



OPEN ACCESS

EDITED BY

Eric Edward Sigmund,
New York University, United States

REVIEWED BY

Kei Shing Ng,
The University of Hong Kong, Hong Kong
SAR, China
Zheng Yuan,
China Academy of Chinese Medical Sciences,
China
Xiaofeng Li,
Tianjin Medical University Cancer Institute and
Hospital, China

*CORRESPONDENCE

Ouwen Li
✉ apocalypse1523@163.com

†These authors have contributed equally to
this work

RECEIVED 28 April 2024

ACCEPTED 31 July 2024

PUBLISHED 19 August 2024

CITATION

Xu H, Yu Y, Chang J, Hu X, Tian Z and Li O
(2024) Precision lung cancer screening
from CT scans using a VGG16-based
convolutional neural network.
Front. Oncol. 14:1424546.
doi: 10.3389/fonc.2024.1424546

COPYRIGHT

© 2024 Xu, Yu, Chang, Hu, Tian and Li. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Precision lung cancer screening from CT scans using a VGG16-based convolutional neural network

Hua Xu^{1†}, Yuanyuan Yu^{2,3†}, Jie Chang³, Xifeng Hu³,
Zitong Tian³ and Ouwen Li^{4*}

¹Department of Infection Control, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Shandong, Jinan, China, ²Data Science Institute, Shandong University, Jinan, Shandong, China, ³Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan, China, ⁴International Center, Jinan Foreign Language School, Shandong, Jinan, China

Objective: The research aims to develop an advanced and precise lung cancer screening model based on Convolutional Neural Networks (CNN).

Methods: Based on the health medical big data platform of Shandong University, we developed a VGG16-Based CNN lung cancer screening model. This model was trained using the Computed Tomography scans data of patients from Pingyi Traditional Chinese Medicine Hospital in Shandong Province, from January to February 2023. Data augmentation techniques, including random resizing, cropping, horizontal flipping, color jitter, random rotation and normalization, were applied to improve model generalization. We used five-fold cross-validation to robustly assess performance. The model was fine-tuned with an SGD optimizer (learning rate 0.001, momentum 0.9, and L2 regularization) and a learning rate scheduler. Dropout layers were added to prevent the model from relying too heavily on specific neurons, enhancing its ability to generalize. Early stopping was implemented when validation loss did not decrease over 10 epochs. In addition, we evaluated the model's performance with Area Under the Curve (AUC), Classification accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), Sensitivity, Specificity and F1 score. External validation used an independent dataset from the same hospital, covering January to February 2022.

Results: The training and validation loss and accuracy over iterations show that both accuracy metrics peak at over 0.9 by iteration 15, prompting early stopping to prevent overfitting. Based on five-fold cross-validation, the ROC curves for the VGG16-Based CNN model, demonstrate an AUC of 0.963 ± 0.004 , highlighting its excellent diagnostic capability. Confusion matrices provide average metrics with a classification accuracy of 0.917 ± 0.004 , PPV of 0.868 ± 0.015 , NPV of 0.931 ± 0.003 , Sensitivity of 0.776 ± 0.01 , Specificity of 0.962 ± 0.005 and F1 score of 0.819 ± 0.008 , respectively. External validation confirmed the model's robustness across different patient populations and imaging conditions.

Conclusion: The VGG16-Based CNN lung screening model constructed in this study can effectively identify lung tumors, demonstrating reliability and effectiveness in real-world medical settings, and providing strong theoretical and empirical support for its use in lung cancer screening.

KEYWORDS

lung cancer screening, medical image recognition, computed tomography scans, VGG16 architecture, convolutional neural network

1 Background

Lung cancer is one of the leading causes of cancer-related deaths worldwide, with an estimated 1.8 million deaths, accounting for 18.7% of the total (1–5). Moreover, both the incidence and mortality rates of lung cancer are on the rise, particularly in developing and middle-income countries (6, 7). Given the vast population of China, this implies a large number of new cases and deaths annually, placing immense strain on the healthcare system. In Taiwan, a study assessed low-dose lung computed tomography (CT) screening criteria among Asian ethnic groups, finding that risk-based strategies more effectively identify high-risk non-smokers, emphasizing the need to optimize screening criteria for better early detection (8). Currently, lung CT scans have become a primary tool in modern medicine for early screening of lung diseases, especially lung cancer (9–12). Henschke et al. demonstrated that CT screening can significantly improve the detection rate of early-stage lung cancer compared to traditional screening methods, which is crucial for patient survival rates (13). This further proves the significance of CT screening in the early detection and treatment of lung cancer. However, these place rigorous demands on a physician's interpretive abilities. To effectively interpret these high-resolution images, physicians must possess extensive clinical medical knowledge and expertise in radiology. Subjectivity and variability in interpretation can lead to inconsistent diagnoses and even misdiagnosis. Therefore, there is a pressing need to seek more advanced methods to assist physicians in achieving more accurate and efficient CT screening interpretations, ultimately enhancing diagnostic outcomes and patients' medical experiences.

In recent years, deep learning, especially Convolutional Neural Networks (CNN), has been proven to be an incredibly effective tool for automating and optimizing the analysis process of medical imaging. It has also been widely applied in various medical imaging systems such as Magnetic Resonance Imaging (MRI), CT, and X-rays. Litjens et al. conducted in-depth research on how CNN can improve the accuracy of image classification, segmentation, and disease detection (14). Rajpurkar et al. applied a deep learning model called CheXNeXt to chest X-ray images, demonstrating that CheXNeXt's performance in automatic diagnosis of various chest diseases is comparable to that of practicing radiologists, and in some cases, it even surpasses the performance of human experts (15). These studies provide

substantial evidence for the application of deep learning techniques in clinical practice, but still face a series of challenges in areas such as data insufficiency, model overfitting, and interpretability of diagnostic results.

