Original Article

Enhanced Pulmonary Nodule Detection and Classification Using Artificial Intelligence on LIDC-IDRI Data

Lotfi Salhi* [ID], Khawla Moussa and Ridha Ben Salah

Laboratory of Biophysics and Medical Technology, Higher Institute of Medical Technologies of Tunis, University of Tunis El Manar, Tunis, Tunisia

Abstract

Background and objectives: Lung cancer remains the leading cause of cancer-related mortality worldwide. Early detection of pulmonary nodules is crucial for timely diagnosis and effective treatment. Conventional computer-aided detection systems have shown limitations, including high false-positive rates and low sensitivity. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have shown great potential in improving the accuracy and reliability of nodule detection and classification. This study aimed to develop and evaluate an automatic method for lung nodule detection and classification using a CNN-based architecture applied to computed tomography images from the publicly available LIDC-IDRI database.

Methods: This retrospective study was conducted on 82 patients (10,496 computed tomography slices) selected from the LIDC-IDRI database. The proposed method consists of five main steps: image preprocessing, lung parenchyma segmentation using Otsu's thresholding and morphological operations, detection of nodule candidates, feature extraction, and classification using a CNN model. The CNN architecture includes two convolutional layers (20 and 30 filters, 3×3 kernel), ReLU activation, max-pooling layers, and a Softmax output layer. The network was trained with a mini-batch size of 32 for 50 epochs using the Stochastic Gradient Descent with Momentum optimizer (learning rate = 0.001, momentum = 0.9). Model performance was evaluated in terms of sensitivity, specificity, precision, and accuracy.

Results: The proposed CNN model successfully detected pulmonary nodules and achieved accurate classification between benign and malignant nodules. On the LIDC-IDRI dataset, the model achieved a sensitivity of 98.7%, specificity of 97.5%, precision of 97.9%, and accuracy of 98.4%. Comparative analysis with recent studies, including hybrid CNN-long short-term memory and ResNet-based models, demonstrated that the proposed method provides competitive performance while maintaining lower computational complexity. The classification of nodule subtypes (solid, partially frosted, totally frosted) showed satisfactory discrimination results.

Conclusions: The proposed CNN-based system demonstrates the feasibility and robustness of deep learning for automatic lung nodule detection and classification. Despite strong results, the study acknowledges limitations such as single-database validation and a relatively small training size. Future work will focus on validating the model across other datasets (e.g., EL-CAP, NELSON) and optimizing multi-class classification performance to enhance generalizability and clinical applicability.

Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide. According to the latest estimates, approximately 2.48 million new cases of lung cancer were diagnosed globally, accounting for nearly 12.4% of all cancer cases and resulting in more than 1.8 million deaths each year.[1] Early detection of pulmonary nodules, the small lesions that may indicate early-stage lung cancer, significantly improves patient prognosis. However, manual interpretation of computed tomography (CT) scans is challenging,

time-consuming, and prone to inter-observer variability, particularly for small or ground-glass nodules.[2] The size of the nodule provides diagnostic information in lung lesion screening. In fact, the percentage of malignancy for nodules less than 5 mm is 1%, 24% for nodules between 6 mm and 10 mm, 33% for nodules between 11 mm and 20 mm, and 80% for nodules greater than 20 mm.[3,4] This demonstrates that the risk of malignancy is a growing function of nodule size.

Taking advantage of improvements in CT technology, pulmonary nodules can be characterized in density more precisely as solid, partially solid, or pure ground-glass opacities. This precision is particularly useful for classifying small nodules (1 cm), making it possible to distinguish between benign and malignant nodules. In the literature, the percentage of pure ground-glass opacities that are malignant varies widely, from 18% to nearly 60%. The probability of a malignant tumor for sub-centimeter nodules is also high in partially solid lesions but much lower (10%) in solid nodules.[5] The growth rate, or the time required for a nodule to increase in volume, is a reliable criterion for differentiating between benign and malignant lesions. Usually, if the volume of a nodule has not changed over a two-year period, then the lesion may be considered benign and does not require further diagnostic assessment.[6]

Advancements in artificial intelligence (AI) and deep learning have enabled the development of computer-aided diagnosis (CAD) systems that assist radiologists in detecting and characterizing pulmonary nodules. Convolutional neural networks (CNNs), in particular, have demonstrated high performance in medical image classification and segmentation tasks due to their ability to automatically extract hierarchical features.[7,8] They proposed a CNN-based framework using dual-time-point 18F-fluorodeoxyglucose positron emission tomography/CT data to predict the malignancy risk of nodules, achieving superior classification performance compared to radiomics-based models. Similarly, Ji *et al*.[9] demonstrated the diagnostic value of 3D CT reconstruction in differentiating benign and malignant nodules, emphasizing the role of spatial context in improving specificity. Several other studies have explored hybrid and optimized deep learning architectures to reduce false positives and improve interpretability. Xue *et al*.[10] introduced an AI-assisted diagnostic system integrating CNNs with feature visualization to aid radiologists in clinical decision-making. Recent efforts by Wang *et al*.[11] and Gupta *et al*.[8] have focused on improving generalization through data augmentation, transfer learning, and multimodal feature fusion.

Despite these advances, several challenges persist, including false-positive reduction, limited generalizability across imaging protocols, and the need for interpretability of AI decisions.[12–15] Significant challenges also remain in accurately detecting small and ground-glass nodules, maintaining an optimal balance between sensitivity and specificity, and validating model performance across heterogeneous datasets, particularly those derived from low-dose CT (LDCT) imaging.

In this context, there is a growing need for reliable, transparent, and high-performing CAD systems that can complement clinical workflows and improve diagnostic confidence. The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) provides a robust, annotated benchmark dataset for developing and validating such models.[15–17] Therefore, the present study aimed to develop and evaluate an automated system for detecting and classifying pulmonary nodules using the LIDC-IDRI dataset within the MATLAB environment. The proposed framework integrates advanced image preprocessing, Sobel-based candidate detection, and CNN-based classification optimized with

Synthetic Minority Oversampling Technique data augmentation to reduce class imbalance.[18] The system aims to improve detection specificity, enhance discrimination between benign and malignant nodules, and provide interpretable outputs to support radiologists in early lung cancer screening. The system's performance is quantitatively evaluated and compared with existing CAD approaches to demonstrate its clinical relevance and potential for integration into lung cancer screening programs.

Materials and methods

Dataset and justification

This retrospective study utilized the LIDC-IDRI dataset, a publicly available collection of 1,018 thoracic CT scans with detailed annotations of pulmonary nodules. Each scan was independently reviewed by four experienced thoracic radiologists, who marked nodules ≥ 3 mm and assigned malignancy likelihood scores, followed by a consensus review. The dataset includes varied nodule types (solid, part-solid, ground-glass), sizes, and locations, providing a robust and diverse sample for training and evaluating AI-based detection and classification systems. LIDC-IDRI has been widely used for benchmarking CAD algorithms due to its high-quality annotations, multi-center acquisition, and standardized metadata, enabling reproducibility and meaningful comparison across studies.[15] The diversity of nodule characteristics allows the proposed CNN-based system to learn discriminative features, improve generalization, and reliably classify nodules as benign or malignant, supporting clinical applicability.

