

An Interpretable Deep Learning Framework for Preoperative Classification of Lung Adenocarcinoma on CT Scans: Advancing Surgical Decision Support

Ann. Ital. Chir., 2025 96, 9: 1206–1217
<https://doi.org/10.62713/aic.4239>

Qiang Shi¹, Yufeng Liao², Jie Li¹, Hongbo Huang¹

¹Department of Thoracic Surgery, Ningbo No.2 Hospital, 315099 Ningbo, Zhejiang, China

²Department of Clinical Laboratory, Ningbo No.2 Hospital, 315099 Ningbo, Zhejiang, China

AIM: Lung adenocarcinoma remains a leading cause of cancer-related mortality, and the diagnostic performance of computed tomography (CT) is limited when dependent solely on human interpretation. This study aimed to develop and evaluate an interpretable deep learning framework using an attention-enhanced Squeeze-and-Excitation Residual Network (SE-ResNet) to improve automated classification of lung adenocarcinoma from thoracic CT images. Furthermore, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to enhance model interpretability and assist in the visual localization of tumor regions.

METHODS: A total of 3800 chest CT axial slices were collected from 380 subjects (190 patients with lung adenocarcinoma and 190 controls, with 10 slices extracted from each case). This dataset was used to train and evaluate the baseline ResNet50 model as well as the proposed SE-ResNet50 model. Performance was compared using accuracy, Area Under the Curve (AUC), precision, recall, and F1-score. Grad-CAM visualizations were generated to assess the alignment between the model's attention and radiologically confirmed tumor locations.

RESULTS: The SE-ResNet model achieved a classification accuracy of 94% and an AUC of 0.941, significantly outperforming the baseline ResNet50, which had an 85% accuracy and an AUC of 0.854. Grad-CAM heatmaps produced from the SE-ResNet demonstrated superior localization of tumor-relevant regions, confirming the enhanced focus provided by the attention mechanism.

CONCLUSIONS: The proposed SE-ResNet framework delivers high accuracy and interpretability in classifying lung adenocarcinoma from CT images. It shows considerable potential as a decision-support tool to assist radiologists in diagnosis and may serve as a valuable clinical tool with further validation.

Keywords: lung adenocarcinoma; deep learning; computed tomography; interpretability; medical image classification

Introduction

Lung adenocarcinoma, the most prevalent histological subtype of non-small cell lung cancer (NSCLC), remains a leading cause of cancer-related morbidity and mortality worldwide [1,2]. In the United States, lung cancer continues to be the leading cause of cancer deaths. However, screening high-risk individuals with low-dose chest computed tomography has been shown to reduce lung cancer mortality by approximately 20% [3]. The critical role of early and accurate diagnosis in improving patient outcomes is well-established, as it facilitates timely surgical or therapeutic intervention, significantly improving 5-year survival rates and reducing the risk of metastasis [4,5].

Computed tomography (CT), particularly low-dose CT (LDCT), has become the standard modality for thoracic imaging and is central to lung cancer screening programs

that have demonstrated a survival benefit [6,7]. Despite its high spatial resolution and rapid acquisition, CT for indeterminate pulmonary nodules remains challenging when dependent solely on visual assessment [8]. Significant inter-reader variability among radiologists and high false-positive rates in screening settings may result in unnecessary anxiety and invasive procedures [9,10]. Consequently, integrating artificial intelligence (AI) and deep learning techniques into the radiological workflows has emerged as a promising strategy to improve diagnostic accuracy, consistency, and efficiency [11].

Convolutional Neural Networks (CNNs) have achieved remarkable success across various medical imaging applications, including diabetic retinopathy in ophthalmology, skin lesion classification in dermatology, and brain tumor segmentation from Magnetic Resonance Imaging (MRI) scans in neuro-oncology [12,13]. These applications encompass tumor classification, segmentation, and prognostic modeling. Among these, ResNet (Residual Network) architectures are particularly valued for their potential to train very deep models without suffering from vanishing gradients, achieved through residual skip connections [14]. However, standard CNN models, including ResNet50, may

Submitted: 27 June 2025 Revised: 16 July 2025 Accepted: 8 August 2025 Published: 10 September 2025

Correspondence to: Hongbo Huang, Department of Thoracic Surgery, Ningbo No.2 Hospital, 315099 Ningbo, Zhejiang, China (e-mail: Hhb20250609@163.com).

have limited capacity to capture inter-channel feature dependencies, potentially overlooking subtle but discriminative textural and morphological cues needed to differentiate benign and malignant lung tissues [15]. To address this limitation, recent advances in attention mechanisms have introduced modules that enhance the model's focus on the most informative features. The Squeeze-and-Excitation (SE) block, which adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies among channels, is a prime example of such mechanisms [16]. Studies have confirmed that incorporating SE blocks or similar channel attention modules, such as the Convolutional Block Attention Module (CBAM), into existing CNN architectures can selectively amplify relevant features while suppressing non-informative ones, enhancing representational power and diagnostic performance [17,18].

In this study, we proposed an attention-enhanced ResNet framework embedding SE blocks into a ResNet50 backbone, for automated classification of lung adenocarcinoma from CT images. Recognizing the imperative for transparency and trust in clinical AI, we further employed eXplainable AI (XAI) methods [19,20]. Specifically, we incorporated Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize the discriminative regions utilized by the model, a widely validated approach for highlighting diagnostically relevant features [21]. We compared the performance of the baseline ResNet50 with the proposed Squeeze-and-Excitation Residual Network (SE-ResNet) across multiple evaluation metrics, including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), demonstrating the superior diagnostic performance and clinical interpretability of our framework.

Methods

In this study, we developed a deep learning pipeline based on an attention-enhanced residual network to classify lung adenocarcinoma from CT images. The overall methodology included data acquisition and preprocessing, model architecture design, training configuration, evaluation protocols, and explainability analysis using Grad-CAM.

Data Collection and Preprocessing

The study population was sourced retrospectively and consisted of 380 subjects in total, comprising a balanced cohort of 190 patients with pathologically confirmed lung adenocarcinoma and 190 non-cancerous control subjects. For each subject, a standardized set of 10 consecutive 1-mm axial slices was selected for analysis, resulting in 3800 slices used in this study. In the adenocarcinoma cohort, this 10-slice set was centered on the axial slice displaying the largest tumor diameter. For the non-cancerous cohort, a representative slice from the lung parenchyma was used as the center for the 10-slice set. This entire dataset was partitioned at the subject level to prevent data leakage, creating a training set (n = 266 subjects; 133 adenocarcinoma, 133

control), a validation set (n = 38 subjects; 19 adenocarcinoma, 19 control), and a test set (n = 76 subjects; 38 adenocarcinoma, 38 control). The complete process from patient screening to dataset construction is outlined in Fig. 1.

All images were acquired using a 128-slice Siemens SOMATOM Definition AS+ scanner (Siemens Healthineers, Erlangen, Germany) with a standard chest CT protocol (tube voltage = 120 kVp, tube current modulation = 100–250 mAs, slice thickness = 1.0 mm, in-plane resolution = 0.625 mm, and matrix size = 512×512). Ethical approval was obtained from the Ethics Committee of Ningbo No. 2 Hospital (Approval No. [YJ-NBEY-KY-2021-087-01]), and all data were anonymized in compliance with data protection regulations. The establishment of Ground-truth labels followed a rigorous, multi-stage process. All tumor regions were independently delineated by two senior thoracic radiologists with 10 and 12 years of experience, respectively. In cases of disagreement, a final consensus was reached through adjudication by a third expert, a senior thoracic radiologist with over 20 years of experience and specialized expertise in pulmonary oncology. The independent delineations of the two radiologists demonstrated a high degree of concordance, with a mean Intersection over Union (IoU) of 0.82 across all malignant cases.