Motivated by the urgent need to improve early lung cancer detection and diagnosis, this study developed a lung cancer screening model based on the CNN architecture. To enhance generalization and reduce overfitting, we employed several strategies. Data augmentation techniques, including random resizing, cropping, horizontal flipping, color jitter, and random rotation, were applied to increase the diversity of the training data. We used five-fold cross-validation to robustly assess performance by ensuring each fold maintained the same ratio of malignant to non-malignant images. The model was fine-tuned with an SGD optimizer (learning rate 0.001, momentum 0.9, and L2 regularization) and a learning rate scheduler. Dropout layers were added to prevent the model from relying too heavily on specific neurons, thus enhancing its ability to generalize. Early stopping was implemented to halt training when validation loss did not decrease over 10 consecutive epochs. In addition, we used a comprehensive set of evaluation metrics, including Area Under the Curve (AUC), Classification accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), Sensitivity, Specificity and F1 score to ensure the model's exceptional performance across diverse criteria. Our aim is to leverage the powerful capabilities of CNNs to provide a cutting-edge, accurate, and effective lung cancer screening method, integrating it into daily medical workflows to improve screening accuracy.

2 Methods

Figure 1 illustrates the comprehensive workflow of our lung cancer screening model using the VGG16 architecture.

2.1 Study population

2.1.1 Training population

2.1.1.1 Construction of the imaging dataset for patients with malignant lung tumors

Based on the health medical big data platform of Shandong University, we gathered a comprehensive dataset from Pingyi

Traditional Chinese Medicine Hospital in Shandong Province, including lung CT scans and EHR data such as age, gender, diagnostic variables, pathological reports, and patient identifiers, collected over January and February 2023 (as shown in Figure 1A). The pathological report clearly indicated whether the patient had a malignant lung tumor. All patient data were anonymized and collected using encrypted identification numbers to protect privacy and comply with ethical standards. As shown in Figure 1B, the inclusion criteria for this study were: 1) Patients with malignant lung

tumors who had undergone procedures such as puncture, bronchoscopy, or surgery to confirm the definitive pathological diagnosis. 2) Patients who must have undergone CT scans prior to percutaneous puncture, bronchoscopy biopsy or surgery. 3) Patients who had lung nodules discovered during the CT scan examination. The exclusion criteria were: 1) Patients with concurrent other lung lesions. 2) Patients with a history of lung diseases. 3) Patients with metastatic tumors. Ultimately, a total of 141 patients with malignant lung tumors were included in this study.

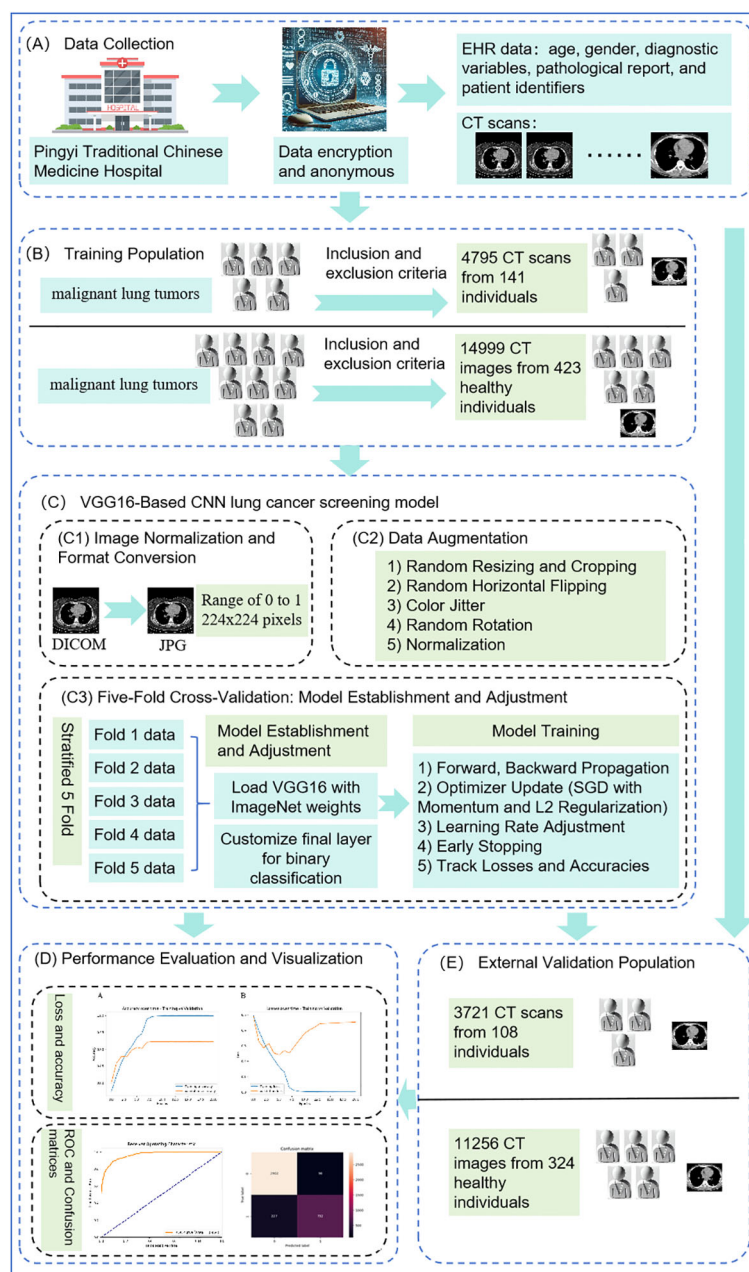


FIGURE 1 Comprehensive workflow of VGG16-based convolutional neural network. (A) shows data collection, (B) indicates the process of identifying the training population, (C) illustrates the construction of the VGG16-Based CNN lung cancer screening model, (D) depicts model evaluation and visualization, and (E) represents external validation.

2.1.1.2 Construction of the dataset for non-malignant lung tumor patients

For the dataset of non-malignant lung tumor patients, we included same lung CT images and detailed metadata from patients at Pingyi Traditional Chinese Medicine Hospital, collected between January and February 2023. The inclusion criteria required patients to have undergone CT scans. Exclusion criteria (Figure 1B), consistent with those for the malignant lung tumor datasets, included patients with concurrent lung lesions, a history of lung diseases, or metastatic tumors. Using a random match method at a ratio of 1:3, a total of 423 non-malignant lung tumor patients were included.

The final dataset was composed of 4795 CT images from 141 individuals with malignant lung tumors and 14999 CT images from 423 individuals with non-malignant lung tumor.