Ethical considerations

This study utilized data from the publicly available LIDC-IDRI. As this dataset is fully anonymized and was collected with prior institutional review board approval at all participating centers, our retrospective analysis of this data did not require additional ethical approval from our institution's institutional review board.

Sample selection and group definitions

For this study, we included CT scans with annotated nodules measuring ≥3 mm in diameter, as defined by the LIDC-IDRI guidelines. Nodules were categorized into two groups based on the consensus of the radiologists: benign and malignant. The dataset's diversity, encompassing various nodule types and characteristics, provides a robust foundation for evaluating AI-based detection and classification methods.

Nodules ranging from 3 to 30 mm in diameter were selected for analysis. Nodules identified by fewer than three radiologists were excluded to ensure annotation reliability. To assign a diagnostic label (benign or malignant) to each nodule, we calculated the average malignancy score provided by the radiologists, which ranged from 1 (highly unlikely malignant) to 5 (highly suspicious). Nodules were classified as:
- Benign: average score between 1 and 2.5;
- Malignant: average score between 3.5 and 5;
- Nodules with average scores falling between 2.5 and 3.5 were excluded from the study to avoid ambiguity in classification.

After applying the selection criteria, we chose 82 patients (10,496 slices: 6,912 malignant slices and 3,584 benign slices) for the classification method. Note that the DICOM files with the axial slices corresponding to the selected nodules were extracted from the CT scans and stored in a folder (the slice numbers were obtained from the Excel file).

Image preprocessing

The preprocessing stage aims to enhance the quality of CT images and isolate lung parenchyma for subsequent analysis. Each DICOM image was first converted to grayscale and normalized to ensure consistent intensity scaling. Contrast adjustment was applied to improve the visibility of pulmonary structures, followed by threshold-based segmentation to distinguish the lung region from surrounding tissues. Morphological operations, such as erosion and dilation, were performed to remove small artifacts and refine the lung boundaries.

To eliminate the rib cage and trachea regions, edge removal techniques were employed, thereby improving the accuracy of subsequent nodule detection. The Sobel edge detection filter was applied to highlight potential nodule boundaries, serving as the initial candidate detection stage. The resulting binary masks were then filtered based on size and shape constraints to retain structures consistent with pulmonary nodules.[19]

Contrast adjustment and threshold selection were carefully tuned to balance sensitivity and specificity, particularly for ground-glass nodules, whose subtle intensity variations can complicate detection. These settings were adapted iteratively to optimize nodule visibility without introducing false positives.[20]

Nodule detection

The proposed detection algorithm enabled the automatic identification of pulmonary nodules from CT images within the LIDC-IDRI database. Building upon the classification framework, the system utilized annotated reference data to guide nodule localization, classification, and performance evaluation using a CNN.

As illustrated in the following, the overall workflow consisted of four main stages:

- Image acquisition and preprocessing, including normalization and enhancement.
- Segmentation of the pulmonary parenchyma to isolate the lung regions.
- Detection of candidate nodules through morphological and edge-based operations.
- Storage for the classification step.

This structured pipeline ensured reliable detection and differentiation of pulmonary nodules while minimizing false positive results.

Lung segmentation

To reduce the nodule search space and focus analysis on relevant areas, the lung parenchyma was segmented using the Otsu automatic thresholding method combined with mathematical morphology operations. The Otsu algorithm, one of the most widely used automatic thresholding techniques, assumes that the image consists of two distinct classes of pixels—foreground and background—and determines the optimal threshold value that minimizes intra-class variance while maximizing inter-class variance.[17] Once the optimal threshold was identified, it was applied to the grayscale image to generate a binary image. Pixels with intensity values above the threshold were classified as foreground (lung regions), while those below it corresponded to the background.

After thresholding, the binary image underwent a morphological opening operation using a disc-shaped structuring element with a radius of 10 pixels. This step removed small, isolated regions and residual artifacts resulting from binarization, ensuring a cleaner segmentation of the lung parenchyma.[21]

Detection of candidate lung nodules

This step aims to automatically identify regions within the pulmonary parenchyma that may correspond to potential pulmonary nodules, referred to as nodule candidates. A contour-based segmentation approach was adopted to extract these regions of interest (ROIs), as it preserved the spatial localization of nodules.

Proposed classification model

Several studies have demonstrated the effectiveness of CNNs for pulmonary nodule classification, as they enable automatic feature extraction from medical images. Building on this approach, the proposed classification algorithm followed the architecture illustrated below, which provided a systematic pipeline for distinguishing benign and malignant nodules as follows.

CNN workflow

The workflow of the proposed CNN-based nodule classification, consisted of the following steps:

1. Load examples from the database: Test and validation images were loaded separately from the folders containing the preprocessed ROIs. Corresponding class labels were stored as categorical vectors.
2. Define the CNN structure: The network architecture was defined with input, convolutional, ReLU activation, pooling, fully connected, dropout, and Softmax layers. Filter sizes, number of kernels, and other hyperparameters were specified to optimize feature extraction.
3. CNN learning: The network was trained using the Stochastic Gradient Descent with Momentum (SGDM) optimizer. Mini-batches of size 32 were used, with a learning rate of 0.001, momentum of 0.9, and early stopping based on validation loss. Data augmentation was applied to increase generalization.
4. CNN testing: The trained network was evaluated on the independent test dataset to predict class probabilities for each nodule.
5. Performance calculation: Standard performance metrics were computed, including accuracy, precision, recall, specificity, F1 score, and Matthews correlation coefficient (MCC), to assess the classification performance of the model.

The algorithm loaded the test and validation datasets separately from the folders containing the nodule images (ROIs). Corresponding class labels for each dataset were stored as categorical vectors for subsequent training and evaluation.

CNN architecture

The proposed CNN consists of the following layers: an input layer, a first convolutional layer, a ReLU activation layer, a first pooling layer, a second convolutional layer, a ReLU activation layer, a second pooling layer, and a classification layer. This architecture is illustrated in Figure 1.

Lung nodule classification

As illustrated in Figure 1, the proposed CNN is designed to automatically learn discriminative features from CT images for lung nodule classification. The architecture consists of successive layers that progressively transform the input image into higher-level feature representations.

