red (high contribution) to blue (low contribution).

In Figure 10. a., the SHAP heatmap highlights a central region near the trachea in red, indicating this area is crucial for the model's lung cancer classification. The red-to-yellow gradient shows strong model attention on this area, suggesting it has detected a potential mass or abnormality. Blue regions, representing minimal contribution, indicate the model deems them irrelevant to the diagnosis.

Figure 10. b. focuses on the mediastinal region, marked in red, showing the model's strong attention here. The concentration suggests possible lymph node involvement or mass near the heart, which are key indicators in lung cancer diagnoses. The red focus confirms the model's confidence in this region's significance for classification.

Figure 12. shifts attention to the periphery, with the red highlight near the rib cage. This suggests the model is detecting a peripheral lesion or tumor, which could otherwise be missed. The red highlight demonstrates the model's ability to classify cancer from various lung areas.

The SHAP heatmaps offer transparency into the model's decision-making, providing insight into areas of medical concern. The yellow hollow squares, representing doctor-identified tumors, serve as benchmarks, affirming the model's focus on clinically significant regions and reinforcing its diagnostic accuracy.

6 Conclusion

The proposed EfficientNetB3 model for lung cancer classification significantly outperforms individual models like ResNet50 and Xception, achieving an accuracy of 96 percent, compared to ResNet50 87 percent and Xception 91 percent. By leveraging compound scaling, EfficientNetB3 optimizes feature extraction, leading to superior classification performance. The model also demonstrates strong precision and recall values of 91 percent and 90 percent, respectively, with an F1 score of 92 percent, signifying a coordinated functioning in precision and recall. Confusion matrix reveals minimal misclassification, particularly in distinguishing early and moderate stages of lung cancer, further supporting the model's robustness.

These results validate the model's potential in improving lung cancer detection, making it a reliable tool for supporting clinical decisions. Future work will aim to enhance the generalizability of the model by incorporating more varied datasets as well as exploring integration into clinical workflows in real time, supporting early diagnosis and intervention. Additionally, further research into explainability techniques, such as SHAP, will enhance transparency in the decision making processes, increasing trust in AI-driven diagnostics for lung cancer.

References

1. X. Wang *et al.*, "Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3950–3962, Sep. 2020, doi: 10.1109/TCYB.2019.2935141.
2. "Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images | IEEE Journals & Magazine | IEEE Xplore." Accessed: Sep. 19, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9395071
3. "Self-Upgraded Cat Mouse Optimizer With Machine Learning Driven Lung Cancer Classification on Computed Tomography Imaging | IEEE Journals & Magazine | IEEE Xplore." Accessed: Sep. 19, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10244209
4. "Enhancing Lung Cancer Classification and Prediction With Deep Learning and Multi-Omics Data | IEEE Journals & Magazine | IEEE Xplore." Accessed: Sep. 19, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10508786
5. F. C. Detterbeck *et al.*, "The Proposed 9th Edition TNM Classification of Lung Cancer," *CHEST*, Jun. 2024, doi: 10.1016/j.chest.2024.05.026.
6. "Empowering Diagnosis: Cutting-Edge Segmentation and Classification in Lung Cancer Analysis," *Computers, Materials and Continua*, vol. 79, no. 3, pp. 4963–4977, Jun. 2024, doi: 10.32604/cmc.2024.050204.
7. R. Rami-Porta *et al.*, *Journal of Thoracic Oncology*, vol. 19, no. 7, pp. 1007–1027, Jul. 2024, doi: 10.1016/j.jtho.2024.02.011.
8. Z. Gao *et al.*, "Seeking multi-view commonality and peculiarity: A novel decoupling method for lung cancer subtype classification," *Expert Systems with Applications*, p. 125397, Sep. 2024, doi: 10.1016/j.eswa.2024.125397.
9. F. C. Detterbeck *et al.*, *Journal of Thoracic Oncology*, vol. 19, no. 7, pp. 1052–1072, Jul. 2024, doi: 10.1016/j.jtho.2024.03.021.
10. W. Zhu *et al.*, "Classification of lung cancer subtypes on CT images with synthetic pathological priors," *Medical Image Analysis*, vol. 95, p. 103199, Jul. 2024, doi: 10.1016/j.media.2024.103199.
11. N. Malligeswari and G. Kavya, "Classification of lung cancer with deep learning Res-U-Net and molecular imaging," *SIViP*, vol. 18, no. 1, pp. 325–333, Feb. 2024, doi: 10.1007/s11760-023-02635-7.
12. U. Prasad, S. Chakravarty, and G. Mahto, "Lung cancer detection and classification using deep neural network based on hybrid metaheuristic algorithm," *Soft Comput*, vol. 28, no. 15, pp. 8579–8602, Aug. 2024, doi: 10.1007/s00500-023-08845-y