2.1.2 External validation population

An external validation dataset was obtained from Pingyi Traditional Chinese Medicine Hospital in Shandong Province, comprising imaging data collected between January and February 2022 (Figure 1E). The inclusion and exclusion criteria were consistent with those used for the primary dataset. This external dataset included 108 patients with malignant lung tumors and 324 patients with non-malignant lung tumors, providing a robust basis for validating the performance of the developed model.

In clinical research, especially with smaller sample sizes, sparse data bias can significantly affect the reliability of statistical estimates. By using a 1:3 random matching design, we increased the sample size of the malignant group, thus reducing the impact of sparse data bias and ensuring more stable and reliable estimates of effect sizes and other statistical parameters (16). In addition, random matching helps to minimize the potential biases that can arise from non-random selection processes. By randomly selecting non-malignant cases at a 1:3 ratio, we aimed to ensure a representative sample of the broader population, thus enhancing the generalizability of our results (17).

Our study utilized both plain and contrast-enhanced CT scans. The choice between these two techniques was based on specific clinical indications and the patient's condition at the time of imaging, ensuring that the most appropriate diagnostic approach was utilized for each case. The imaging was performed using three different models of CT scanners, including a Philips 256-slice CT, a 64-slice CT, and a 16-slice CT. This diverse array of scanning equipment enabled a comprehensive assessment of the conditions under study, ensuring a robust analysis through varied imaging capabilities.

This study was approved by the Institutional Review Board of Pingyi Traditional Chinese Medicine Hospital in Shandong Province, China. The ethical approval number for our study is PYX2YYLLWYh2023030602. It is important to note that no informed consent was required from the participants for this study, as only anonymized lung CT images and detailed metadata were used. The anonymization process was rigorously conducted before the data was accessed for research purposes, ensuring that all personal identifiers were removed to protect patient confidentiality.

2.2 VGG16-based CNN lung cancer screening model

The Convolutional Neural Network (CNN) is a cornerstone algorithm in deep learning, especially adept at image processing tasks. Since the advent of LeNet in 1998, CNN has become mainstream in computer vision tasks (18). This study attempts to use the VGG16 model to construct a lung cancer screening model. The choice of the VGG16 architecture for lung cancer screening is driven by its distinguished capabilities in handling complex image data and its historical success in diverse image recognition tasks (18). VGG16's architectural depth and uniformity are ideal for medical imaging, where precision and reliability are paramount. The architecture features 13 convolutional layers that are exceptionally effective at extracting multi-scale features, a fundamental requirement for identifying subtle and critical anomalies in medical images (19, 20). Moreover, VGG16's robustness and adaptability in processing new and varying datasets make it an exemplary choice for the dynamic requirements of medical diagnostics, as demonstrated by its proven efficacy in diagnosing conditions such as pneumonia from chest X-rays, papillary thyroid carcinomas from cytological images, and brain tumors from MRI scans (15, 21, 22). This combination of deep learning efficiency and versatility underscores why VGG16 is uniquely suited for developing a lung cancer screening model (Figure 1C).

2.2.1 Data preparation

2.2.1.1 Image normalization and format conversion

We use *pydicom* to extract DICOM images and normalize their intensity values to a range of 0 to 1, which is essential for consistent CNN performance. Subsequently, we utilize the *Figure* module from *matplotlib* to display and manipulate image plots, aiding in converting these images to JPG format (23, 24).

2.2.2 Image pre-processing

2.2.2.1 Resizing and normalization

All images were resized to 224x224 pixels to meet the input requirements of the VGG16 model and were normalized to a range of 0 to 1 to standardize the input (19).

2.2.2.2 Data augmentation

To enhance the model's generalization capabilities and prevent overfitting, data augmentation techniques were applied using the *torchvision.transforms* library. The techniques used included 1) Random Resizing and Cropping: Images were randomly resized and cropped to provide a variety of image sizes and perspectives to the model. 2) Random Horizontal Flipping: Images were randomly flipped horizontally to make the model invariant to left-right orientation. 3) Color Jitter: Random adjustments to brightness, contrast, saturation, and hue to introduce variability. 4) Random Rotation: Images were randomly rotated to make the model invariant to orientation. 5) Normalization: Pixel values were normalized to ensure consistent intensity values across all images (19). By applying these augmentations, we increased the diversity of

our training data, prevented overfitting, and ensured that the model could generalize well to unseen data.

2.2.3 Five-fold cross-validation

We implemented the five-fold cross-validation method to reduce the risk of overfitting to a particular subset of the data. This method involved splitting the dataset into five parts, ensuring each fold maintained the same ratio of malignant to non-malignant images. Each fold served as a validation set once, while the remaining four folds constituted the training set (25).

2.2.4 Model establishment and adjustment

We used the VGG16 architecture, which is a well-known deep CNN model pre-trained on the ImageNet dataset. VGG16 consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. Given our task of binary classification (malignant lung tumors vs. non-malignant lung tumors), we modified the final fully connected layer of the VGG16 model to output two classes (19). Specific layers of the pre-trained model were unfrozen to allow fine-tuning during training (26).

2.2.5 Model training

During each iteration, the model underwent forward and backward propagation on the training data. The SGD optimizer, with a learning rate of 0.001 and momentum of 0.9, and weight decay of 0.0005 (L2 regularization), was used to update the model weights. The binary cross-entropy loss function was employed to calculate the loss and both training and validation losses and accuracies were recorded for performance evaluation (27).

2.2.6 Preventing overfitting

Several strategies were employed to prevent overfitting, including: 1) Dropout layers were added to the model to randomly drop neurons during training, which helps in preventing overfitting. 2) A learning rate scheduler was used to dynamically adjust the learning rate during training, helping to fine-tune the model and avoid overfitting. 3) Early stopping was implemented, terminating training if there was no significant decrease in validation loss over 10 consecutive epochs (28).