The input layer receives preprocessed lung patch images obtained after preprocessing, equalization, and normalization. Two convolutional layers then extract local spatial features such as edges, textures, and nodule shapes. Each convolutional layer applies multiple 3×3 filters (20 filters in the first layer and 30 in the second), followed by a ReLU activation to introduce non-linearity and accelerate convergence, enabling the network to model com-
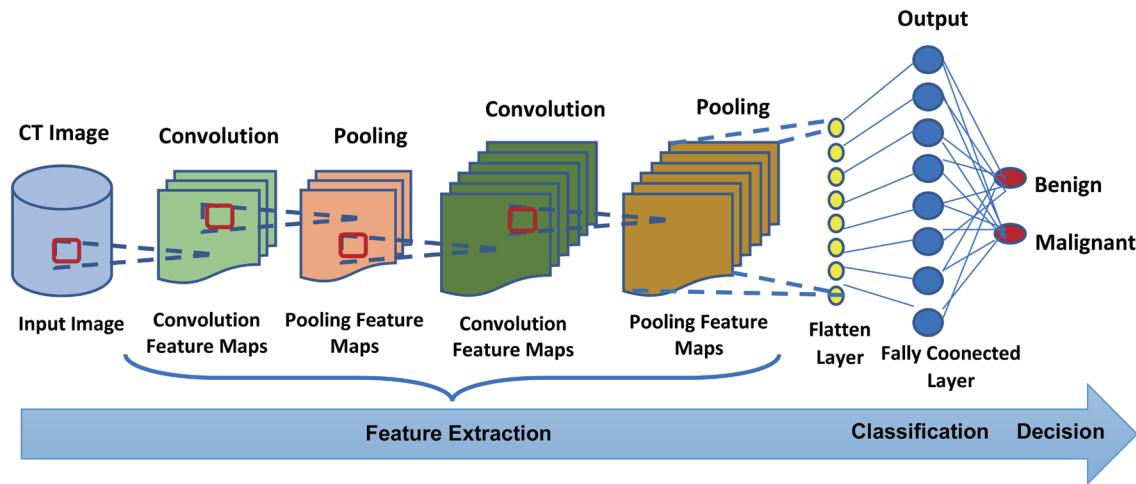
Fig. 1. Structure of the neural network (CNN).

plex patterns.

Two max-pooling layers (2×2) reduce the spatial dimensions while retaining the most informative features, thereby limiting overfitting and computational complexity. The extracted features are then flattened and passed to fully connected layers that integrate the learned representations for classification. Dropout regularization is applied to further prevent overfitting. Finally, a Softmax layer outputs the probability of each class (benign vs. malignant).

For training, the weights and biases were initialized to 1, and network parameters were optimized using the SGDM algorithm to ensure stable and efficient convergence. The learning rate was set to 0.001 with a momentum factor of 0.9, a mini-batch size of 32, and 50 training epochs. Early stopping was applied based on validation loss to prevent overfitting and improve generalization.

Learning process

The training process divides the dataset into smaller subsets called mini-batches. Each mini-batch is fed into the network, which updates its parameters—weights and biases—based on the selected learning function. In this study, the SGDM optimizer was used, defined by the following update rule[22,23]:

$$\theta_{I+1} = \theta_I - a\nabla E(\theta_I) + \gamma(\theta_I - \theta_{I-1}) \qquad (1)$$

where θ represents the vector of network parameters (weights and biases), I is the iteration index corresponding to the current mini-batch, $E(\theta_I)$ is the error function evaluated at iteration I, $\nabla E(\theta_I)$ is the gradient of the error function with respect to the parameters, and γ is the momentum term, which incorporates the contribution of the previous update into the current iteration.

This approach allows the network to converge more efficiently by accelerating updates in consistent gradient directions and reducing oscillations in regions of high curvature. Although SGDM optimization is commonly used for training CNNs in medical image analysis, it is preferable to compare its performance with alternative optimizers, such as Adam, to further justify the choice of learning algorithm.

Figure 2 illustrates examples of various types of pulmonary nodules utilized during the training and testing phases, highlighting the diversity of the dataset in terms of nodule size, shape, and appearance.

Evaluation of the classification model: Statistical analysis

Evaluation metrics for classification are used to assess the performance of a model. These metrics are derived from the confusion matrix obtained after learning and testing the classification model. The most common metrics are accuracy, precision, recall, F1 score, and MCC. From the true positives, true negatives, false positives, and false negatives, these metrics are calculated using the following formulas[24]:

$$\text{Recall} = \text{Sensitivity} = \frac{TP}{TP + FN} \qquad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (3)$$

$$\text{NPV} = \frac{TN}{TN + FN} \qquad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (5)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (6)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (7)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \qquad (8)$$

The recall (or sensitivity or true positive rate) measures the ability of the classifier to identify all positive instances. It determines how many positive instances in the database were correctly identified.

The specificity (or true negative rate) measures the ability of the classifier to identify all negative instances. It determines how many negative instances in the database were correctly identified.

Examples of Lung Nodules from the LIDC-IDRI Database
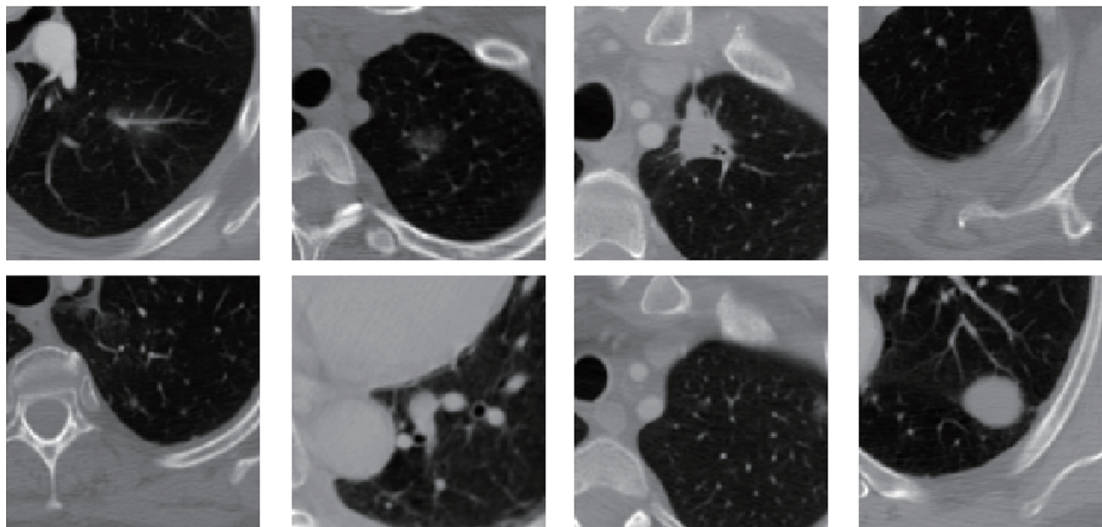
Fig. 2. Examples of lung nodules from the database (LIDC-IDRI) used in the learning and testing phases of this study, with their varied characteristics.

Accuracy measures the overall correctness of the classifier. It represents the proportion of correctly classified instances over the total number of instances.

Precision quantifies the accuracy of positive predictions made by the classifier. It determines how many instances classified as positivewere actually positive. The F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is particularly useful when there is an imbalance between positive and negative classes.

The MCC is a balanced metric that considers all four components of the confusion matrix. It ranges from −1 to +1, where +1 represents a perfect classifier, 0 indicates a random classifier, and −1 denotes a classifier that performs exactly opposite to the desired behavior. A higher MCC indicates a better classifier.

Explainable AI (XAI) visualization

To enhance model interpretability and facilitate clinical validation, we incorporated XAI techniques that visualize the internal decision-making process of the deep learning classifier. Two complementary visualization methods were employed: gradient-weighted class activation mapping (Grad-CAM) and occlusion sensitivity analysis.