The diagnostic reference standard was strictly defined: all malignant cases (adenocarcinoma) were unequivocally confirmed by histopathological analysis of surgical or biopsy specimens, while all negative cases were verified through longitudinal imaging follow-up for at least two years, demonstrating either nodule stability or complete resolution. To ensure consistency across deep learning models, all images were resized to 224×224 pixels and normalized to a [0, 1] intensity range using min-max normalization. Notably, the full, uncropped axial slices were used as input for the models. This approach was intentionally chosen over a segmentation-based region-of-interest (ROI) pipeline to evaluate the end-to-end performance of the model and specifically challenge the attention mechanism's capacity to identify relevant pathological features within a complex anatomical background.

Given the relatively limited dataset size, data augmentation techniques were applied to improve the generalization capability of the model. These included random rotations within $\pm 15^\circ$, horizontal and vertical flipping, Gaussian noise injection with a standard deviation range of 0.01–0.05, and contrast-limited adaptive histogram equalization (CLAHE). The dataset was divided by patient level into a training set (70%, n = 266), validation set (10%, n = 38) and test set (20%, n = 76). One split was performed using a fixed random seed to ensure subject independence and prevent data leakage.

Baseline Model: ResNet50

As a baseline, we adopted the ResNet50 architecture, a widely used convolutional neural network recognized for

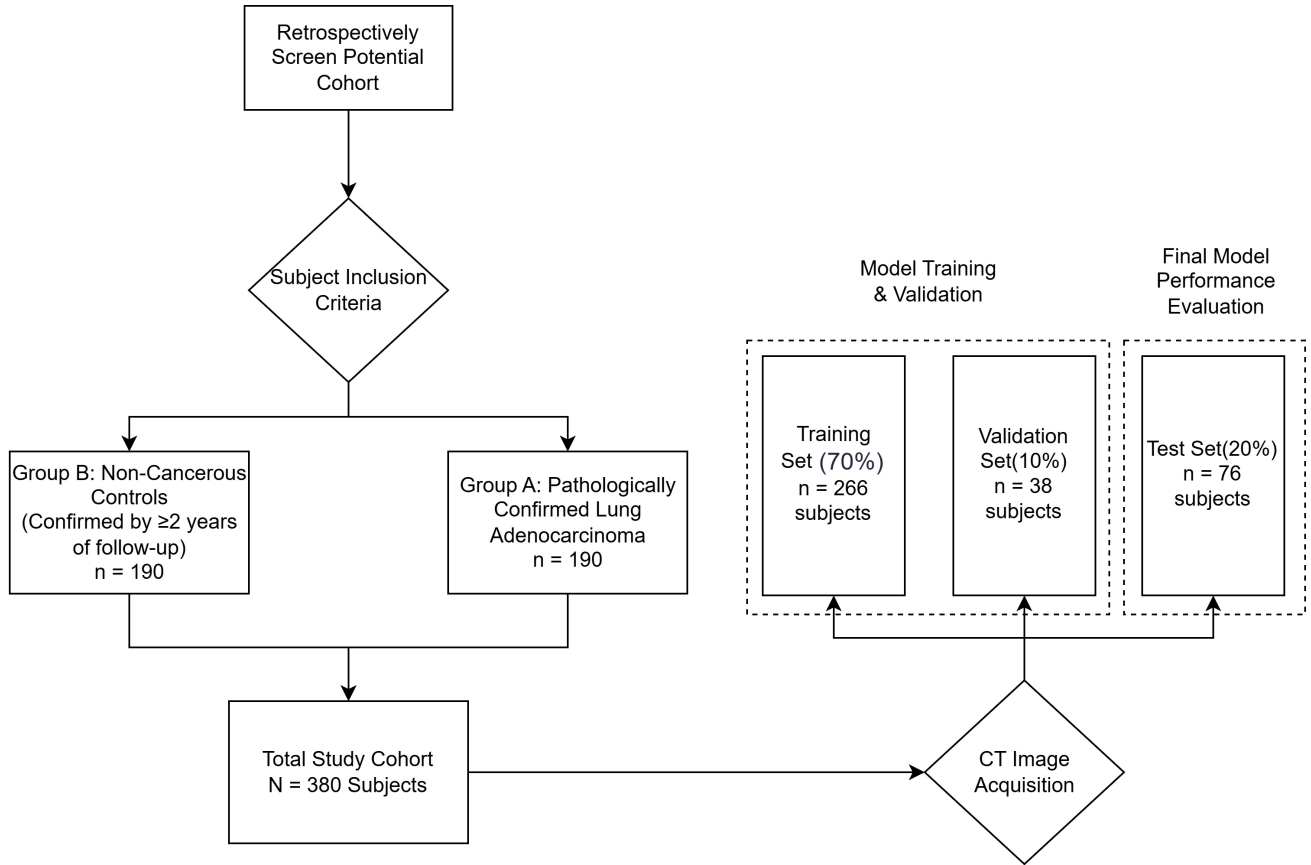


Fig. 1. Flowchart of research data acquisition and cohort construction. This diagram illustrates the process from retrospective subject screening to the final construction of training, validation, and test datasets, with the number of subjects indicated at each stage. CT, computed tomography.

its residual connections that mitigate the vanishing gradient challenge in deep networks. The model was initialized with ImageNet-pretrained weights and fine-tuned on our dataset. The network was modified to accept single-channel grayscale CT inputs, and the final fully connected layer was adapted to produce binary outputs corresponding to adenocarcinoma or normal tissue. Model training was conducted using the PyTorch framework. The binary cross-entropy loss function was minimized with the Adam optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$). The initial learning rate was set to 1×10^{-4} and reduced by a factor of 10 when the validation loss plateaued for more than five consecutive epochs. Each model was trained for up to 50 epochs with early stopping based on validation performance, using a batch size of 32.

Attention-Enhanced Model: SE-ResNet50

To improve the capacity of the model to focus on channel-wise salient features, we incorporated Squeeze-and-Excitation (SE) blocks into the ResNet50 architecture. These blocks were inserted after each residual unit to adaptively recalibrate feature responses. The SE module operates in three sequential steps: squeeze, excitation, and scale. In the squeeze step, a global average pooling operation is applied across the spatial dimensions of the feature maps

to generate a compact channel descriptor. Given a feature map $X \in \mathbb{R}^{(C \times H \times W)}$, the channel-wise descriptor z is computed as:

$$z_c = \frac{1}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W X_{c,i,j}$$

This descriptor is then passed through two fully connected layers with ReLU and sigmoid activations, forming the excitation operation:

$$s = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot z))$$

Where r is the reduction ratio, set to 16 in our implementation. Finally, in the scale step, the original feature maps were reweighted using the learned attention vectors:

$$\tilde{X}_c = s_c \cdot X_c$$

This operation enabled the model to emphasize the most informative channels while suppressing less relevant ones, thus improving representational efficiency without increasing spatial complexity. The overall structure of the pro-

posed SE-ResNet model is illustrated in Fig. 2. The model received preprocessed thoracic CT slices as input, processed them through a sequence of convolutional and residual blocks, and integrated SE blocks after key stages to enable adaptive channel-wise recalibration. These modules highlight features critical for distinguishing between malignant and normal tissues. The extracted high-level representations were then passed through a fully connected layer and a final softmax classifier to yield the probability of lung adenocarcinoma versus normal tissue. This architecture maintains a streamlined and modular design, facilitating both predictive accuracy and interpretability.

Evaluation Metrics and Training Procedure

To comprehensively evaluate model performance, we computed accuracy, precision, recall (sensitivity), F1-score, and the area under the receiver operating characteristic curve (ROC curve). Confusion matrices were also generated to visualize classification outcomes in terms of true positives, true negatives, false positives, and false negatives. All experiments were conducted using an NVIDIA RTX 3090 GPU, and each model was trained three times to ensure statistical robustness. Final metrics were reported as averages over these runs. The 95% confidence intervals for AUC values were estimated using the percentile bootstrap method with 2000 resamples.

Model Interpretability With Grad-CAM

To enhance the interpretability of the model and provide visual insight into its decision-making process, we applied Gradient-weighted Class Activation Mapping (Grad-CAM). This method highlights regions within the input image that contribute most to the predicted class by computing the gradient of the class score with respect to feature maps in the final convolutional layer. The resulting heatmaps were superimposed onto the original CT slices to identify regions of diagnostic significance and evaluate whether the attention of the model aligned with known pathological areas.