2.2.7 Performance evaluation and results visualization

Training and validation loss and accuracy were recorded at the end of each epoch. These metrics were plotted to visualize the learning progress and identify potential overfitting or underfitting issues (27). As shown in Figure 1D, the ROC curve was plotted to evaluate the trade-off between sensitivity and specificity at various threshold settings (29). The area under the ROC curve (AUC) was calculated to quantify the overall ability of the model to discriminate between positive and negative cases (29). A confusion matrix was generated to provide a detailed breakdown of the model's predictions and their alignment with the actual outcomes. This matrix helped in understanding the distribution of true positives, true negatives, false positives, and false negatives, which is crucial for evaluating the performance of the classification

model (29, 30). At the end of Area Under the Curve (AUC), classification accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), Sensitivity, Specificity and F1 score were calculated as evaluation metrics (30).

2.2.8 External validation

To ensure the model's generalizability and robustness, the trained model was further evaluated on the external validation dataset obtained from Pingyi Traditional Chinese Medicine Hospital in Shandong Province (Figure 1E). The AUC was computed, and results were visualized using ROC curves to provide a comprehensive evaluation of the model's performance on an independent dataset.

2.3 Software details

The code was written in Python, and executed on the Jupyter server at the Health and Medical Big Data Research Institute of Shandong University.

3 Result

3.1 General characteristics of patients with malignant lung tumors

A total of 141 patients with malignant lung tumors were included, consisting of 97 males and 44 females. Their ages ranged from 44 years old to 89 years, with a median age of 69 years and an interquartile range (IQR) of 63 to 74 years. Additionally, 423 individuals with no-lung tumors were included, comprising 225 males and 198 females. The age of those without lung tumors ranged from 20 to 95 years, with a median age of 67 years and an IQR of 54.5 to 76.6 years. The frequency distribution of ages for both groups is detailed in Supplementary Material Tables 1 and 2.

3.2 Results for pulmonary imaging based on CNN

Figure 2 shows the loss and accuracy of the training and validation sets during each iteration. As can be seen from Figure 2A, the accuracy of both the training and validation sets generally increases with the number of iterations. By the time the number of iterations reaches 10, the accuracy of both the training and validation sets is at its highest, with both achieving an accuracy of over 90%. As shown in Figure 2B, when the number of iterations reaches 15, the program determines that the validation error has not improved for 10 consecutive epochs. Therefore, we chose an iteration count of 15 for early stopping. This early stopping strategy helps prevent overfitting and ensures the model's robustness by terminating training when no significant improvement in validation loss is observed over 10 consecutive epochs.

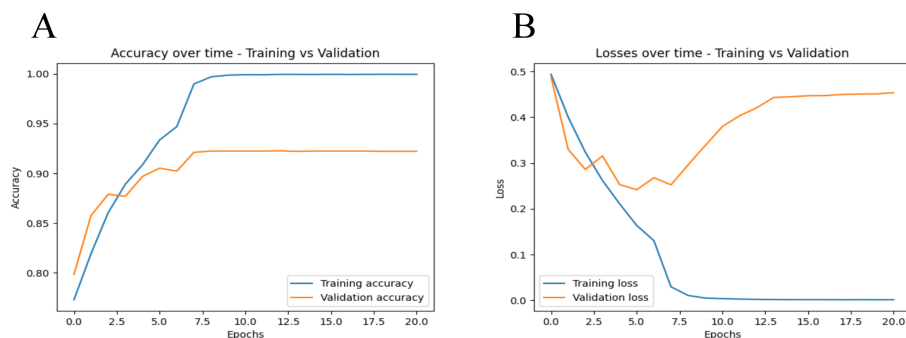


FIGURE 2

Training and validation metrics over iterations. (A) displays the accuracy metrics for both training and validation sets across epochs. The blue line represents the training accuracy, while the orange line represents the validation accuracy. (B) shows the loss values for both the training and validation datasets over each training epoch. The blue line represents the training loss, while the orange line represents the validation loss.

Figure 3- displays the ROC curves for the five-fold cross-validation and the external dataset, while Table 1 lists the evaluation indices along with their 95% Confidence Intervals (CIs) for five-fold cross-validation. The CNN-based pulmonary imaging diagnostic model consistently demonstrated high diagnostic accuracy, evidenced by an average AUC of 0.963 ± 0.004 across five-fold cross-validation, as depicted in Figures 3A–E. This robust performance is presented in the first row of Table 1, where each fold's AUC score approximates 0.96, with narrow 95% confidence intervals. Such consistently high AUC values across multiple validation folds attest to the model's reliable capability to differentiate between malignant and non-malignant cases.

Additionally, the confusion matrices for all 5 runs are provided in the Supplementary Material Figures 1–5. Based on the confusion matrix, the classification accuracy, PPV and NPV were 0.917 ± 0.004 , 0.868 ± 0.015 , and 0.931 ± 0.003 , respectively, with each fold's accuracy, PPV, and NPV along with their 95% CIs, detailed in lines two to four of Table 1. Classification accuracy measures the overall effectiveness of the model in correctly identifying both positive and negative cases, with a reported accuracy of approximately 91.1%. PPV specifically assesses the accuracy of the model's positive predictions, indicating that approximately 86.6% of the lung cancer diagnoses made by the model are accurate. An NPV of 93.1% ensures a high probability that negative diagnoses are correct, minimizing the risk of missed diagnoses. Together, high PPV and NPV highlight the model's reliability, making it suitable for clinical lung cancer screening. According to the confusion matrix of the five-fold cross-validation, the Sensitivity, Specificity and F1 score were 0.776 ± 0.01 , 0.962 ± 0.005 and 0.819 ± 0.008 , respectively, with these results for each fold shown in lines five to seven of Table 1. The sensitivity of 77.6% indicates a robust capability to detect true cases of lung cancer, effectively minimizing the risk of missing diagnoses (false negatives). The specificity of 96.2% demonstrates the model's precision in identifying individuals who do not have lung cancer, significantly reducing the occurrence of false positives. Additionally, an F1 score of 81.9% reflects a well-balanced trade-off between precision and sensitivity, ensuring the model's overall accuracy and reliability in medical diagnostics. This

comprehensive validation across multiple metrics confirms the model's effectiveness in lung cancer detection.

The external validation showed that our model maintained robust performance (Figure 3F), with an AUC and 95%CI of 0.7564 [0.7563, 0.7564]. The result confirms the model's ability to generalize well across different patient populations and imaging conditions.

