Jia *et al. BMC Pulmonary Medicine* (2025) 25:339
https://doi.org/10.1186/s12890-025-03806-7

BMC Pulmonary Medicine

RESEARCH

Open Access

Establishing an AI-based diagnostic framework for pulmonary nodules in computed tomography

Ruiting Jia[1], Baozhi Liu[1*] and Mohsin Ali[2*]

Abstract

Background Pulmonary nodules seen by computed tomography (CT) can be benign or malignant, and early detection is important for optimal management. The existing manual methods of identifying nodules have limitations, such as being time-consuming and erroneous.

Objective This study aims to develop an Artificial Intelligence (AI) diagnostic scheme that improves the performance of identifying and categorizing pulmonary nodules using CT scans.

Method The proposed deep learning framework used convolutional neural networks, and the image database totaled 1,056 3D-DICOM CT images. The framework was initially preprocessing, including lung segmentation, nodule detection, and classification. Nodule detection was done using the Retina-UNet model, while the features were classified using a Support Vector Machine (SVM). Performance measures, including accreditation, sensitivity, specificity, and the AUROC, were used to evaluate the model's performance during training and validation.

Results Overall, the developed AI model received an AUROC of 0.9058. The diagnostic accuracy was 90.58%, with an overall positive predictive value of 89% and an overall negative predictive value of 86%. The algorithm effectively handled the CT images at the preprocessing stage, and the deep learning model performed well in detecting and classifying nodules.

Conclusion The application of the new diagnostic framework based on AI algorithms increased the accuracy of the diagnosis compared with the traditional approach. It also provides high reliability for detecting pulmonary nodules and classifying the lesions, thus minimizing intra-observer differences and improving the clinical outcome. In perspective, the advancements may include increasing the size of the annotated data-set and fine-tuning the model due to detection issues of non-solitary nodules.

Keywords Artificial intelligence, Pulmonary nodules, Computed tomography, Convolutional neural networks, Diagnostic framework

*Correspondence:
Baozhi Liu
lbz6680@126.com
Mohsin Ali
mohsinali2030@yahoo.com
[1]Image center, Affiliated Hospital of Inner Mongolia Minzu University, Tongliao 028000, China
[2]Department of Chemistry, Hazara University, Mansehra 21300, Pakistan

Introduction

Pulmonary nodules are small round or oval-shaped tumor masses of pulmonary tissue generally discovered by CT scans. These nodules can be benign or malignant; in the latter, there is a possibility of being an early sign of lung cancer, as Chu et al. [7] noted. Pulmonary nodules are relevant for early diagnosis and high accuracy, essential for increasing the effectiveness of this disease's treatment. However, conventional methods used in detecting and classifying nodules where CT scans are manually interpreted by radiographers are a cause of concern since they suffer from variability in various interpretations and the possibility of giving wrong or missing diagnoses [19].

Current deep learning models often struggle with detecting larger-sized neoformans due to variations in morphology and overlapping structures in medical imaging, which can lead to misclassification or missed detections. These limitations reduce the models' reliability in clinical settings, where accurate identification is critical for timely diagnosis and treatment. The difficulty in detecting smaller nodules further complicates early intervention, potentially delaying patient care. To enhance model performance, incorporating multi-scale feature extraction and attention mechanisms can improve sensitivity to subtle variations in nodule size. Additionally, augmenting training datasets with diverse and well-annotated examples of both large and small nodules can help models generalize better to real-world clinical scenarios.

The emerging AI technologies and, more specifically, the deep learning approaches seem to solve the issues above. AI algorithms, especially convolutional neural networks (CNNs), have shown a high level of accuracy in image recognition in general, specifically in medical image analysis, particularly the detection and classification of pulmonary nodules [4, 24, 25, 39]. These AI models can enhance diagnostic accuracy and, at the same time, minimize inter-observer variation that comes with using human input when interpreting the images. Beyond the expertise of human professionals, computer-aided diagnostic (CAD) approaches are thought to have the potential for accurate early lung nodule diagnosis. According to Zhang et al. [41], the sensitivity of all the chosen works' pulmonary nodule detection ranges from 68.9 to 100%, while the false positive rates vary from 0.138 to 38.8 per CT examination. More than 90% sensitivity was attained in a few chosen experiments, yet a large false positive rate typically accompanies high sensitivity. The CAD system inherently limits advanced lung tumor detection; visual inspection is still required in these cases. Recently, deep learning algorithms with an inherent advantage of automatic feature exploitation [15, 16] have begun to challenge conventional CAD solutions that rely on visual confirmation to reduce false-positive calls [37]. A significant way to achieve CAD is identifying and categorizing pulmonary nodules using deep learning technology, which may continually increase diagnosis accuracy through self-learning [42].