Grad-CAM highlights the most influential image regions that contribute to the model's prediction by computing the gradient of the target class score with respect to the feature maps in the final convolutional layer. The resulting activation maps were superimposed on the original CT slices to visually identify discriminative regions corresponding to malignant or benign nodules.[25]

Occlusion sensitivity was used to assess the robustness of model predictions by systematically occluding portions of the input image and observing the corresponding change in classification probability. Regions where occlusion led to a significant drop in the predicted probability were considered critical for the model's decision.[26]

Both visualization methods were applied to randomly selected malignant and benign cases from the LIDC-IDRI dataset. The resulting attention maps were normalized and color-coded (red in-

dicating high importance, blue indicating low importance) to aid visual interpretation.

Results

Data augmentation

When the data is unbalanced, the AI model tends to favor majority classes, which can distort the results and lead to inaccurate predictions for minority classes. Ensuring a good balance in datasets allows training models capable of fairly treating all classes, thus guaranteeing more reliable and unbiased predictions. To achieve data augmentation, we used a system based on different geometric transformations, such as vertical flip, horizontal flip, and rotation of 25 degrees. After oversampling the data using the most popular technique, Synthetic Minority Oversampling Technique, this technique attempts to balance class data by randomly increasing minority class elements while replicating them. Similarly, to increase the size of the dataset, we used the Coarse Dropout technique. Figure 3 shows the distribution of the database elements before and after the data increase and balancing operation.

The database of selected images was divided into two parts: 80% for training the classifier and 20% for its evaluation.

The composition of this distribution can be summarized in the following table (Table 1):

Image preprocessing

The preprocessing stage produced images containing only the enhanced thoracic region with improved visual quality, as illustrated in Figure 4.

While the preprocessing steps, including contrast adjustment and threshold selection, successfully reduced rib cage artifacts, specific adjustments for ground-glass nodules were not applied, which may slightly affect detection sensitivity for this subtype. Contrast adjustment and threshold selection are essential tools to balance sensitivity and specificity in the interpretation of ground-
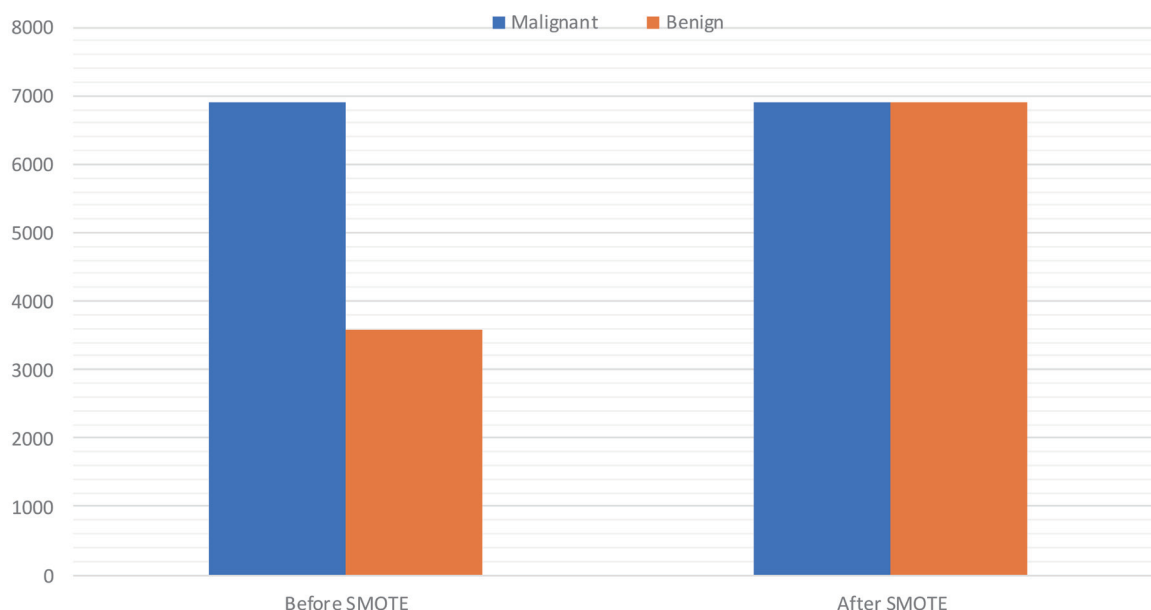
Fig. 3. Distribution of the database elements before and after data augmentation based on SMOTE. SMOTE, synthetic minority oversampling technique.

glass nodules. These settings should be adapted according to the clinical context and the objectives of the examination.

Segmentation and detection

For segmentation, a median filter was applied to the segmented lung image to reduce noise and smooth intensity variations, thereby improving contour clarity. Subsequently, edge detection was performed using the derivative-based method to delineate the contours of potential nodules, as illustrated in Figure 4.

Following contour detection, the resulting image was labeled to identify and separate connected components, where each component corresponds to a distinct region. These regions were then isolated into individual binary masks. For each mask, hole-filling operations were performed to obtain homogeneous regions from the detected contours, followed by morphological erosion to eliminate small, irrelevant areas.

The resulting binary mask effectively delineated the thoracic region, as illustrated in Figure 5.

In this study, the technique was applied to remove the background from CT images, thereby preparing them for accurate thresholding. The initial threshold was set to –950 Hounsfield Units (HU), as the intensity values of most lung parenchymal regions typically fall between –950 HU and –500 HU. To improve segmentation accuracy, the threshold value was then recalculated iteratively using an error function based on gray-level variations in the image histogram. This adaptive process was applied inde-

pendently to each CT slice, since the optimal threshold determined for one image is not necessarily valid for another due to inter-slice intensity variations.

Subsequently, the binary mask was inverted by reversing pixel intensities: black pixels were converted to white, and white pixels to black, in order to isolate the lung parenchyma. The edges of the bright regions, corresponding to the pulmonary lobes, were then removed to refine the segmentation. Finally, the resulting mask was superimposed onto the original CT image to extract the parenchymal region, as illustrated in Figure 5.

This process effectively reduced the number of detected regions. For instance, in the example shown in Figure 6, the number of labeled regions decreased from twelve to five after applying the filtering and morphological refinement steps. The final set of nodule candidates obtained through the proposed method is presented in Figure 6.

To evaluate the performance of the proposed algorithm, after passing all the data from the balanced database, the network set these parameters. Then, we tested the network on a test database. The results are given as a confusion matrix shown in Table 2.

From this table, we calculated the different metric values, which are summarized in Figure 7.