4 Discussion

Lung cancer has become a significant public health problem in China, with rising incidence and mortality rates as highlighted in the "Cancer incidence and mortality in China, 2016" report and recent data from the American Cancer Society (ACS) (31, 32). With the advancement of technology and medicine, radiological lung screening has become a pivotal means for early detection and evaluation of pulmonary diseases, especially lung cancer. Despite extensive research and practice, existing methods still have limitations. The survival rate for patients with advanced lung cancer remains low (33, 34), and early diagnosis and treatment are becoming increasingly important (35–37). Aberle et al. have pointed out that compared to traditional chest X-rays, lung cancer screenings using low-dose CT can significantly reduce lung cancer mortality rates (38). This pivotal discovery laid a solid scientific foundation for the promotion and application of low-dose CT screening (39–41). However, traditional imaging analysis methods have certain limitations, such as high missed diagnosis rate, misdiagnosis rate, and limited ability to interpret intricate images (42, 43). Therefore, with the increasing complexity and resolution of medical imaging, there's a growing reliance on advanced computer-aided diagnostic systems to assist in understanding and interpreting these images (44–46). Deep learning techniques, especially CNNs, have revolutionized medical imaging analysis (47).

Our study constructed a precise lung cancer screening model based on CNN. Initial preprocessing and data augmentation ensured the quality and consistency of lung cancer imaging data. Overfitting was mitigated through multiple strategies including data

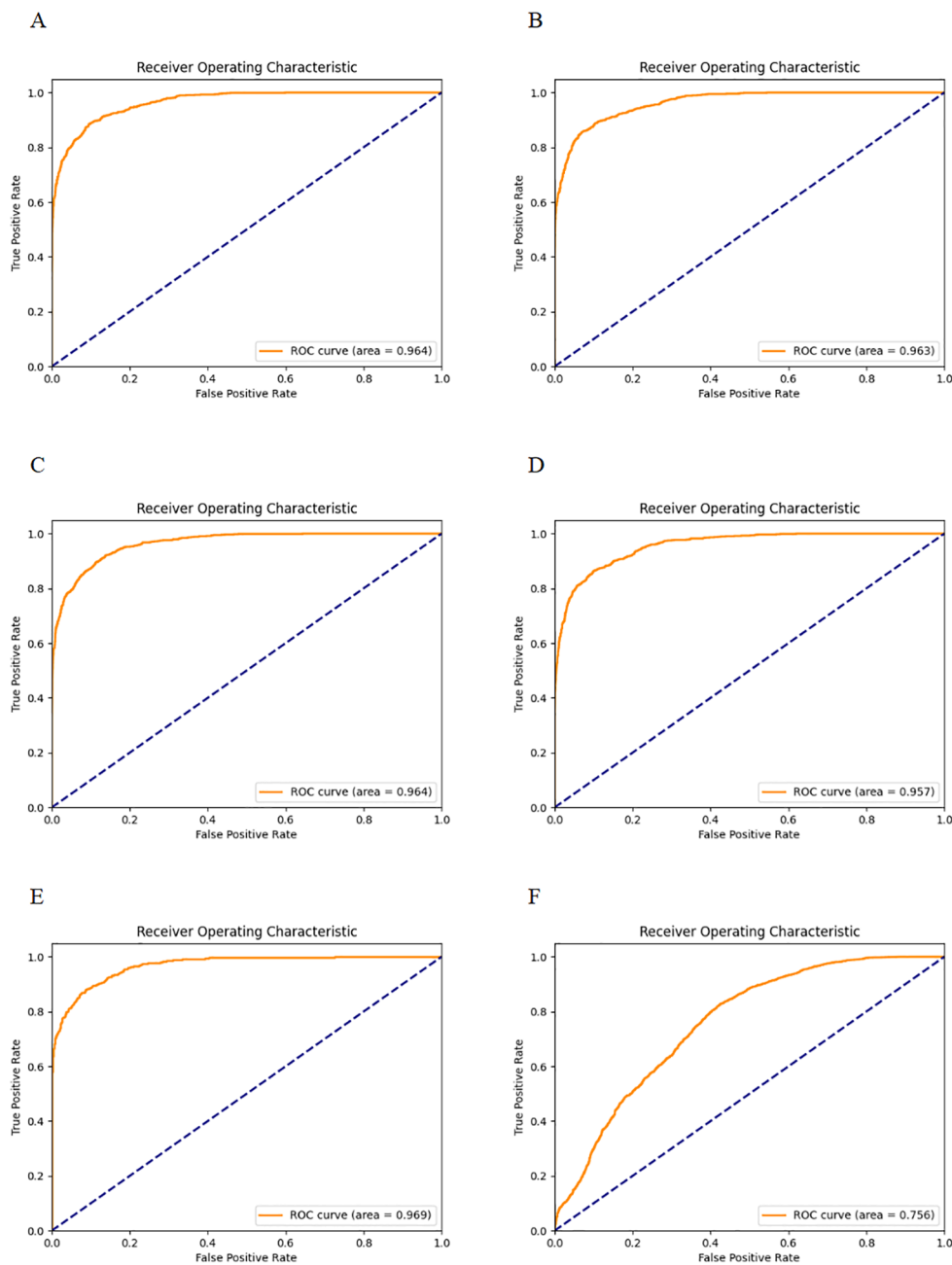


FIGURE 3 ROC curves for five-fold cross-validation and external validation. (A–E) display the ROC curves for each fold of the five-fold cross-validation. (F) shows the ROC curve for the external validation dataset.

augmentation techniques such as random resizing and cropping, random horizontal flipping, color jitter, and random rotation. L2 regularization was applied via weight decay in the optimizer, and a learning rate scheduler dynamically adjusted the learning rate during training. Dropout layers were added to the model to prevent reliance on specific neurons, and early stopping was implemented to halt training if validation loss did not decrease over 10 consecutive epochs. Five-fold cross-validation demonstrated the model’s robustness. Our study achieved an

AUC of 0.963 ± 0.004 on the validation set, indicative of excellent diagnostic capability. This high AUC value demonstrated the model’s robust capacity to distinguish accurately between lung cancer and those without, minimizing both false positives and false negatives. These performances are consistent with a classification accuracy of 0.917 ± 0.004 . Additionally, the PPV, NPV, Sensitivity, Specificity, and F1 score were 0.868 ± 0.015 , 0.931 ± 0.003 , 0.776 ± 0.01 , 0.962 ± 0.005 , and 0.819 ± 0.008 , respectively. These performance metrics collectively demonstrate the model’s

TABLE 1 The evaluation index and its 95% confidence interval for five-fold cross-validation.