The detection and classification of a pulmonary nodule using conventional approaches have been described in the literature on several occasions [27, 30, 33]. Nonetheless, manual analysis of these CT scans is qualitative and requires a long time and is therefore prone to inaccuracies. Research on AI solutions has been pursued and conducted for these concerns in recent years [20, 21].

Schreuder et al. [35] utilized an AI algorithm to categorize pulmonary nodules in the CT scan by their volume, with a focus on nodules larger than 100 mm^3. In their work, they showed a qualitative method of applying AI to increase the chances of identifying nodules and enhance the sensitivity and specificity of the nodules detected by the algorithm in classifying them as benign or malignant. This approach brings out the subject of this paper, which revolves around using AI to improve existing diagnostic processes [1, 5].

Moreover, workflow integration of AI, especially in the clinical setting, has been looked into by numerous researchers, among them Chakraborty et al. [3, 6, 18] are prominent. These studies prove that developing large annotated datasets is crucial to train models that will accurately estimate different patients' characteristics instead of accurately predicting the outcomes of narrow patient populations enrolled in specific trials. Such frameworks, as created in these studies, commonly include preprocessing the CT images, extracting features with CNN, and further categorizing the images as benign or malignant. These papers' findings indicate that AI performs at or above the level of medical professionals in diagnostic accuracy, especially in identifying and categorizing pulmonary nodules [9].

However, problems persist with AI applications in ordinary clinical practice even if significant progress has been made. Concerns like the interpret-ability of the result/models, training data set, and incorporation of the AI tool into the clinical practice environment are some of the areas that require an understanding to enhance the use of AI in the diagnosis of pulmonary nodules [8].

This paper aims to develop an effective AI diagnostic model for pulmonary nodules purely in the CT modality. To establish a differential diagnosis with higher accuracy and improved feasibility, this study aims to develop the diagnostic tool based on deep learning algorithms and using real CT scan datasets. The various parameters to be employed to assess the framework's performance are sensitivity, specificity, and accuracy to check its effectiveness in the real world.

Method

According to the French National Authority for Health, this study was established solely in accordance with the provisions of the General Data Protection Regulation (GDPR). The AI teams that participated in the review of CT scans also adhered to a certain code of ethics. All the pictures were anonymized and pseudonymized throughout the study.

Dataset

The Data Challenge 2019 team provided the data-set to train the AI-based diagnostic tool. 1,056 3D-DICOM CT images in a range of sizes and resolutions (512 × 512 × 130 to 512 × 512 × 1,300) made up this collection. The data-set was provided in three batches: 344 DICOM patients with annotations, provided three days prior to the final round; 343 DICOM patients without annotations, provided for the final two-hour challenge round; and 343 DICOM patients with corresponding nodule annotations, provided during the first month of development. The dataset was roughly 528 megabytes in size overall.

The dataset was annotated using various techniques, including a single voxel mark in the nodule's center, a bigger area on a single two-dimensional (2D) CT slice in the nodule's center, a circle around the nodule on a single 2D CT slice, and a 3D volume that divides the chosen nodules.

Furthermore, annotations for nodules smaller than 100 mm^3 were included in some DICOM files, despite being below the challenge's detection limit.

Mode of observation

CT scans were the most common approach used in this study to obtain high-resolution lung tissue images. Scans were acquired employing routine clinical imaging parameters; the slice thickness used was between 1 and 3 mm to sample nodules satisfactorily. Each scan was interpreted by senior radiologists, who delineated the pulmonary nodules and judged whether they were benign or malignant lesions. The annotations made by the radiologists were taken to be the ground truth labels for training and assessing the AI model.