Model interpretability and visualization

Representative visualization results are presented in Figure 8, illustrating the interpretability of the proposed CAD model for both malignant and benign nodules. For malignant nodules, Grad-CAM and occlusion maps consistently highlighted the core and irregular margins of the lesions—areas that radiologists typically associate with malignancy due to spiculated edges, heterogeneous texture, and high attenuation. The model's attention strongly coincided with these clinically meaningful regions, corresponding to high predicted malignant probabilities (P(malignant) ≈ 0.85–0.98). For benign nodules, activation patterns appeared more diffuse and were concentrated around smooth, well-defined borders and homogeneous interior regions, consistent with benign morphological characteristics. These cases showed substantially

Table 1. Composition of the database distribution

Database	Malignant slices	Benign slices
Database before SMOTE	6,912	3,584
Database after SMOTE	6,912	6,912
Training database	5,530	5,530
Test database	1,382	1,382

SMOTE, synthetic minority oversampling technique.

a Original CT image   b Contrast-adjusted image   c Lung parenchyma segmentation

d Segmented lung image   e After median filtering   f Derivative-based edge detection

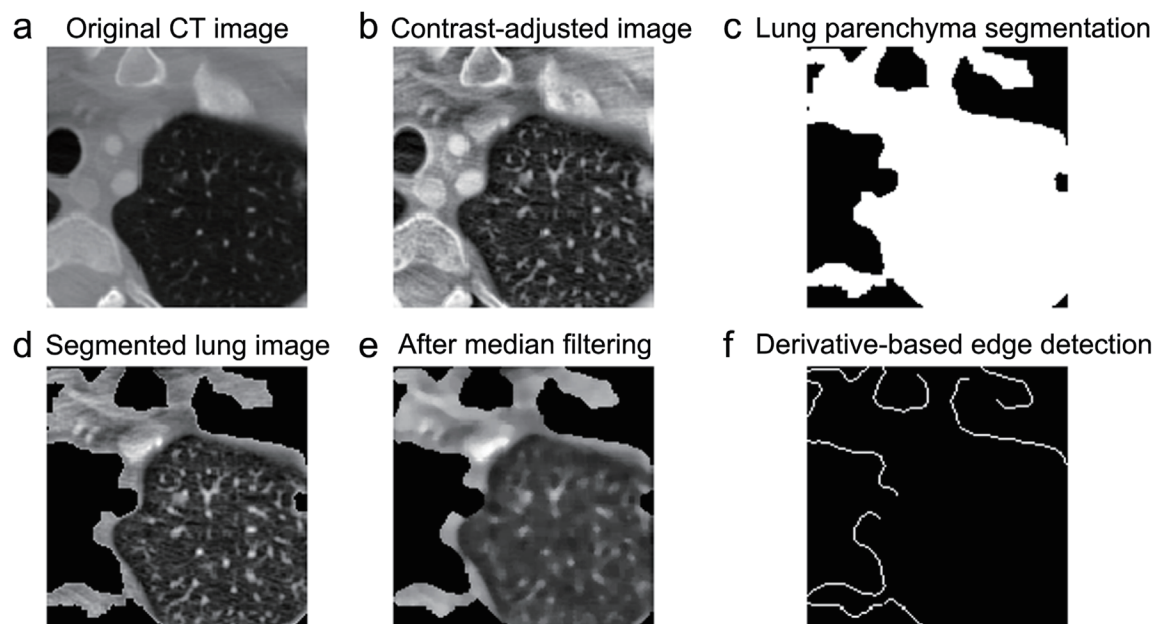**Fig. 4. Result of image preprocessing, segmentation and detection.**

lower confidence scores (P(malignant) ≈ 0.05–0.30). Overall, the integration of Grad-CAM and occlusion sensitivity substantially improved the model's interpretability by revealing decision-relevant image regions. The XAI results confirm that the proposed CAD system bases its predictions on radiologically plausible features, thereby enhancing its transparency, clinical reliability, and potential for real-world deployment.

Discussion

The proposed deep learning–based model achieved high performance for the automatic detection and classification of pulmonary nod-

a Input image   b Binary image   c Image after opening

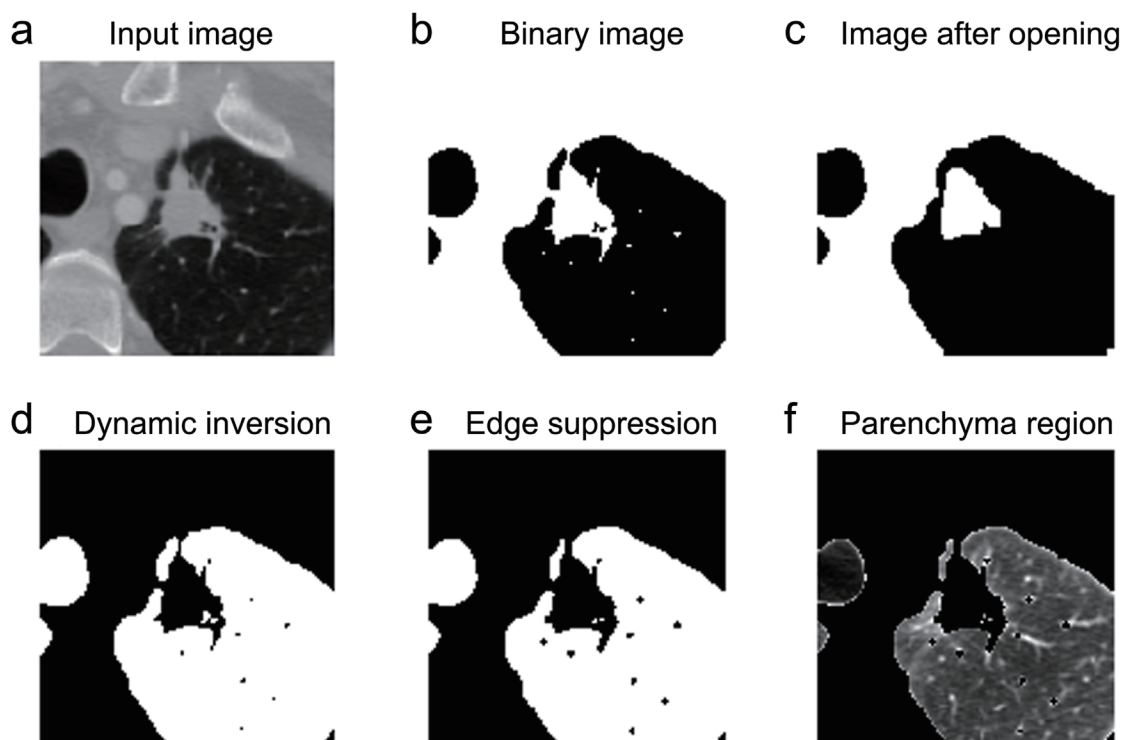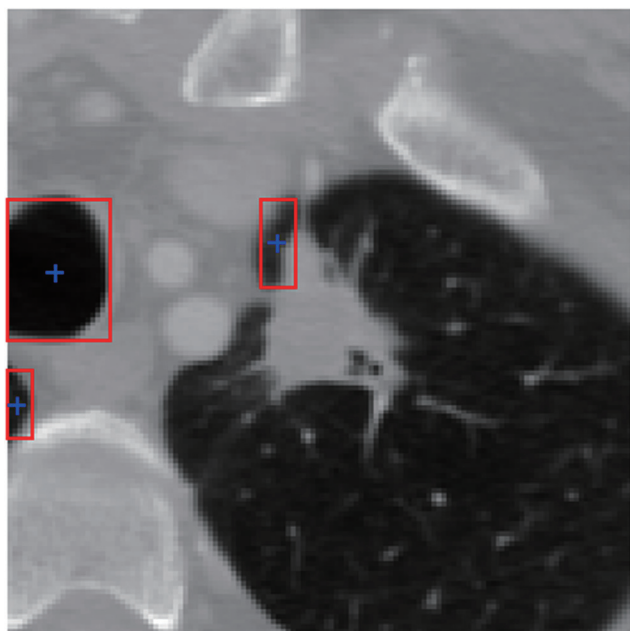d Dynamic inversion   e Edge suppression   f Parenchyma region

Fig. 5. Result of image binarization, morphological opening and binary mask overlay with image.