Evaluation index[95%CI]	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
AUC	0.9643 [0.9642, 0.9643]	0.9628 [0.9628, 0.9629]	0.9640[0.9639, 0.9640]	0.9566[0.9565, 0.9567]	0.9690[0.9690, 0.9691]
Classification accuracy	0.9179 [0.9089, 0.9261]	0.9179 [0.9089, 0.9261]	0.9154[0.9063, 0.9237]	0.9113[0.9021, 0.9198]	0.9222[0.9134, 0.9301]
PPV	0.8819[0.8582, 0.9021]	0.8747[0.8507,0.8953]	0.8537[0.8289,0.8755]	0.8486[0.8233,0.8709]	0.8798[0.8563,0.8999]
NPV	0.9275[0.9178,0.936]	0.9296[0.9201,0.9381]	0.9331[0.9237,0.9414]	0.9291[0.9195,0.9376]	0.9339[0.9246,0.9421]
Sensitivity	0.7633[0.7354, 0.7891]	0.7716 [0.7440, 0.7971]	0.7852[0.7581, 0.8100]	0.7716 [0.7440, 0.7971]	0.7862[0.7592, 0.8110]
Specificity	0.9673[0.9604, 0.9731]	0.9647 [0.9574, 0.9707]	0.9570[0.9491, 0.9637]	0.9560[0.9481, 0.9628]	0.9657 [0.9585, 0.9716]
F1 score	0.8183[0.8058, 0.8298]	0.8199 [0.8076, 0.8316]	0.8180 [0.8056, 0.8296]	0.8083[0.7957, 0.8202]	0.8304[0.8182, 0.8416]

high degree of diagnostic accuracy and reliability in identifying lung cancer. The classification accuracy of 91.7% indicates a strong overall ability to correctly classify cases as either having lung cancer or not, supported by a high PPV (86.8%) and NPV (93.1%), which ensure that the positive and negative diagnoses made by the model are likely correct. The sensitivity of 77.6% shows that the model is capable of identifying a substantial majority of true positive cases, critical for early and accurate disease detection. Furthermore, the specificity of 96.2% underscores the model's effectiveness in correctly ruling out disease in healthy individuals, reducing the likelihood of unnecessary treatments. Finally, an F1 score of 81.9% reflects a balanced trade-off between precision and recall, validating the model's utility in clinical settings where both detecting cases and avoiding false alarms are equally important. Together, these metrics not only underline the model's capability but also highlight its potential to significantly enhance patient management and treatment outcomes in clinical practice.

Lu et al. created a CNN model for predicting the long-term incidence of lung cancer, achieving an AUC of 74.9% (48). Ardila et al. reached an AUC of 94.4% using a 3D CNN model for lung nodule detection (49). Cellina et al. reviewed numerous studies on AI applications in lung cancer imaging and diagnosis, reporting AUC values ranging from 87% to 95% (50, 51). This is slightly lower than the AUC of 96.3% achieved by our model, indicating superior performance in distinguishing between lung cancer and non-cancer cases. Additionally, different models reported classification accuracies typically around 90% to 97%, with sensitivity and specificity values ranging from 75% to 95%. Our model's classification accuracy of 91.7% and sensitivity of 77.6% fall within this range. However, our model's sensitivity is slightly lower, suggesting a need for improvement in detecting true positive cases. In contrast, our model's specificity of 96.2% is significantly higher than other studies, indicating its superior ability to correctly identify individuals without lung cancer, minimizing false positives. Comparing positive predictive value (PPV), Hsu et al. achieved 15.0%, whereas our study reported a much higher PPV of 86.8%, reflecting greater accuracy in positive predictions and ensuring fewer false positives. Both studies demonstrated high negative predictive

values (NPV), with Hsu et al. at 99.0% and our study at 93.1%, highlighting effectiveness in correctly predicting negative cases (51). Overall, our model's robust performance metrics indicate its potential to significantly enhance lung cancer screening and diagnostic accuracy in clinical settings.

The data collection period from January to February 2023 at Pingyi Traditional Chinese Medicine Hospital was relatively short, which may limit the diversity of our dataset and affect the generalizability of our model. To address this issue, we employed extensive data augmentation techniques, such as random resizing, cropping, horizontal flipping, color jitter, random rotation, and normalization. Despite these efforts, it remains essential to increase the sample size to enhance the model's generalizability. We have planned additional data collection efforts and propose implementing a continuous learning framework to periodically retrain the model with updated data from our hospital and datasets from various regions and medical institutions, ranging from community clinics to large tertiary hospitals. This approach aims to enhance the model's adaptability to variations in clinical practice and its generalizability.

In addition, using data from the same institution collected at different times for external validation may not be ideal. However, existing literature supports the effectiveness of temporal external validation (52, 53), which led us to choose this approach for our study. In addition, the external validation confirmed an AUC of 0.7564 across various patient populations and imaging conditions, which is significantly lower than the training model's AUC of 0.963. The two cohorts, although originating from the same institution, represent different time periods. This temporal difference could potentially affect the model's performance, influenced by changes in patient demographics, CT scanning protocols, or other clinical practices over time. Despite implementing several strategies during model training, such as dropout and L2 regularization, the generalizability of our model across different geographical regions and medical institutions remains a significant challenge. To address this, we outline planned future studies intended to apply our model to datasets collected from different regions, including both urban and rural settings, and from various types of medical institutions

ranging from community clinics to large tertiary hospitals. This will allow us to assess its performance and adaptability in diverse healthcare environments.