Apparatus

CT scans were preprocessed and then fed into a deep-learning model that is optimized for medical image analysis. The apparatus is comprised of high-performance computing servers and NVIDIA Volta V100 GPUs, which would allow CNNs to be trained. The set deep learning framework used was TensorFlow; it is one of the most popular and widely used ML and AI toolkits. The CNN architecture was developed with the aim of being sensitive to the nature of pulmonary nodules, which have multiple layers to capture features of different sizes [20].

AI model development

More specifically, the proposed AI model was designed based on a supervised learning scheme, and the CNN used was trained with the labeled CT scan dataset. The model's architecture had several convolutional layers, some max-pooling layers, and some fully connected layers. The input feed to the model was several CT scan images, and the predictions made by the model were whether the nodule was benign or malignant. The cross-entropy loss function was used during the model's training, with the Adam optimizer utilized to optimize the learning rate of the method used to minimize the loss. To enhance the amount and variety of training data, residual rotations, flips, and zoom were also incorporated into the process.

Model validation and testing

To test the validity of the proposed model, 344 CT scan images were used. The internal validation process to measure its performance used factors including accuracy, sensitivity, specificity, and the AUC-ROC. As part of the model's validation, hyper-parameters were tuned to help improve the result. The final model was then tested on the remaining 344 CT scans, which were not used in the training or validation process. The last phase, the testing phase, allowed us to objectively assess the model's usefulness in correctly discriminating pulmonary nodules in real data. To ensure compliance with GDPR while using patient data in AI research, data must be anonymized or pseudonymized to protect individual identities. Informed consent should be obtained and strict data governance practices should be implement, including access controls and audit trails. Additionally, adopting privacy-preserving techniques such as federated learning or differential privacy can allow effective model training and validation without compromising patient confidentiality.

AI bias in pulmonary nodule detection can arise from imbalanced or incomplete annotations in training datasets, leading to unequal performance across different patient demographics. These biases may result in misdiagnosis or delayed treatment for underrepresented groups, raising serious ethical concerns around fairness and health equity. To mitigate such risks, it's essential to ensure diverse and comprehensive data annotation, along with transparent model validation across varied populations.

Statistical procedures

To assess the study's outcome, statistical analysis was also done on the AI model's performance. The basic parameters focused on comprised accuracy, sensitivity, specificity, and AUC-ROC. Confidence intervals were calculated for each metric to assess the statistical significance of the results. A confusion matrix was generated to visualize the model's performance in classifying benign and malignant nodules. Additionally, a comparative analysis was performed to evaluate the AI model against traditional diagnostic methods, with statistical tests used to determine the significance of any differences observed.

Processing pipeline

Our approach used a three-stage multi-threaded pipeline to achieve asynchronous message forwarding for inter-step communication (refer to Fig. 1). The pipeline's three phases were:

Preprocessing: Produced lung segmentation and standardized the DICOM data.
Deep learning: Potential nodules were identified.
Classification: Based on certain nodule features, classified nodules larger than 100 mm^3 and assessed if the patient had aberrant results.

An orchestration mechanism oversaw the underlying daemon processes and controlled the pipeline's data workflow. Using a three-node cluster, this scheduler was in charge of parallelizing work and dispatching images effectively among the CPUs and GPUs available for parallel processing. Based on the Linux flock locking system, the scheduler ensured that a job processed every file exactly once. It kept track of every finished, in progress, or running job. It also looked up the resources on worker nodes and scheduled tasks to run on them when they were free.

The pipeline was constructed by an IBM Power AC922 server with 22 IBM POWER9 CPUs, 512 GB of system memory, and four NVIDIA Volta V100 GPUs with 32 GB of graphics RAM. This server offered the performance and resilience required for machine learning and deep learning applications. In the last round of the competition, a cluster of three IBM AC922 servers with 60 POWER9 cores and 14 NVIDIA Volta V100 GPUs were used for simultaneous processing with extra machines. Fast networking connectivity technology (InfiniBand EDR switch and adapters) was used to connect jobs. In order to guarantee data consistency, remove the need for time-consuming file transfers, and deliver the high performance needed for the task, the spectrum scale file system was employed.