Reproducibility and Code Availability

The source code and trained model weights supporting the findings of this study are available from the corresponding author upon reasonable request and are subject to a material transfer agreement. Clinical data cannot be publicly shared due to patient privacy restrictions imposed by the institutional review board.

To ensure methodological transparency and facilitate replication, we provide the following details of our computational environment. Hardware: All models were trained and evaluated on an NVIDIA RTX 3090 GPU. Software: Experiments were conducted using Python (v3.8.10, Python Software Foundation, Beaverton, OR, USA), PyTorch (v1.10.1, Meta AI, Menlo Park, CA, USA), CUDA (v11.3, NVIDIA Corporation, Santa Clara, CA, USA), Albumentations (v1.1.0, open-source library developed

by Albumentations Team including Alexander Buslaev *et al.*), NumPy (v1.21.2, open-source library maintained by NumPy Steering Council, associated with NumFOCUS, Austin, TX, USA), and Scikit-learn (v1.0.2, open-source library maintained by scikit-learn community, associated with Inria, Saclay, France), and imgaug (v0.4.0, open-source library developed by Alexander Jung). Random Seeds: To ensure deterministic results, a global random seed of 42 was set for PyTorch, NumPy, and Python's random module at the beginning of each experiment.

Results

To evaluate the effectiveness of the proposed SE-ResNet architecture in classifying lung adenocarcinoma from CT scans, we conducted a series of experiments comparing its performance to a baseline ResNet50 model. The results demonstrate that incorporating Squeeze-and-Excitation (SE) blocks significantly improves classification accuracy and robustness across multiple evaluation metrics. Fig. 3 illustrates the receiver operating characteristic (ROC) curves for both models on the test set. The baseline ResNet50 achieved an Area Under the Curve (AUC) of 0.854 (95% CI, 0.825–0.881), whereas the SE-ResNet model exhibited a markedly improved and statistically distinct AUC of 0.941 (95% CI, 0.920–0.962). These findings indicate that attention-enhanced architecture has a superior discriminative capability for distinguishing malignant tissue from normal tissue.

For a comprehensive quantitative evaluation, the performance of both the baseline ResNet50 and the proposed SE-ResNet50 was assessed across the training, validation, and test sets, with detailed metrics presented in Table 1. This multi-set analysis was crucial for evaluating not only final model efficacy but also generalization capability. The test set results reveal a clear performance advantage for the proposed model: the baseline ResNet50 achieved an overall accuracy of 85% and an AUC of 0.854, whereas the SE-ResNet model achieved a significantly higher accuracy of 94% and an AUC of 0.941. These consistent improvements across all test metrics confirm the benefit of channel-wise attention modeling.

Beyond these final performance scores, the comprehensive data in Table 1 provide critical insights into model generalization. The baseline ResNet50 exhibited a significant performance drop from a training accuracy of 0.96 to a test accuracy of 0.85, indicating a notable degree of overfitting. In contrast, the SE-ResNet50 model showed a much smaller performance gap, declining from a training accuracy of 0.99 to a test accuracy of 0.94. This markedly superior generalization underscores the effectiveness of the SE blocks as a form of architectural regularization, enabling the model to learn more robust and transferable features while reducing the risk of overfitting.

To provide a detailed view of classification errors, confusion matrices for both models on the test subset ($n = 760$)

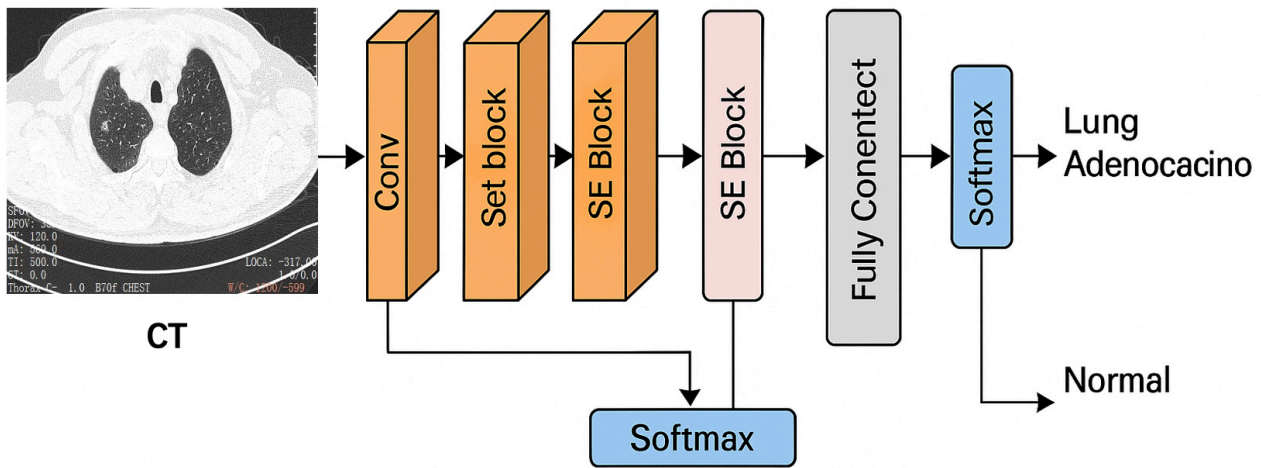


Fig. 2. Schematic architecture of the proposed SE-ResNet model for lung adenocarcinoma classification from thoracic CT scans. The model integrates Squeeze-and-Excitation (SE) blocks into the ResNet50 backbone, enabling adaptive feature recalibration to enhance discriminative performance. SE-ResNet, Squeeze-and-Excitation Residual Network. Conv, Convolutional.

are presented in Figs. 4,5 (n indicates the number of CT slices, not the number of patients). The baseline ResNet50 correctly identified 313 negative samples and 319 positive samples, while misclassifying 67 negatives as positives and 61 positives as negatives. In contrast, the SE-ResNet model demonstrated improved error control, correctly identifying 350 negatives and 348 positives, with only 30 false positives and 32 false negatives. This improvement in misclassification distribution highlights the enhanced sensitivity and specificity of the proposed method.

In addition to accuracy-related metrics, we examined the distribution of model prediction confidence scores across all test samples (Fig. 6). The histogram illustrates that the SE-ResNet model tends to produce more polarized confidence values, with higher certainty for both positive and negative predictions compared to the baseline ResNet50. This behavior indicates that the SE-enhanced model achieves better class separability and generates more decisive predictions, which is particularly valuable in clinical decision-making scenarios.

To further assess the interpretability of the proposed SE-ResNet model, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize spatial regions that contributed most significantly to the predictions of the model. To provide an intuitive, real-world demonstration of the utility of our model, we presented a representative clinical case from the test set in Fig. 7. Fig. 7A shows the original axial CT slice containing a lung adenocarcinoma, with the ground-truth tumor region delineated by an expert radiologist. When processed by the baseline ResNet50 (Fig. 7B), the resulting Grad-CAM heatmap was diffuse and poorly localized. In contrast, when processed by the proposed SE-ResNet model (Fig. 7C), the model correctly classified the lesion as adenocarcinoma and produced a highly focused activation map that precisely highlighted the tumor,

demonstrating strong alignment with expert interpretation and underscoring its potential as a clinical decision-support tool.

As illustrated, the baseline ResNet50 frequently generated diffuse and nonspecific activation maps, often highlighting broad areas extending beyond tumor boundaries or emphasizing unrelated anatomical structures such as the chest wall or mediastinum. Conversely, the SE-ResNet model consistently produced more compact and sharply localized attention regions that closely matched the radiologist-identified lesions. Notably, in correctly classified adenocarcinoma cases, SE-ResNet Grad-CAM maps exhibited high spatial concordance with annotated tumor areas, effectively capturing peripherally located nodules, spiculated margins, and heterogeneous intensities, features of high clinical relevance for lung adenocarcinoma diagnosis.