We observed an increase in validation loss after the 7th epoch, as shown in Figure 2. To enhance the model's generalization and prevent overfitting, we have implemented various measures, including data augmentation, dynamic adjustment of learning rates, integration of dropout layers and L2 regularization and so on. However, the behavior of machine learning models often remains influenced by inherent characteristics of the training data, such as latent noise and complex nonlinear relationships. These factors may cause the model to quickly adapt to these characteristics in the early stages of training, which could pose challenges as training progresses. Despite extensive efforts to prevent overfitting, the model may still exhibit heightened sensitivity to certain specific features of the training data, especially after prolonged training periods. Looking forward, we plan to further enhance our model's performance by expanding our dataset and exploring the use of updated and more sophisticated model architectures.

In practical clinical applications, deep learning models, including ours, have shown remarkable potential for improving the accuracy and speed of medical diagnoses, as supported by research from Erickson et al. and others (54). Our CNN-based model demonstrates high accuracy in the early detection and diagnosis of lung cancer, offering the potential for continuous improvement through updates and retraining with new data. In addition, challenges such as data privacy, model interpretability, and acceptance within the medical community remain. Therefore, future work will also need to focus on strengthening data privacy measures and enhancing the interpretability of the model. Addressing these challenges comprehensively will be key to fully integrating advanced diagnostic tools like ours into clinical practice, ultimately improving early lung cancer diagnosis and patient outcomes.

Data availability statement

The datasets presented in this article are not readily available because the data is individual-level data restricted for sharing due to ethical requirements. Requests to access the datasets should be directed to apocalypse1523@163.com.

Ethics statement

The studies involving humans were approved by Ethics review committee of Pingyi county Hospital of Traditional Chinese Medicine (PYX2YYYYLLWYh2023030602). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent

for participation from the participants or the participants' legal guardians/next of kin because Data for this study were collected and anonymized by Pingyi County Hospital of Traditional Chinese Medicine. It was retrospectively acquired from all patients who underwent imaging examinations during a specified period. Due to the retrospective nature of the study and the anonymization of the data, obtaining informed consent from specific subjects was not feasible.

Author contributions

HX: Funding acquisition, Validation, Writing – original draft, Writing – review & editing. YY: Methodology, Validation, Writing – original draft, Writing – review & editing. OL: Methodology, Software, Writing – original draft, Writing – review & editing. JC: Data curation, Validation, Writing – review & editing. XH: Data curation, Writing – review & editing. ZT: Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Research on optimizing the informatization and standardization of hospital infection monitoring systems in medical institutions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2024.1424546/full#supplementary-material>.