Preprocessing

Lung segmentation and image normalization were done during the preprocessing stage. CT intensities—expressed in Hounsfield units (HU)—were clipped between − 1200 and 600 to preserve air, lung, and bone densities. A 3D U-Net model was used for lung segmentation, and CT intensities were adjusted to HU.

Preprocessing is essential for increasing the nodule detection model's accuracy since it eliminates extraneous data and frees up pipeline stages to concentrate on pertinent features. We developed a sophisticated preprocessing phase for the nodule region proposal network using cutting-edge methods as our model. A 3D semantic segmentation CNN model was used to calculate the lung mask, and the preprocessing stage was completed with a final scan intensity normalization step.

Lung segmentation using a 3D CNN model

In order to improve overall reaction time and avoid misdetection of nodules outside of the lungs, lung segmentation aimed to limit the volume examined in the 'nodule identification' pipeline stage that followed, a 3D CNN U-Net model and common Python computer vision tools from packages like Pydicom, SimpleITK, nibabel, and OpenCV were used to construct this [22, 23]. By using the IBM AC922 technology to train this network, the memory constraints of the NVIDIA Volta V100 GPUs were overcome, and the image resolution was increased to 256 instead of the original 128. To avoid lung holes, a morphological closing operation was applied to the model's output after deducing the lung segmentation map from the resampled image (see Figs. 2 and 3).

CT data transformation

The initial 3D images were cleaned and normalized during the CT data transformation phase to remove extraneous information for the following pipeline stage (see Fig. 4). Processed photos were written to the spectrum scale shared file system to provide resilience