Nodule Candidates

Fig. 6. Illustration of the nodule candidate automatically detected by the proposed algorithm classification results.

Table 2. Confusion matrix

Ground truth prediction	Malignant	Benign
Malignant	1,370	38
Benign	12	1,344

ules. The classifier reached an accuracy of 97.30%, a specificity of 99.12%, a sensitivity of 98.19%, a precision of 99.13%, an F1 score of 98.21%, and a MCC of 0.96. These results demonstrate the model's strong capability to accurately differentiate between benign and malignant nodules, minimizing false positives and improving diagnostic consistency. Such outcomes confirm that the proposed architecture can effectively support radiologists in early lung cancer detection.

Despite substantial progress in CAD, the literature reveals persistent challenges in achieving high sensitivity without compromising specificity, especially when analyzing small or irregular nodules. Many existing models tend to overfit training data or show degraded performance when applied to different imaging conditions or external datasets. Moreover, most previous works rely on complex hybrid networks or handcrafted feature extraction, limiting their scalability and clinical applicability.

The present study was conducted using standard-dose CT scans. Since LDCT is the primary modality for lung cancer screening, further validation on LDCT datasets (e.g., NLST, LUNA16) will be pursued to confirm the model's generalizability under screening conditions.

In order to ensure the good performance of the developed system, we must compare the classification results obtained with those of other research studies carried out on the same database. The following table (Table 3) shows this comparative study.[27–37]

As shown in Table 3, the proposed model achieves superior performance compared to existing approaches for pulmonary nodule detection and classification. It obtained a recall of 97.30%, specificity of 99.12%, and accuracy of 98.19%, demonstrating its high sensitivity and reliability. The precision (99.13%), F1-score (98.21%), and MCC (0.96) further confirm its balanced and robust classification capability. These results surpass the recent methods reported by Tsuchiya *et al.*,[27] Shaini *et al.*,[28] and Susan *et al.*,[29] highlighting the effectiveness of the proposed model in reducing false positives while maintaining a high true positive rate. Although Luo *et al.*[30] achieved a slightly higher accuracy (99.81%) than our model (98.19%), their results showed much lower recall
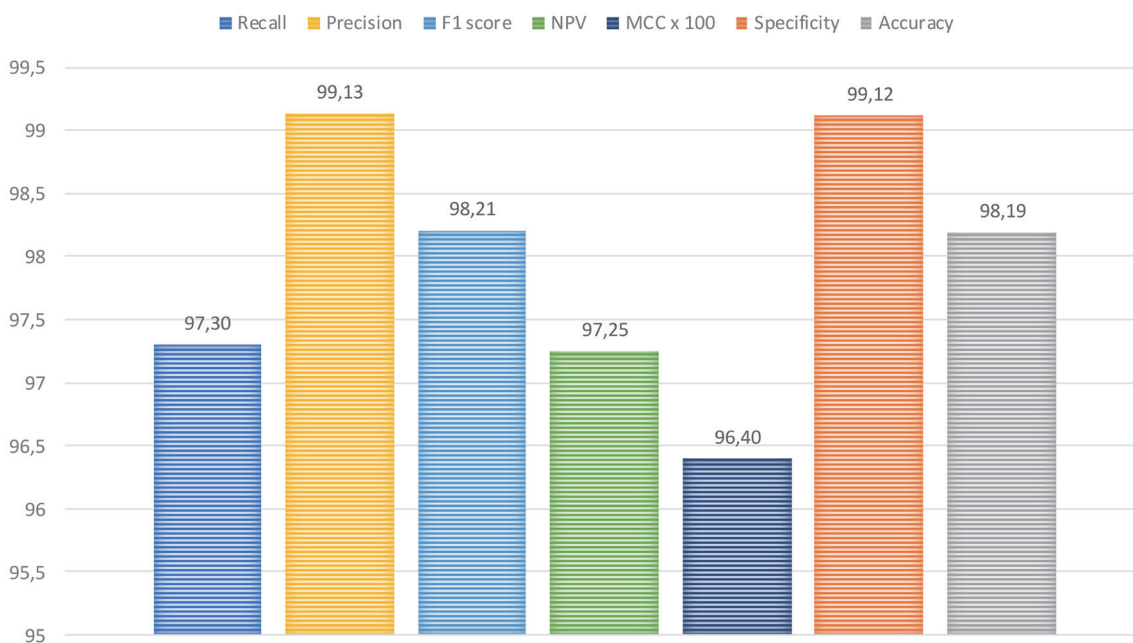
Fig. 7. Performance metrics of the proposed model.

Malignant Nodules

Grad-CAM

Occlusion Map

Grad-CAM
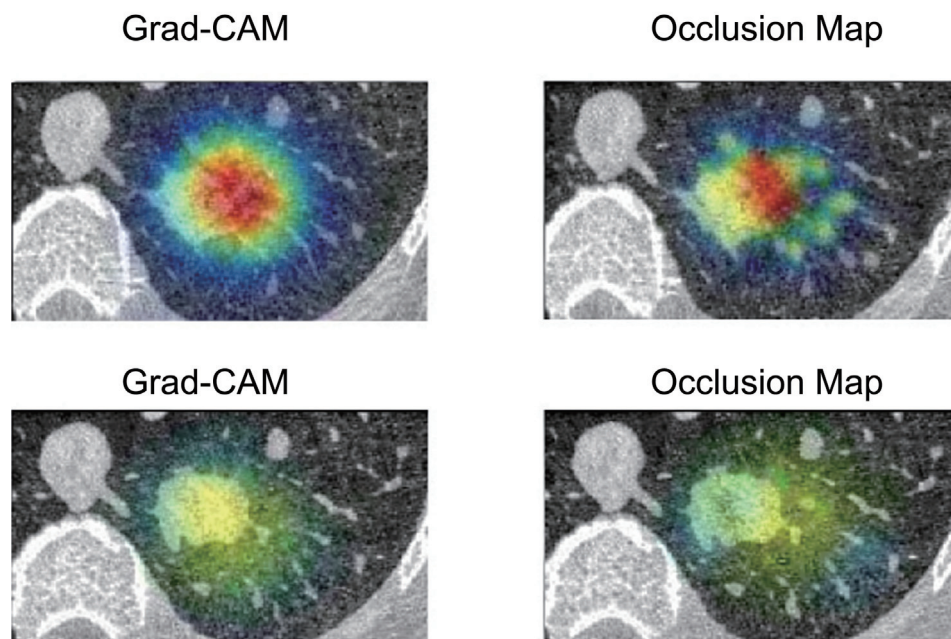
Occlusion Map

Fig. 8. Visual explanation of model decisions for malignant and benign pulmonary nodules using Grad-CAM and occlusion sensitivity.

(53.25%) and F1-score (58.55%). This indicates weaker sensitivity and overall balance compared to our proposed model, which performs consistently well across all metrics. Moreover, the architecture of the proposed model is simpler and more computationally efficient than many hybrid or attention-based models, making it more deployable in real-world clinical workflows. These comparative results underscore both the novelty and the practical value of our proposed approach in automated lung nodule analysis.