To quantify the alignment between model attention and radiological ground truth, we computed the average Intersection over Union (IoU) between Grad-CAM heatmaps and expert-annotated tumor regions (Fig. 8). The SE-ResNet achieved a higher average IoU of 0.75 compared to 0.68 for the baseline ResNet50. This quantitative result further supports that channel-wise attention improves not only classification performance but also the spatial reliability of model explanations.

Overall, these findings indicate that the SE-ResNet model enhances the extraction of salient diagnostic features from CT images while providing more reliable and interpretable predictions suitable for clinical decision support.

Discussion

In this study, we developed and evaluated an attention-enhanced convolutional neural network, SE-ResNet, for automated classification of lung adenocarcinoma from tho-

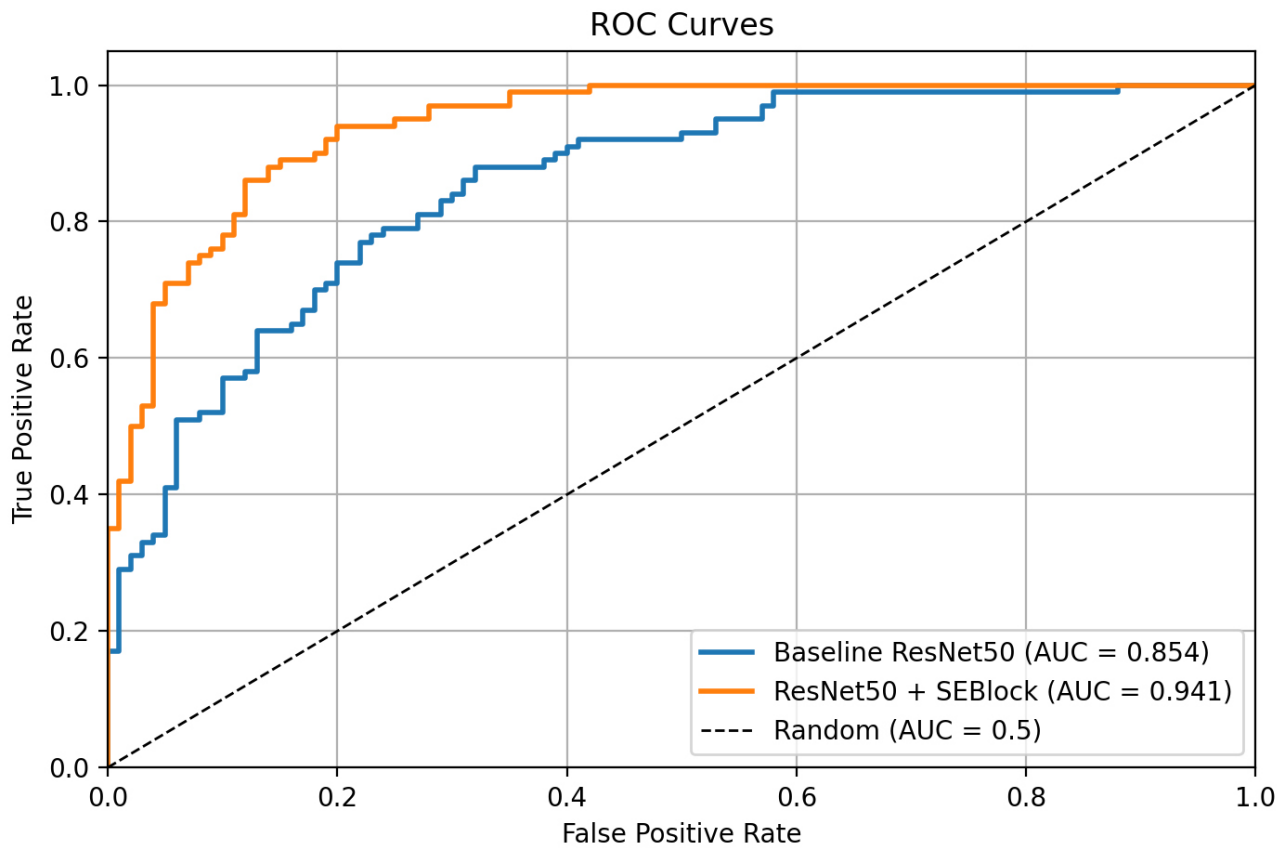


Fig. 3. Receiver operating characteristic (ROC) curves comparing baseline ResNet50 and SE-ResNet models on the test set. The baseline ResNet50 achieved an Area Under the Curve (AUC) of 0.854 (95% CI: 0.825–0.881), while the SE-ResNet achieved an AUC of 0.941 (95% CI: 0.920–0.962). Confidence intervals were estimated using 2000 bootstrap resamples.

Table 1. Comparative performance of baseline ResNet50 and attention-enhanced SE-ResNet50 across training and test sets.

Model	Data set	Accuracy	Precision	Recall	F1-score
Baseline ResNet50	Training set	0.96	0.95	0.97	0.96
	Test set	0.85	0.82	0.84	0.83
SE-ResNet50	Training set	0.99	0.98	0.99	0.98
	Test set	0.94	0.92	0.92	0.92

racic CT images. Compared to the baseline ResNet50 model, the proposed architecture demonstrated marked improvements across all key performance metrics [22], including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). These findings provide strong evidence that integrating Squeeze-and-Excitation (SE) blocks into conventional residual networks significantly enhances the capacity of the model to focus on diagnostically relevant features, consistent with reports from other medical imaging domains [23].

It is important to clarify the scope of our contribution within the context of surgical decision support. While our framework does not perform surgical staging or assess resectability, it addresses the foundational prerequisite for such downstream tasks: the accurate and reliable identification of lung adenocarcinoma. In the clinical pathway, a confident radiological diagnosis, supported by objective

and interpretable evidence, is the critical trigger for initiating surgical consultation and planning. By providing not only a highly accurate classification but also interpretable visual evidence via Grad-CAM aligned with expert annotations, our model aims to enhance the quality and confidence of the radiological reports informing surgical teams. Therefore, while not a comprehensive surgical planning solution, this work represents a tangible step toward strengthening the data-driven foundation upon which surgical decisions are ultimately based.

One of the major challenges in thoracic imaging is the subtle and heterogeneous presentation of early-stage lung adenocarcinoma, which often exhibits complex radiological features [24]. Traditional radiological assessment, though grounded in human expertise, is prone to significant inter-reader variability, even among experienced specialists [9]. Our results indicate that deep learning models, especially