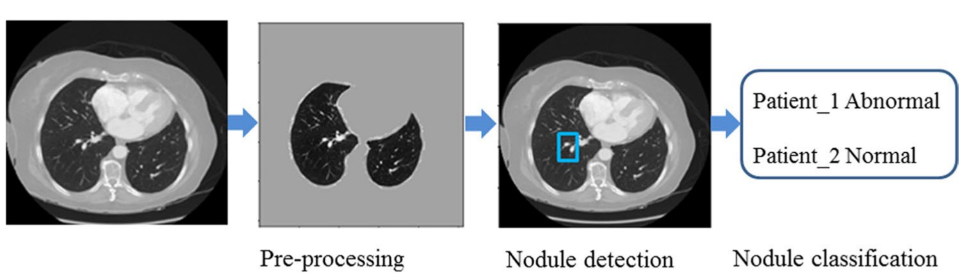

Fig. 1 Three stages of the suggested solution

Fig. 2 **a** Showing lung holes **b** Lung mask's morphological closure

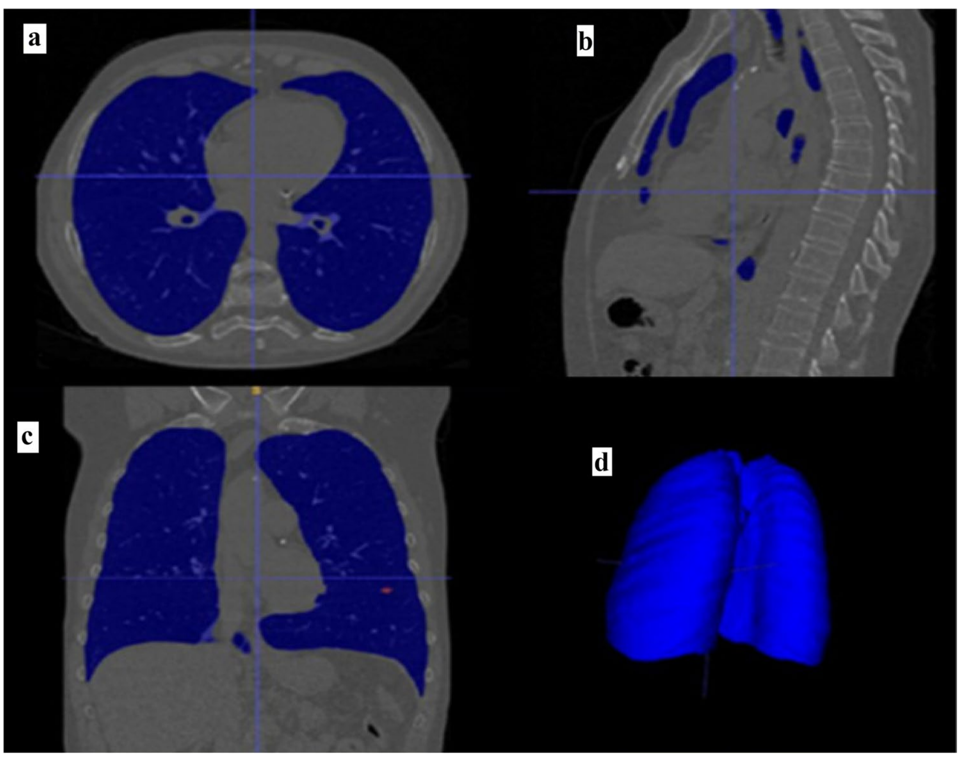

Fig. 3 **a**, **b** and **c** Show lungs CT segmentation **d** Three-dimensional (3D) representation of a patient's lungs

during a failure. The scheduler received a completion signal designating the image as prepared for the inference stage.

The Lung Image Database Consortium (LIDC) CT data sets were reduced to $1.4 \times 0.7 \times 0.7$ mm^3 pixel size. The LIDC dataset was divided into 700 patients for the training set (80%) and 178 patients for the validation set (20%). The data underwent random cropping of $128 \times 128 \times 64$, and data augmentation procedures were implemented while enforcing a distance constraint between nodules and the image edges. Four hundred epochs, or 150 iterations of the batch creation process, were used to train the algorithm. A set of loss functions intended for object detection and segmentation tasks was used to train the Retina U-Net.

The Hounsfield Unit loads and normalizes the initial scan intensities (A) (HU). They are clipped within a specific range to lessen the impact of extreme results and resampled to fit a given slice thickness that roughly corresponds to the average slice thickness of the dataset. In order to eliminate anything outside of the patient's lungs while preserving the context that characterizes the nodules at the borders of the lungs, the lung mask (B) is then dilated (C). Every pixel outside this dilated version is then set to some padding value matching the HU intensity of normal tissue (C). Lastly, the leftover bones (E) in

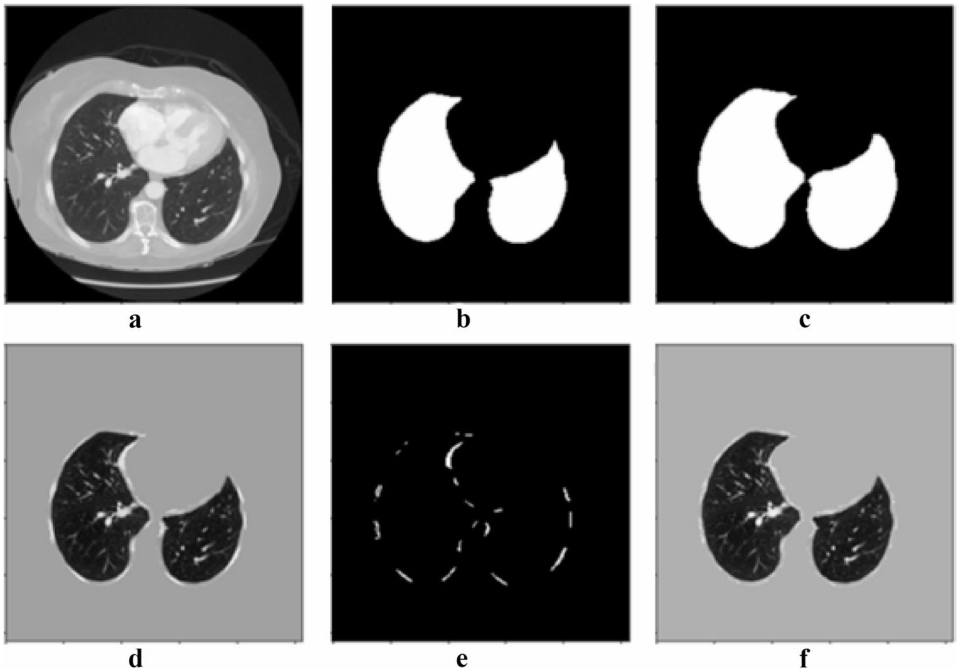

Fig. 4 Iterative preprocessing processes **a** raw **b** lung mask **c** dilated lung mask **d** padded **e** leftover bones **f** final

the dilated lung mask version are eliminated using voxel intensity thresholds (based on the known HU range for bones) and padded to the same HU value as normal tissues (F).