References

- Oliver AL. Lung cancer: epidemiology and screening. *Surg Clin North Am.* (2022) 102:335–44. doi: 10.1016/j.suc.2021.12.001
- Weber A, Morgan E, Vignat J, Laversanne M, Pizzato M, Rungay H, et al. Lung cancer mortality in the wake of the changing smoking epidemic: A descriptive study of the global burden in 2020 and 2040. *BMJ Open.* (2023) 13:e065303. doi: 10.1136/bmjopen-2022-065303
- Lu S, Cheng Y, Huang D, Sun Y, Wu L, Zhou C, et al. Efficacy and safety of seliperatinib in chinese patients with advanced RET fusion-positive non-small-cell lung cancer: A phase II clinical trial (LIBRETTO-321). *Ther Adv Med Oncol.* (2022) 14:17588359221105020. doi: 10.1177/17588359221105020
- 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 J Clin.* (2024) 74:229–63. doi: 10.3322/caac.21834
- Yuan Z, Yu X, Wu S, Wu X, Wang Q, Cheng W, et al. Instability mechanism of osimertinib in plasma and a solving strategy in the pharmacokinetics study. *Front Pharmacol.* (2022) 13:928983. doi: 10.3389/fphar.2022.928983
- Huang J, Deng Y, Tin MS, Lok V, Ngai CH, Zhang L, et al. Distribution, risk factors, and temporal trends for lung cancer incidence and mortality: A global analysis. *Chest.* (2022) 161:1101–11. doi: 10.1016/j.chest.2021.12.655
- Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. *Nat Rev Clin Oncol.* (2023) 20:624–39. doi: 10.1038/s41571-023-00798-3
- Wu FZ, Huang YL, Wu CC, Tang EK, Chen CS, Mar GY, et al. Assessment of selection criteria for low-dose lung screening CT among asian ethnic groups in Taiwan: from mass screening to specific risk-based screening for non-smoker lung cancer. *Clin Lung Cancer.* (2016) 17:e45–56. doi: 10.1016/j.clcc.2016.03.004
- Li C, Wang H, Jiang Y, Fu W, Liu X, Zhong R, et al. Advances in lung cancer screening and early detection. *Cancer Biol Med.* (2022) 19:591–608. doi: 10.20892/jissn.2095-3941.2021.0690
- Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: current application and future perspective. *Semin Cancer Biol.* (2023) 89:30–7. doi: 10.1016/j.semcancer.2023.01.006
- Thong LT, Chou HS, Chew HSJ, Lau Y. Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: A systematic review and meta-analysis. *Lung Cancer.* (2023) 176:4–13. doi: 10.1016/j.lungcan.2022.12.002
- Feng J, Jiang JJ. Deep learning-based chest CT image features in diagnosis of lung cancer. *Comput Math Methods Med.* (2022) 2022:4153211. doi: 10.1155/2022/4153211
- Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med.* (2006) 355:1763–71. doi: 10.1056/NEJMoa060476
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the cheXNeXt algorithm to practicing radiologists. *PLoS Med.* (2018) 15:e1002686. doi: 10.1371/journal.pmed.1002686
- Peduzzi P, Concato J, Kemper E, Kemper E, Holford TR, Feinstein AR, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* (1996) 49:1373–9. doi: 10.1016/S0895-4356(96)00236-3
- Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci.* (2010) 25:1–21. doi: 10.1214/09-STS313
- LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* (1998) 86:2278–324. doi: 10.1109/5.726791
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv.* (2014) 1409.1556. arXiv:1409.1556.
- Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev BioMed Eng.* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442
- Guan Q, Wang Y, Ping B, Li D, Du J, Qin Y, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. *J Cancer.* (2019) 10:4876–82. doi: 10.7150/jca.28769
- Rohini A, Praveen C, Mathivanan SK, Muthukumar V, Mallik S, Alqahtani MS, et al. Multimodal hybrid convolutional neural network based brain tumor grade classification. *BMC Bioinf.* (2023) 24:382. doi: 10.1186/s12859-023-05518-3
- pydicom. pydicom documentation. Available online at: <https://pydicom.github.io/> (Accessed 2024-07-16).
- matplotlib. matplotlib documentation. Available online at: <https://matplotlib.org/stable/contents.html> (Accessed 2024-07-16).
- Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proceedings of the 14th International Joint Conference on Artificial Intelligence.* (1995) Volume 2 (IJCAI'95). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1137–1143.
- Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst.* (2014) 27:3320–8.
- Goodfellow I, Bengio Y, Courville A. *Deep learning [M]*. MIT Press (2016) Cambridge, Massachusetts, USA. <https://www.deeplearningbook.org/>.
- Prechelt L. Early stopping - but when? *Neural Networks: Tricks Trade.* (1998) 1524:55–69. doi: 10.1007/3-540-49430-8_3
- Fawcett T. An introduction to ROC analysis. *Pattern Recognition Lett.* (2006) 27:861–74. doi: 10.1016/j.patrec.2005.10.010
- Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol.* (2011) 2:37–63.
- Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2016. *J Natl Cancer Center.* (2022) 2:1–9. doi: 10.1016/j.jncc.2022.02.002
- Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA: A Cancer J Clin.* (2023) 73:17–48. doi: 10.3322/caac.21763
- Yang D, Woodard G, Zhou C, Wang X, Liu Z, Ye Y, et al. Significance of different response evaluation criteria in predicting progression-free survival of lung cancer with certain imaging characteristics. *Thorac Cancer.* (2016) 7:535–42. doi: 10.1111/1759-7714.12363
- Wei N, Zheng B, Que W, Zhang J, Liu M. The association between proton pump inhibitor use and systemic anti-tumour therapy on survival outcomes in patients with advanced non-small cell lung cancer: A systematic review and meta-analysis. *Br J Clin Pharmacol.* (2022) 88:3052–63. doi: 10.1111/bcp.15276
- Saha P, Nyarko RO, Lokare P, Lokare P, Boateng A, Kahwa I, et al. Effect of covid-19 in management of lung cancer disease: A review. *Asian J Pharm Res Dev.* (2022) 10(3):58–64. doi: 10.22270/ajprd.v10i3.1131
- Kerpel-Fronius A, Tammemägi M, Cavic M, Henschke C, Jiang L, Kazerooni E, et al. Screening for lung cancer in individuals who never smoked: an international association for the study of lung cancer early detection and screening committee report. *J Thorac Oncol.* (2022) 17:56–66. doi: 10.1016/j.jtho.2021.07.031
- Herath S, Sadeghi Rad H, Radfar P, Ladwa R, Warkiani M, O'Byrne K, et al. The role of circulating biomarkers in lung cancer. *Front Oncol.* (2022) 11:801269. doi: 10.3389/fonc.2021.801269
- Aberle DR, Adams AM, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
- Grenier PA, Brun AL, Mellot FJ. The potential role of artificial intelligence in lung cancer screening using low-dose computed tomography. *Diagnostics (Basel).* (2022) 12:2435. doi: 10.3390/diagnostics12102435
- Reck M, Dettmer S, Kauczor H-U, Kaaks R, Reinmuth N, Vogel-Claussen J, et al. Lung cancer screening with low-dose computed tomography: current status in Germany. *Dtsch Arztebl Int.* (2023) 120(23):387. doi: 10.3238/arztebl.m2023.0099
- Nekolla EA, Brix G, Griebel J. Lung cancer screening with low-dose CT: radiation risk and benefit-risk assessment for different screening scenarios. *Diagnostics (Basel).* (2022) 12:364. doi: 10.3390/diagnostics12020364
- Donovan K, Tustison NJ, Linn KA, Shinohara RT. Multivariate residualization in medical imaging analysis. *bioRxiv: Preprint Server Biol.* (2023). doi: 10.1101/2023.02.15.528657
- Reznikov N, Buss DJ, Provencher B, McKee MD, Piché N. Deep learning for 3D imaging and image analysis in biomineralization research. *J Struct Biol.* (2020) 212:107598. doi: 10.1016/j.jsb.2020.107598
- Hyun Y, Kim D. Development of deep-learning-based single-molecule localization image analysis. *Int J Mol Sci.* (2022) 23:6896. doi: 10.3390/ijms23136896
- Zhang H, Ma R, Zhao Y, Zhang Q, Sun Q, Ma Y, et al. Optimized convolutional neural network recognition for athletes' Pneumonia image based on attention mechanism. *Entropy (Basel Switzerland).* (2022) 24:1434. doi: 10.3390/e24101434
- Efimenko M, Ignatev A, Koshechkin K. Review of medical image recognition technologies to detect melanomas using neural networks. *BMC Bioinf.* (2020) 21:270. doi: 10.1186/s12859-020-03615-1
- Miotto R, Wang F, Wang S, Jiang X, Dudley JT, et al. Deep learning for healthcare: review, opportunities and challenges. *Briefings Bioinf.* (2017) 19:1236–46. doi: 10.1093/bib/bbx044
- Lu MT, Raghu VK, Mayrhofer T, Aerts HJWL, Hoffmann U. Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening computed tomography: development and validation of a prediction model. *Ann Intern Med.* (2020) 173:704–13. doi: 10.7326/M20-1868
- Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med.* (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x
- Cellina M, Cè M, Irmici G, Ascenti V, Khenkina N, TotoBrocchi M, et al. Artificial intelligence in lung cancer imaging: unfolding the future. *Diagnostics (Basel).* (2022) 12:2644. doi: 10.3390/diagnostics12112644

51. Hsu YC, Tsai YH, Weng HH, Hsu LS, Tsai YH, Lin YC, et al. Artificial neural networks improve LDCT lung cancer screening: a comparative validation study. *BMC Cancer*. (2020) 20:1023. doi: 10.1186/s12885-020-07465-1
52. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. (2009) 338:b605. doi: 10.1136/bmj.b605
53. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol*. (2022) 22:316. doi: 10.1186/s12874-022-01801-8
54. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. (2017) 37:505–15. doi: 10.1148/rg.2017160130