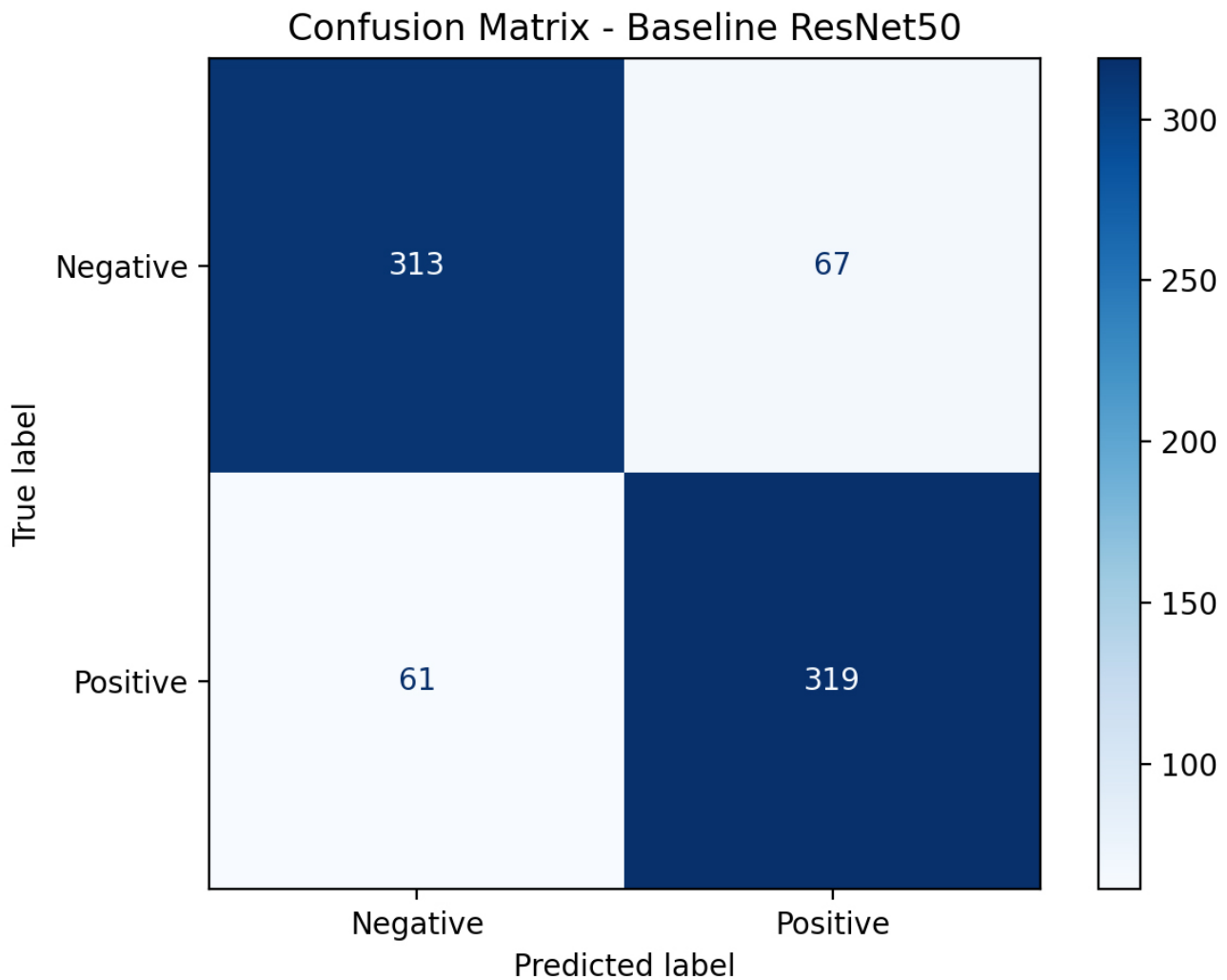


Fig. 4. Confusion matrix of the baseline ResNet50 model.

those augmented with attention mechanisms, can capture fine-grained radiological patterns that may be overlooked, serving as an effective second reader [25]. SE blocks, in particular, allowed the network to learn inter-channel dependencies, leading to stronger class-discriminative feature representations. This was evidenced in the significantly higher AUC (0.941 vs. 0.854) and reduced false positives and negatives, addressing key limitations in diagnostic AI [10].

Beyond improving classification accuracy, the SE-ResNet model exhibited more confident and polarized predictions, as evidenced by the distribution of confidence scores. Compared to the baseline, SE-ResNet produced clearer separation between high- and low-confidence outputs, suggesting better-calibrated decision boundaries and reduced model ambiguity [26]. This behavior is highly valuable in clinical practice, where well-calibrated, high-confidence predictions can support stronger diagnostic decisions, while low-confidence outputs can appropriately flag cases for further human review [27].

The application of Grad-CAM further strengthened the interpretability and clinical trustworthiness of our proposed model [19]. Visual inspection of Grad-CAM heatmaps revealed that the SE-ResNet model not only improved classification accuracy but also exhibited more anatomically and pathologically meaningful attention. In many cases, the attention maps of the model closely aligned with radiologist-annotated tumor regions, suggesting its potential for integration into radiological workflows to provide real-time visual cues, an approach increasingly advocated in human-AI collaboration [28]. To quantitatively evaluate the spatial reliability of model attention, we computed the average Intersection over Union (IoU) between Grad-CAM heatmaps and expert-annotated tumor regions. The SE-ResNet achieved a higher average IoU of 0.75, compared to 0.68 for the baseline ResNet50. This metric objectively supports the qualitative Grad-CAM findings and reinforces that SE-based attention mechanisms not only enhance prediction accuracy but also align more closely with radiologically relevant anatomy. Such alignment is critical for build-

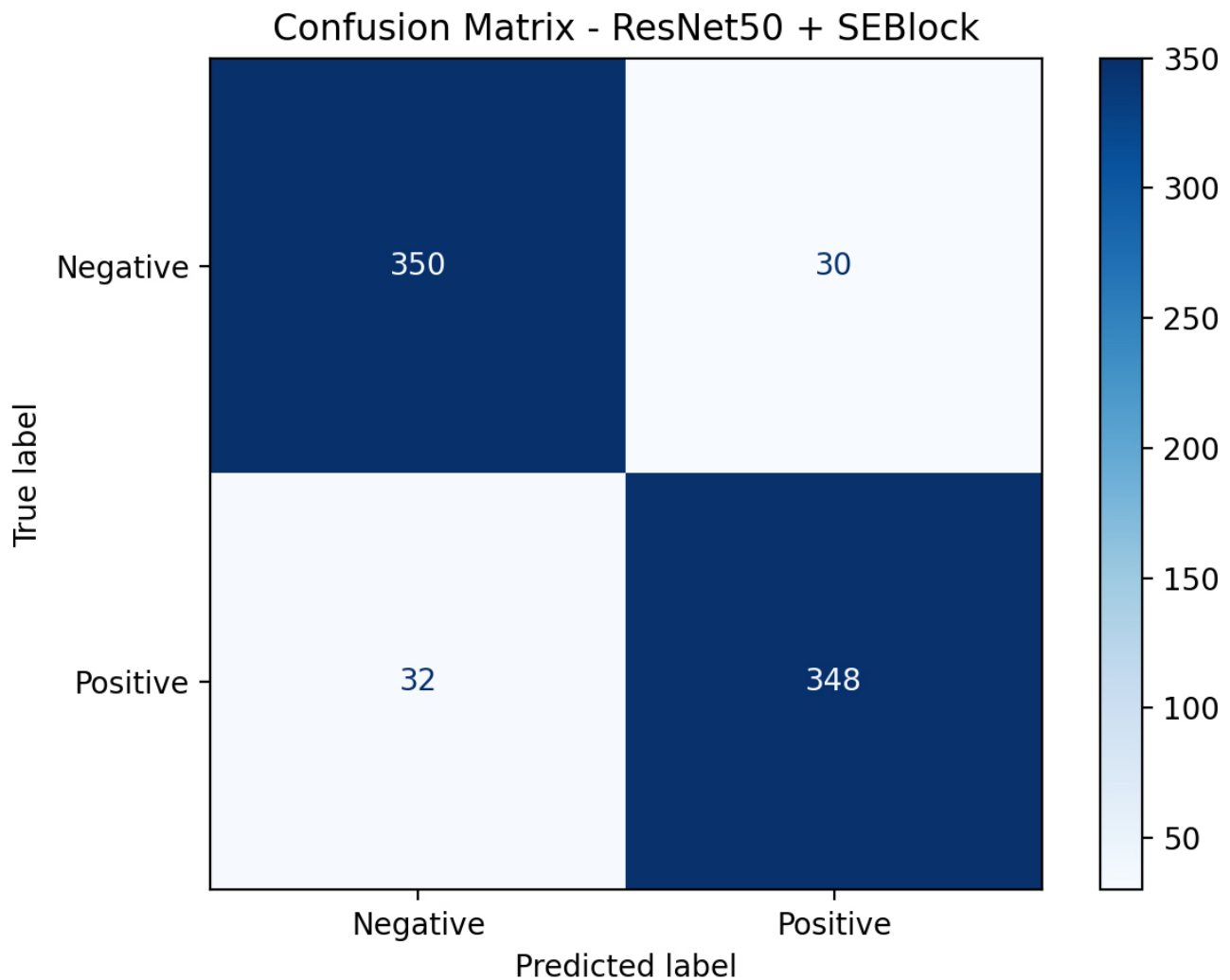


Fig. 5. Confusion matrix of the SE-ResNet model.

ing clinical trust and facilitating the adoption of AI-assisted diagnostic tools [29].