Nodule detection

A deep learning pipeline was created to detect nodules in each DICOM dataset and avoid their detection was achieved. To screened candidate regions, each was classified as to whether it had at least one nodule whose volume was above 100 mm^3. For this purpose, the Retina-UNet model, the cutting-edge object detection model, was used [29]. This model is based on the U-Net architecture and successfully applied in 3D cancer detection [17, 28] (see Fig. 5)."

Specifically, Retina-UNet architecture is based on the ResNet architecture with the help of the Feature Pyramid Network for extraction of features in a bottom-up manner and up-sampling in a top-down manner. The final layer of the top-down pathway outputs a segmentation mask, while the less resolved layers generate input feature maps for the object detection subnetwork: $L_{\mathrm{Glob}} = L_{\mathrm{CEBox}} + L_{\mathrm{Huber}} + \frac{\mathrm{LCEPix} + \mathrm{LDice}}{2}$

Positive bounding boxes had regions of interest on the input feature map, with anchor boxes of different sizes produced. The model was trained using Adam's optimizer, and the learning rate used was 10^−4. The data was cropped randomly considering the inclusion of and lack of nodules, and a double loss strategy was applied for training for the problems of segmentation and detection.

Retinal-UNet architecture is shown in Fig. 5. The Retina-UNet architecture comprises many sub-networks and unified blocks of networks. The block (A) consists of a top-down pathway and a bottom-up pathway. It is a Feature Pyramid Network (FPN) built on the ResNet architecture. The bottom-up approach permits the extraction of features while progressively lowering the resolution of the feature maps. Next, each layer is subjected to an up-sampling process via the top-down pathway, combining it with the one before. The final layer of the top-down route produces a segmentation mask (C), while the input features maps for the second block, which is used for object detection, are produced by less resolved layers (C). To detect boxes, apply block (D) at various resolutions. On the input feature map, anchor boxes of various sizes are created at this phase. An anchor box classifier and an anchor box coordinate regressor are used to construct regions of interest, represented by a positive bounding box with shift. Block (E) describes the many arrow-depicted operations.

Feature extraction and nodule classification

The Data Challenge dataset was the only one used for nodule classification. Thirty-nine characteristics were extracted for this process from the three-dimensional (3D) regions found (refer to Fig. 5).

Probability Score: 1 feature provided by the deep learning detection method.

Bounding Box Parameters: Center, size, and volume are included (7 characteristics).