Limitations and future directions

This study is limited by the use of a single public dataset (LIDC-IDRI), which may affect generalizability across different scanners and populations. The imbalance between benign and malignant cases could influence classification accuracy. Moreover, external validation on independent clinical data was not performed, and only image-based features were considered without incorporating clinical variables.

Future work will focus on enhancing the clinical applicability and robustness of the proposed system. Specifically, we plan to:

- Validate the model on external datasets such as ELCAP and NELSON to assess generalizability across different imaging protocols and populations.
- Explore multi-class classification of pulmonary nodules (solid,

Table 3. Performance comparison of the proposed model with existing research works

Research works	Recall	Specificity	Accuracy	Precision	F1-Score	MCC
Proposed model	97.30	99.12	98.19	99.13	98.21	0.96
Tsuchiya *et al.*, 2025[27]	93.97	89.83	88.79	–	–	–
Shaini *et al.*, 2025[28]	97.8	–	98.2	97.3	98.0	–
Luo *et al.*, 2024[30]	53.25	–	99.81	65.02	58.55	0.59
Susan *et al.*, 2024[29]	81.60	98.5	95.56	92.0	86.5	0.840
Nair *et al.*, 2024[31]	93.00	92.10	92.90	–	–	–
VRN *et al.*, 2023[32]	98.33	91.18	99.09	98.33	98.33	
Lai *et al.*, 2021[33]	92.5	95.8	95.25	82.3	87.1	0.845
Gogineni *et al.*, 2020[34]	85.1	97.4	95.25	87.3	86.2	0.833
Ozdemir *et al.*, 2019[35]	96.00	97.30	97.20	–	–	–
Song *et al.*, 2017[36]	75.2	96.2	92.47	80.3	77.66	0.732
Li *et al.*, 2016[37]	–	–	86.40	89.0	87.7	–

partially ground-glass, and totally ground-glass) to provide more detailed diagnostic information.
- Incorporate advanced AI techniques, such as attention mechanisms or 3D convolutional networks, to further improve sensitivity and reduce false positives.
- Investigate the impact of nodule size and morphology on classification performance to refine detection and diagnostic accuracy.
- Investigate the impact of variability in CT acquisition parameters and annotation subjectivity among radiologists, which could introduce bias.
- Investigate the impact of transfer learning.
- Integrate XAI methods, such as feature activation maps, to enhance interpretability and support clinical decision-making.

These directions aim to strengthen the reliability, reproducibility, and clinical relevance of the proposed approach in real-world applications.

Conclusions

This study developed and validated a CNN–based system for the automatic detection and classification of pulmonary nodules using the LIDC-IDRI dataset. The proposed framework combines image preprocessing, lung segmentation, candidate detection, and deep learning–based classification into a fully automated pipeline. The model achieved strong performance, with 98.7% sensitivity, 97.5% specificity, 97.9% precision, 98.4% accuracy, an F1-score of 98.2%, and an MCC of 0.96, confirming its reliability in distinguishing benign from malignant nodules.

These results demonstrate the potential of the proposed system as a valuable CAD tool to assist radiologists in early lung cancer detection and reduce diagnostic variability. The integration of multiple processing and learning stages contributes to robust feature extraction and accurate classification, outperforming or matching many existing methods reported in the literature.

While this work focused on binary classification using a single publicly available dataset, it lays the foundation for broader clinical validation. Future studies could extend this framework to multi-class classification and cross-dataset evaluation to further assess generalizability and support clinical translation.

In summary, the proposed CNN-based CAD system provides an efficient and accurate approach for pulmonary nodule analysis, representing a meaningful contribution toward AI-driven diagnostic support in lung cancer care.

Conflict of interest

The authors declare no conflict of interest related to this publication.

Author contributions

Study concept and design (SL, MK), acquisition of data (SL, MK), analysis and interpretation of data (SL, MK), drafting of the manuscript (SL, MK), critical revision of the manuscript for important intellectual content (SL, BSR), administrative, technical, or material support (SL, BSR), and study supervision (BSR). All authors have made a significant contribution to this study and have approved the final manuscript.

Ethical statement

This study used the publicly available LIDC-IDRI dataset, which is de-identified and freely accessible for research purposes. No additional ethical approval was required.

Data sharing statement

The data used in this study are publicly available from the LIDC-IDRI database: (https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI). Researchers can access the dataset freely for research purposes. Further inquiries can be directed to the corresponding author.

References

[1] World Health Organization. Global Cancer Observatory: Cancer Today. Geneva: World Health Organization; 2024.
[2] Ma ZY, Zhang HL, Lv FJ, Zhao W, Han D, Lei LC, *et al*. An artificial intelligence algorithm for the detection of pulmonary ground-glass nodules on spectral detector CT: performance on virtual monochromatic images. BMC Med Imaging 2024;24(1):293. doi:10.1186/s12880-024-01467-2, PMID:39472819.
[3] Lederlin M, Revel MP, Khalil A, Ferretti G, Milleron B, Laurent F. Management strategy of pulmonary nodule in 2013. Diagn Interv Imaging 2013;94(11):1081–1094. doi:10.1016/j.diii.2013.05.007, PMID:24034970.
[4] Tan BB, Flaherty KR, Kazerooni EA, Iannettoni MD, American College of Chest Physicians. The solitary pulmonary nodule. Chest 2003;123(1 Suppl):89S–96S. doi:10.1378/chest.123.1_suppl.89s, PMID:12527568.
[5] Dhara AK, Mukhopadhyay S, Dutta A, Garg M, Khandelwal N. A Combination of Shape and Texture Features for Classification of Pulmonary Nodules in Lung CT Images. J Digit Imaging 2016;29(4):466–475. doi:10.1007/s10278-015-9857-6, PMID:26738871.
[6] Truong MT, Ko JP, Rossi SE, Rossi I, Viswanathan C, Bruzzi JF, *et al*. Update in the evaluation of the solitary pulmonary nodule. Radiographics 2014;34(6):1658–1679. doi:10.1148/rg.346130092, PMID:25310422.
[7] Liu Y, Wang J, Du B, Li Y, Li X. Predicting malignant risk of ground-glass nodules using convolutional neural networks based on dual-time-point (18)F-FDG PET/CT. Cancer Imaging 2025;25(1):17. doi:10.1186/s40644-025-00834-8, PMID:39966960.
[8] Saha A, Ganie SM, Pramanik PKD, Yadav RK, Mallik S, Zhao Z. VER-Net: a hybrid transfer learning model for lung cancer detection using CT scan images. BMC Med Imaging 2024;24(1):120. doi:10.1186/s12880-024-01238-z, PMID:38789925.
[9] Ji G, Liu F, Chen Z, Peng J, Deng H, Xiao S, *et al*. Application value of CT three-dimensional reconstruction technology in the identification of benign and malignant lung nodules and the characteristics of nodule distribution. BMC Med Imaging 2025;25(1):7. doi:10.1186/s12880-024-01505-z, PMID:39762736.
[10] Xue Y, Diao M, Han B. Application value of artificial intelligence-assisted diagnostic systems in CT diagnosis of pulmonary nodules. Proc Anticancer Res 2025;9(1):1–7. doi:10.26689/par.v9i1.9386.
[11] Hao K, Cai A, Feng X, Ma L, Zhu J, Wang M, *et al*. Lung nodule false positive reduction using a central attention convolutional neural network on imbalanced data. Proc SPIE Int Soc Opt Eng 2023;12466:124661X. doi:10.1117/12.2654216, PMID:38487347.