To better understand model limitations, we performed a retrospective analysis of misclassified test set cases. Among the 30 false positives (non-cancerous slices classified as adenocarcinoma), most (60%, 18/30) were focal inflammatory lesions or granulomas, which frequently exhibit high density and irregular margins resembling malignancy. Another subset (27%, 8/30) involved post-inflammatory scarring or atelectasis. Regarding the 32 false negative cases (adenocarcinoma slices classified as non-cancerous), lesion subtlety was the primary challenge. A substantial proportion (62.5%, 20/32) consisted of pure ground-glass nodules (GGOs) with low attenuation, while others (25%, 8/32) were small nodules (<6 mm) adjacent to pulmonary vessels, posing a significant challenge for differentiation. These findings suggest that future model improvements should focus on enhancing specificity against inflammatory mimics and improving sensitivity for subtle, low-density lesions.

Our findings align with a growing body of literature highlighting the value of attention mechanisms in medical image analysis [30]. Prior studies have successfully implemented attention-based networks for CT-based lung cancer detection [31], brain tumor segmentation from MRI [13], and breast lesion classification in mammography [32]. While several studies have investigated attention-augmented networks for CT-based lung cancer diagnosis, our findings further reinforce their effectiveness in thoracic imaging [13,33]. By focusing specifically on CT data, this study contributes robust evidence supporting the integration of attention mechanisms into widely adopted imaging modalities for pulmonary oncology.

Despite the promising results, several limitations must be acknowledged. First, while our IoU analysis provides a measure of localization on radiological images, a direct comparison with co-registered histopathology slides, the ultimate ground truth, was not performed due to the technical complexities of image registration. This remains a critical avenue for future validation. Second, the dataset used in

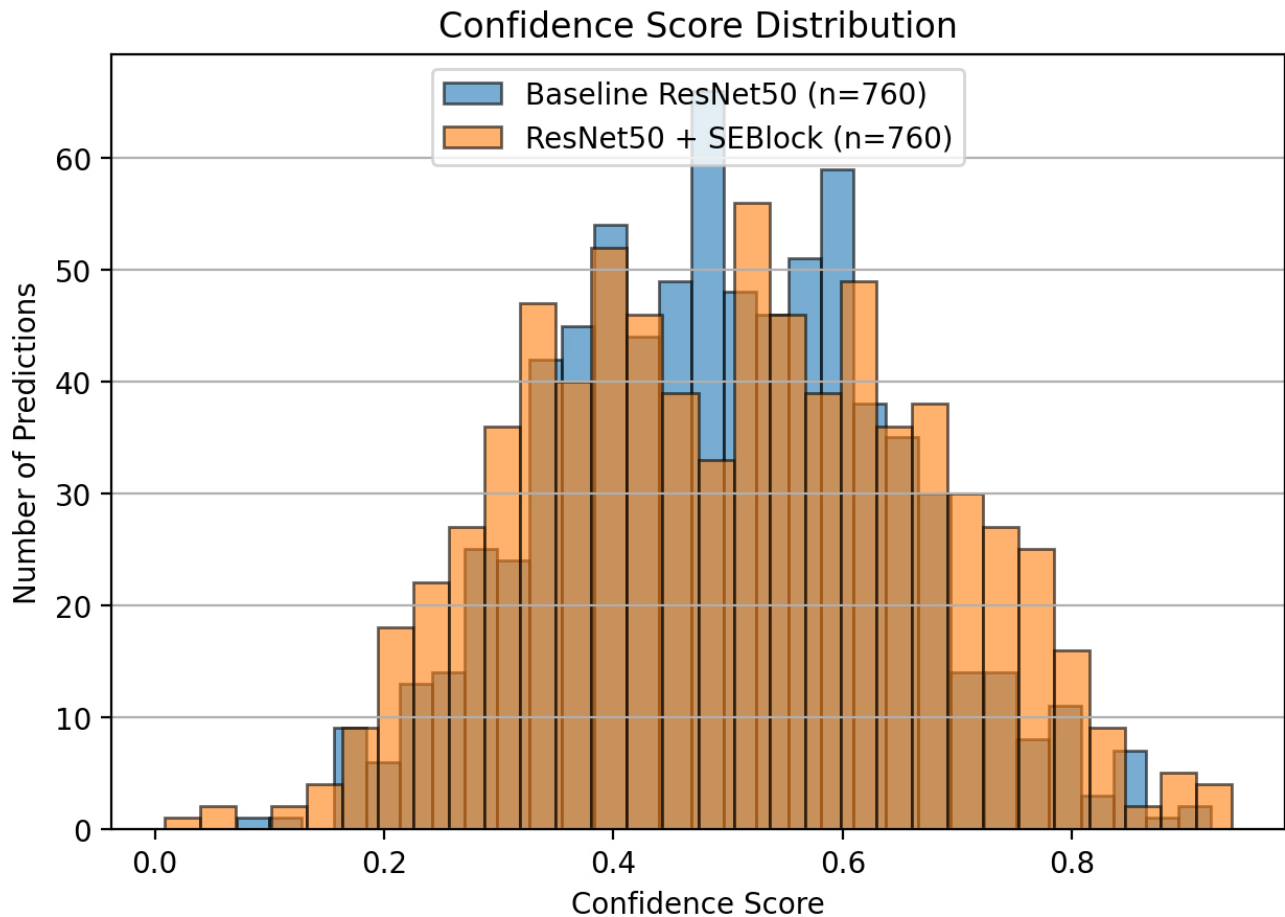


Fig. 6. Distribution of predicted confidence scores. Histogram of prediction confidence values for all test samples. The SE-ResNet model demonstrates more polarized outputs, indicating stronger class separability compared to the baseline ResNet50.

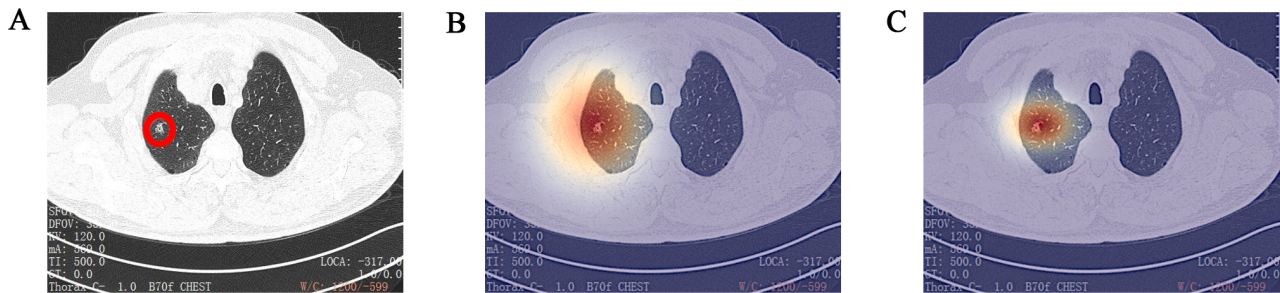


Fig. 7. Representative Gradient-weighted Class Activation Mapping (Grad-CAM) visualisations comparing expert annotation and model attention mechanisms. (A) Original CT slice with red circles indicating pathologically confirmed tumour regions. (B) Grad-CAM heatmap from baseline ResNet50 showing diffusion of attention mechanisms. (C) Grad-CAM heatmap from SE-ResNet showing more focused attention mechanisms, aligned to the real tumour region. Red areas indicate higher class activation.

this study was acquired from a single institution, which may limit the generalizability of the model across different scanners and patient populations, a well-recognized challenge in medical AI development [34]. Third, although SE blocks improved performance, other advanced attention mechanisms, such as self-attention or transformer-based architectures, may yield even greater gains and warrant further investigation [35]. Fourth, while Grad-CAM provides valu-

able insights, its resolution is inherently coarse and may not capture voxel-level pathology, suggesting that a more fine-grained saliency approach, such as Score-CAM or Eigen-CAM, should be explored in future work [36].

Additionally, our current evaluation was performed at the slice level, which, while appropriate for initial benchmarking, does not fully capture the clinical decision-making process that relies on patient-level diagnosis [37]. In real-

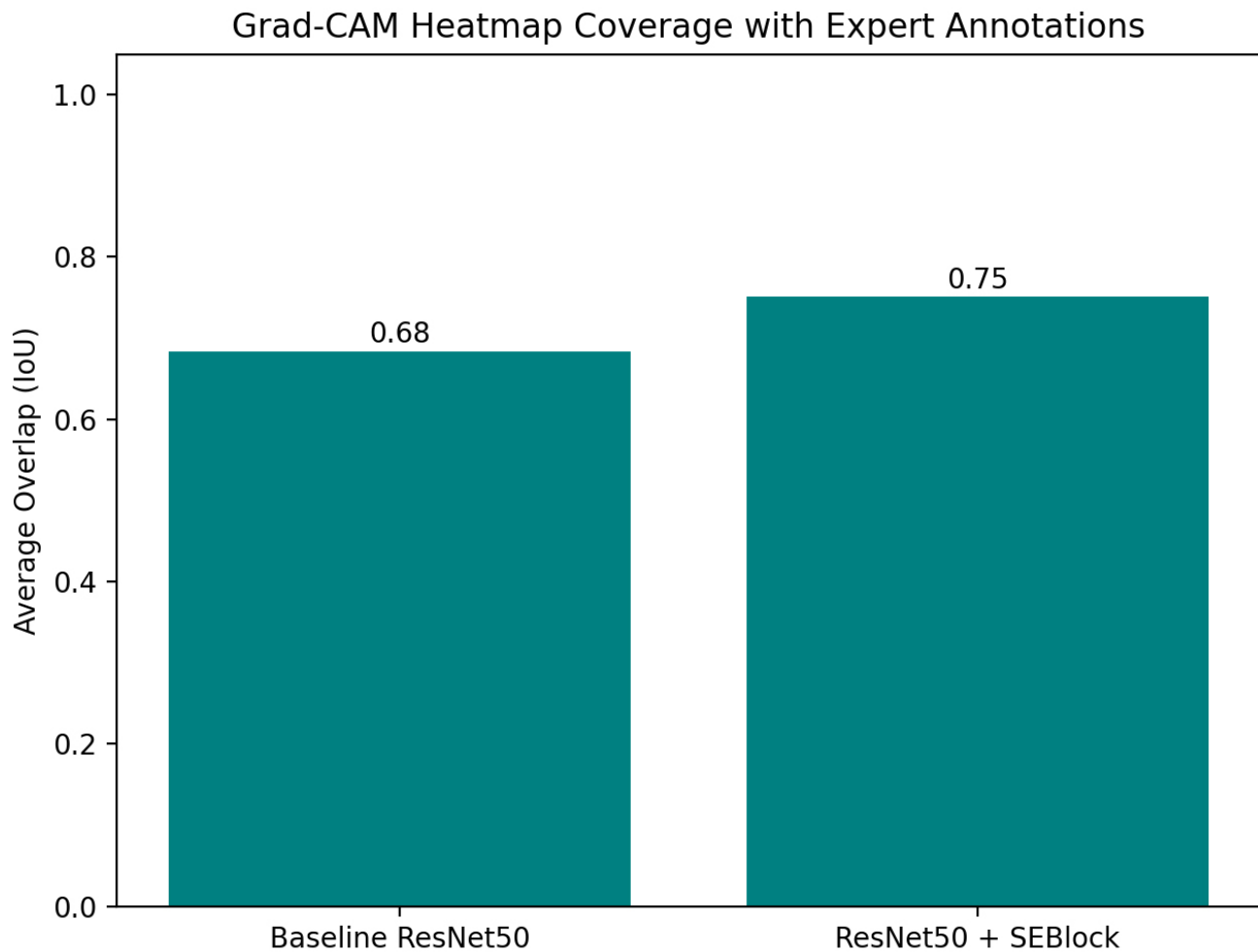


Fig. 8. Alignment of model attention with expert annotations measured by Intersection over Union (IoU). This figure compares the average IoU between Grad-CAM heatmaps and expert-annotated tumor regions for the baseline ResNet50 and the proposed SE-ResNet. The higher IoU of the SE-ResNet model correlates with its superior classification performance and indicates more reliable model interpretability.

world practice, radiologists assess multiple slices, considering spatial continuity and contextual cues across entire scans. Future research should extend the proposed framework to aggregate predictions across all slices belonging to the same patient using techniques such as max-pooling, weighted fusion, or more sophisticated temporal modeling approaches, including Long Short-Term Memory (LSTM) networks [38]. Such patient-level classification strategies will be essential for translating this model into clinically actionable tools [39].

Furthermore, as a technical proof-of-concept focused solely on image-based classification, this study did not incorporate detailed patient characteristics, such as specific tumor subtypes, treatment modalities, or comorbidities, into the model. This is a recognized limitation, as these clinical variables are critical for a comprehensive diagnostic and prognostic assessment. While this was beyond the scope of our current algorithmic investigation, future research

should prioritize the development of multi-modal fusion models. Integrating the proposed deep learning framework with structured clinical data from electronic health records (EHR) represents a crucial next step to enhance model robustness and create tools more closely aligned with the holistic clinical decision-making process [40,41].

Conclusions

This study developed and validated an attention-enhanced deep learning framework, SE-ResNet, for the classification of lung adenocarcinoma from thoracic CT images. By incorporating Squeeze-and-Excitation (SE) blocks, the SE-ResNet model significantly outperformed the baseline ResNet50, achieving an accuracy of 94% and an AUC of 0.941. Moreover, the application of Grad-CAM provided valuable visual interpretability, demonstrating that the decision-making process of the model aligns with clinically relevant tumor regions, thereby enhancing its trust-

worthiness. Collectively, these findings establish the SE-ResNet framework as a robust and interpretable approach for automated lung adenocarcinoma classification. It shows significant potential as a clinical decision-support tool designed to assist radiologists in improving the accuracy and efficiency of thoracic CT interpretation. Future research should focus on validating the model using multi-center datasets, extending the approach to 3D volumetric inputs, and integrating additional clinical variables to enable more comprehensive diagnostic decision-making.

Availability of Data and Materials

The source code and trained model weights supporting the findings of this study are available from the corresponding author upon reasonable request and are subject to a material transfer agreement. Clinical data cannot be publicly shared due to patient privacy restrictions imposed by the institutional review board.

Author Contributions

QS and YL designed the research study; JL and HH performed the research and data collection; QS analyzed the data and drafted the initial manuscript. QS, YL, JL, HH contributed to critical revision of the manuscript for important intellectual content, read and approved the final manuscript, agreeing to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Ethical approval was obtained from the Ethics Committee of Ningbo No. 2 Hospital (Approval No. [YJ-NBEY-KY-2021-087-01]). The study was conducted in accordance with the principles of the Declaration of Helsinki. Due to the retrospective nature of the study, informed consent was waived and all data were anonymised in compliance with data protection regulations.

Acknowledgment

Not applicable.

Funding

This research was supported by the Medical Scientific Research Foundation of Zhejiang Province (Grant No. 2020KY838) and the Ningbo Natural Science Foundation (Grant No. 2024J335).

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians*. 2024; 74: 229–263. <https://doi.org/10.3322/caac.21834>.

[2] Mattiuzzi C, Lippi G. Cancer statistics: a comparison between World Health Organization (WHO) and Global Burden of Disease (GBD). *European Journal of Public Health*. 2020; 30: 1026–1027. <https://doi.org/10.1093/eurpub/ckz216>.

[3] Bandi P, Star J, Ashad-Bishop K, Kratzer T, Smith R, Jemal A. Lung Cancer Screening in the US, 2022. *JAMA Internal Medicine*. 2024; 184: 882–891. <https://doi.org/10.1001/jamainternmed.2024.1655>.

[4] Saman H, Raza A, Patil K, Uddin S, Crnogorac-Jurcevic T. Non-Invasive Biomarkers for Early Lung Cancer Detection. *Cancers*. 2022; 14: 5782. <https://doi.org/10.3390/cancers14235782>.