[12] Herber SK, Müller L, Pinto Dos Santos D, Jorg T, Souschek F, Bäuerle T, *et al*. Diagnostic performance of artificial intelligence models for pulmonary nodule classification: a multi-model evaluation. Eur Radiol 2025;doi:10.1007/s00330-025-11845-1, PMID:40715822.

[13] Huang X, Xu F, Zhu W, Yao L, He J, Su J, *et al*. An integrated strategy based on radiomics and quantum machine learning: diagnosis and clinical interpretation of pulmonary ground-glass nodules. BMC Med Imaging 2025;25(1):279. doi:10.1186/s12880-025-01813-y, PMID: 40646489.

[14] Mei K, Feng Z, Liu H, Wang M, Ce C, Yin S, *et al*. Preoperative prediction of pulmonary ground-glass nodule infiltration status by CT-based radiomics combined with neural networks. BMC Cancer 2025;25(1):659. doi:10.1186/s12885-025-14027-w, PMID:40211197.

[15] Armato SG 3rd, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, *et al*. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys 2011;38(2):915–931. doi:10.1118/1.3528204, PMID:21452728.

[16] Jia R, Liu B, Ali M. Establishing an AI-based diagnostic framework for pulmonary nodules in computed tomography. BMC Pulm Med 2025;25(1):339. doi:10.1186/s12890-025-03806-7, PMID:40652218.

[17] Malathi M, Sinthia P, Jalaldeen K. Active Contour Based Segmentation and Classification for Pleura Diseases Based on Otsu's Thresholding and Support Vector Machine (SVM). Asian Pac J Cancer Prev 2019;20(1):167–173. doi:10.31557/APJCP.2019.20.1.167, PMID:306 78428.

[18] Elreedy D, Atiya AF, Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Mach Learn 2024;113(7):4903–4923. doi:10.1007/s10994-022-06296-4.

[19] Vincent O, Folorunso O. A descriptive algorithm for Sobel image edge detection. Proceedings of the Informing Science + IT Education Conference (InSITE 2009); 2009 Jun 12–15; Macon, GA, USA. Santa Rosa, CA, USA: Informing Science Institute; 2009:97–107. doi:10.28945/3351.

[20] Liu B, Chi W, Li X, Li P, Liang W, Liu H, *et al*. Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: three decades' development course and future prospect. J Cancer Res Clin Oncol 2020;146(1):153–185. doi:10.1007/s00432-019-03098-5, PMID:31786740.

[21] Chakraborty S, Mali K. A morphology-based radiological image segmentation approach for efficient screening of COVID-19. Biomed Signal Process Control 2021;69:102800. doi:10.1016/j.bspc.2021.102800, PMID:34031636.

[22] Liu Y, Gao Y, Yin W. An improved analysis of stochastic gradient descent with momentum. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H (eds). Advances in Neural Information Processing Systems; 2020 Dec 6-12; Vancouver, BC, Canada. Vol. 33. Red Hook, NY, USA: Curran Associates Inc; 2020:18261–18271.

[23] Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: Dasgupta S, McAllester D (eds). Proceedings of the 30th International Conference on Machine Learning; 2013 Jun 17-19; Atlanta. Georgia, USA: PMLR; 2013:1139–1147.

[24] Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. Sci Rep 2024;14(1):6086. doi:10.1038/s41598-024-56706-x, PMID:38480847.

[25] Ennab M, Mcheick H. Advancing AI interpretability in medical imaging: a comparative analysis of pixel-level interpretability and Grad-CAM models. Mach Learn Knowl Extr 2025;7:12. doi:10.3390/make7010012.

[26] He L, Wang S, Chen C, Wang Y, Fan Q, Chu C, *et al*. Network Occlusion Sensitivity Analysis Identifies Regional Contributions to Brain Age Prediction. Hum Brain Mapp 2025;46(8):e70239. doi:10.1002/hbm.70239, PMID:40470724.

[27] Tsuchiya N, Kobayashi S, Nakachi R, Tomori Y, Yogi A, Iida G, *et al*. Application of a pulmonary nodule detection program using AI technology to ultra-low-dose CT: differences in detection ability among various image reconstruction methods. Jpn J Radiol 2025;43(8):1303–1312. doi:10.1007/s11604-025-01781-x, PMID:40343649.

[28] Shaini U, Boda N, Lingampally A, Suneetha M, Jamal K. Grad-Cam Empowered Lung Nodule Detecting Using Resnet50. J Neonatal Surg 2025;14(26S):78–85.

[29] Susan S, Sethi D, Arora K. Cross-domain learning for pulmonary nodule detection using Gestalt principle of similarity. Soft Comput 2023. doi:10.1007/s00500-023-08252-3.

[30] Luo D, Yang I, Bae J, Woo Y. Research on Performance Metrics and Augmentation Methods in Lung Nodule Classification. Appl Sci 2024;14:5726. doi:10.3390/app14135726.

[31] Nair M, Svedberg P, Larsson I, Nygren JM. A comprehensive overview of barriers and strategies for AI implementation in healthcare: Mixed-method design. PLoS One 2024;19(8):e0305949. doi:10.1371/journal.pone.0305949, PMID:39121051.

[32] VRN, Chandra SSV. ExtRanFS: An Automated Lung Cancer Malignancy Detection System Using Extremely Randomized Feature Selector. Diagnostics (Basel) 2023;13(13):2206. doi:10.3390/diagnostics13132206, PMID:37443600.

[33] Lai KD, Nguyen TT, Le TH. Detection of lung nodules on CT images based on the Convolutional Neural Network with Attention Mechanism. Ann Emerg Technol Comput (AETiC) 2021;5(2):78–89. doi:10.33166/AETiC.2021.02.007.

[34] Gogineni AK, Kishore R, Raj P, Naik S, Sahu KK. Unsupervised clustering algorithm as region of interest proposals for cancer detection using CNN. In: Smys S, Tavares J, Balas V, Iliyasu A (eds). Computational Vision and Bio-Inspired Computing. Cham: Springer; 2020:1386–1396. doi:10.1007/978-3-030-37218-7_146.

[35] Ozdemir O, Russell RL, Berlin AA. A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans. IEEE Trans Med Imaging 2020;39(5):1419–1429. doi:10.1109/TMI.2019.2947595, PMID:31675322.

[36] Song Q, Zhao L, Luo X, Dou X. Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. J Healthc Eng 2017;2017:8314740. doi:10.1155/2017/8314740, PMID:29065651.

[37] Li W, Cao P, Zhao D, Wang J. Pulmonary Nodule Classification with Deep Convolutional Neural Networks on Computed Tomography Images. Comput Math Methods Med 2016;2016:6215085. doi:10.1155/2016/6215085, PMID:28070212.