[5] Bolowia NF. Computed Tomography (CT) Scans: Advancements in Oncology Diagnosis and Treatment. *Derna Academy Journal for Applied Sciences*. 2025; 3: 140–147. <https://doi.org/10.71147/t415y560>.

[6] De Koning HJ, van der Aalst CM, Ten Haaf K, Oudkerk M. PL02.05 Effects of Volume CT Lung Cancer Screening: Mortality Results of the NELSON Randomised-Controlled Population Based Trial. *Journal of Thoracic Oncology*. 2018; 13: S185. <https://doi.org/10.1016/j.jtho.2018.08.012>.

[7] Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, Caughey AB, et al. Screening for Lung Cancer. *Journal of the American Medical Association*. 2021; 325: 962–970. <https://doi.org/10.1001/jama.2021.1117>.

[8] MacMahon H, Austin JHM, Gamsu G, Herold CJ, Jett JR, Naidich DP, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*. 2005; 237: 395–400. <https://doi.org/10.1148/radiol.2372041887>.

[9] Kim H, Park CM, Kim SH, Lee SM, Park SJ, Lee KH, et al. Persistent pulmonary subsolid nodules: model-based iterative reconstruction for nodule classification and measurement variability on low-dose CT. *European Radiology*. 2014; 24: 2700–2708. <https://doi.org/10.1007/s00330-014-3306-7>.

[10] van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European Radiology*. 2021; 31: 3797–3804. <https://doi.org/10.1007/s00330-021-07892-z>.

[11] Yu S, Mahil A. Introduction to AI in Radiology. *Artificial Intelligence in Clinical Medicine*. 2023; 318–320.

[12] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nature Medicine*. 2019; 25: 24–29. <https://doi.org/10.1038/s41591-018-0316-z>.

[13] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. *International MICCAI BrainLes Workshop* (pp. 272–284). Springer: Cham, Switzerland. 2021.

[14] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). 2016.

[15] Li X, Li M, Yan P, Li G, Jiang Y, Luo H, et al. Deep learning attention mechanism in medical image analysis: Basics and beyonds. *International Journal of Network Dynamics and Intelligence*. 2023; 93–116. <https://doi.org/10.53941/ijndi0201006>.

[16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018.

[17] Woo S, Park J, Lee J, Kweon IS. CBAM: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3–19). 2018.

[18] Painuli D, Bhardwaj S, Köse U. Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review. *Computers in Biology and Medicine*. 2022; 146: 105580. <https://doi.org/10.1016/j.combiomed.2022.105580>.

- [19] He Y. Diving deep: The role of deep learning in medical image analysis, today and tomorrow. In Proceedings of the 2024 9th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (pp. 537–540). 2024.
- [20] Garg A, Singh A, Kumar A. Mental disorders management using explainable artificial intelligence (XAI). Explainable Artificial Intelligence for Biomedical and Healthcare Applications (pp. 113–138). CRC Press: Boca Raton, FL, USA. 2024.
- [21] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 618–626). 2017.
- [22] Miller RA. Reference standards in evaluating system performance. Journal of the American Medical Informatics Association: JAMIA. 2002; 9: 87–88. <https://doi.org/10.1136/jamia.2002.0090087>.
- [23] Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: A survey. Medical Image Analysis. 2023; 88: 102802. <https://doi.org/10.1016/j.media.2023.102802>.
- [24] Xu Y, Sun H, Zhang Z, Song A, Wang W, Lu X. Assessment of Relationship Between CT Features and Serum Tumor Marker Index in Early-stage Lung Adenocarcinoma. Academic Radiology. 2016; 23: 1342–1348. <https://doi.org/10.1016/j.acra.2016.06.005>.
- [25] Lång K, Josefsson V, Larsson AM, Larsson S, Högborg C, Sartor H, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. The Lancet. Oncology. 2023; 24: 936–944. [https://doi.org/10.1016/S1470-2045\(23\)00298-X](https://doi.org/10.1016/S1470-2045(23)00298-X).
- [26] Pernot P. Calibration in machine learning uncertainty quantification: Beyond consistency to target adaptivity. APL Machine Learning. 2023; 1. <https://doi.org/10.1063/5.0174943>.
- [27] Ouane K, Farhah N. Effectiveness of Artificial Intelligence (AI) in Clinical Decision Support Systems and Care Delivery. Journal of Medical Systems. 2024; 48: 74. <https://doi.org/10.1007/s10916-024-02098-4>.
- [28] Soomro TA, Zheng L, Afifi AJ, Ali A, Soomro S, Yin M, et al. Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review. IEEE Reviews in Biomedical Engineering. 2023; 16: 70–90. <https://doi.org/10.1109/RBME.2022.3185292>.
- [29] Ghassemi M, Hightower M, Nsoesie EO. Settling the score on algorithmic discrimination in health care. New England Journal of Medicine AI. 2024; 1: 731–733. <https://doi.org/10.1056/AI.p2400583>.
- [30] Gupta S, Sharma S, Sharma R, Chandra J. Healing with hierarchy: Hierarchical attention empowered graph neural networks for predictive analysis in medical data. Artificial Intelligence in Medicine. 2025; 165: 103134. <https://doi.org/10.1016/j.artmed.2025.103134>.
- [31] Tang T, Zhang R, Lin K, Li F, Xia X. SM-RNet: A scale-aware-based multiattention-guided reverse network for pulmonary nodules segmentation. IEEE Transactions on Instrumentation and Measurement. 2023; 72: 1–14.
- [32] Kuttan GO, Elayidom MS. Review on Computer Aided Breast Cancer Detection and Diagnosis using Machine Learning Methods on Mammogram Image. Current Medical Imaging. 2023; 19: 1361–1371. <https://doi.org/10.2174/1573405619666230213093639>.
- [33] Zhang Y, Feng W, Wu Z, Li W, Tao L, Liu X, et al. Deep-Learning Model of ResNet Combined with CBAM for Malignant-Benign Pulmonary Nodules Classification on Computed Tomography Images. Medicina (Kaunas, Lithuania). 2023; 59: 1088. <https://doi.org/10.3390/medicina59061088>.
- [34] Liu H, Cai H, Yang D, Zhu W, Wu G, Chen J. Learning pyramidal multi-scale harmonic wavelets for identifying the neuropathology propagation patterns of Alzheimer's disease. Medical Image Analysis. 2023; 87: 102812. <https://doi.org/10.1016/j.media.2023.102812>.
- [35] Aburass S, Dorgham O, Al Shaqsi J, Abu Rumman M, Al-Kadi O. Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases. Journal of Imaging Informatics in Medicine. 2025; 10.1007/s10278–10.1007/s10278–025–01481–y. <https://doi.org/10.1007/s10278-025-01481-y>.
- [36] Singh A, Mishra D. CoBooM: Codebook guided bootstrapping for medical image representation learning. Lecture Notes in Computer Science (pp. 23–33). Springer: Cham, Switzerland. 2024.
- [37] Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. Medical Image Analysis. 2021; 72: 102125. <https://doi.org/10.1016/j.media.2021.102125>.
- [38] Takahashi K, Usuzaki T, Inamori R. Vision Transformer-based Deep Learning Models Accelerate Further Research for Predicting Neurosurgical Intervention. Radiology. Artificial Intelligence. 2024; 6: e240117. <https://doi.org/10.1148/ryai.240117>.
- [39] Wang LV, Gao L. Photoacoustic microscopy and computed tomography: from bench to bedside. Annual Review of Biomedical Engineering. 2014; 16: 155–185. <https://doi.org/10.1146/annurev-bioeng-071813-104553>.
- [40] Wang H, Subramanian V, Syeda-Mahmood T. Modeling uncertainty in multi-modal fusion for lung cancer survival analysis. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 1169–1172). 2021.
- [41] Da-Ano R, Tankyevych O, Andrade-Miranda G, Prasoon A, Meriaudeau F, Chen T, et al. Multi-modal PET/CT fusion for automated PD-L1 status prediction in lung cancer. In Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI) (pp. 1–5). 2024.

© 2025 The Author(s).