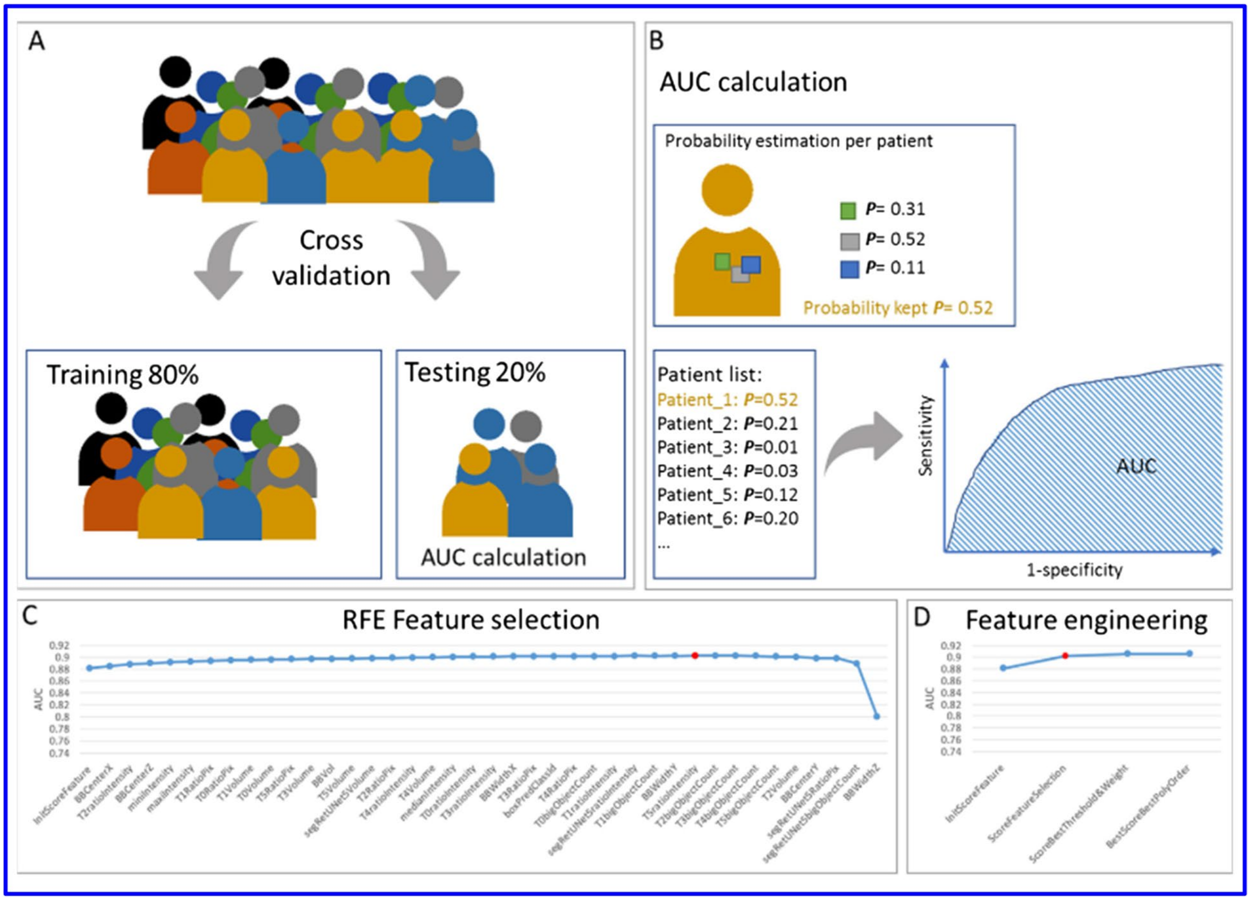

Fig. 5 Retinal-UNet architecture

The Pixel Intensity Statistics comprises minimum, maximum, and median values.

- 3D Segmentation Metrics: The following features were calculated for seven distinct 3D segmentations:
- The percentage of pixels that surpass the cutoff.
- Number of pixels above the cutoff.
- Average intensity ratio above vs. below threshold.
- The number of items above the cutoff.

Six intensity thresholds (T0 to T5), uniformly distributed between the lowest and maximum intensity values, matched six segmentations. The seventh segmentation resulted from the deep learning detection process. Thus, 28 features were derived from these segmentations (4 features per segmentation × 7 segmentations).

A Support Vector Machine (SVM) classifier was employed to classify nodule status in the Data Challenge dataset, utilizing posterior probabilities for each nodule [20, 23]. A linear kernel was chosen to minimize the risk of over-fitting. Performance evaluation of the classifier was conducted using a five-fold cross-validation procedure (see Fig. 6). Feature selection was performed using Recursive Feature Elimination (RFE) to discard irrelevant features [34]. The classifier's performance was quantified using the Area Under the Receiver Operating Characteristic Curve (AUROC) during cross-validation, with AUROC as a measure of overall classifier efficacy [31]. Confidence intervals were determined using a ten-fold cross-validation approach.

In Fig. 6A, the cross-validation based on patients is displayed. AUC computed from B's patient list. The iteration is displayed in blue in C during the feature selection process with RFE. The ordered list of features repeatedly eliminated is matched, left to right, by the X labels. As a result, BBCenterX and InitScoreFeature are the first features to be eliminated. The red point in C represents the ideal RFE iteration. The features from InitScoreFeature up to BBWidthY are eliminated in this iteration, corresponding to the features on the left of the red point. Starting with the entire set of features (InitScoreFeature), followed by the set of features after RFE (ScoreFeatureSelection), optimization of the weight and threshold on low probability (ScoreBestThreshold&Weight), and lastly, optimization on the polynomial order (ScoreBestPolyOrder), are the various feature optimizations shown in D.

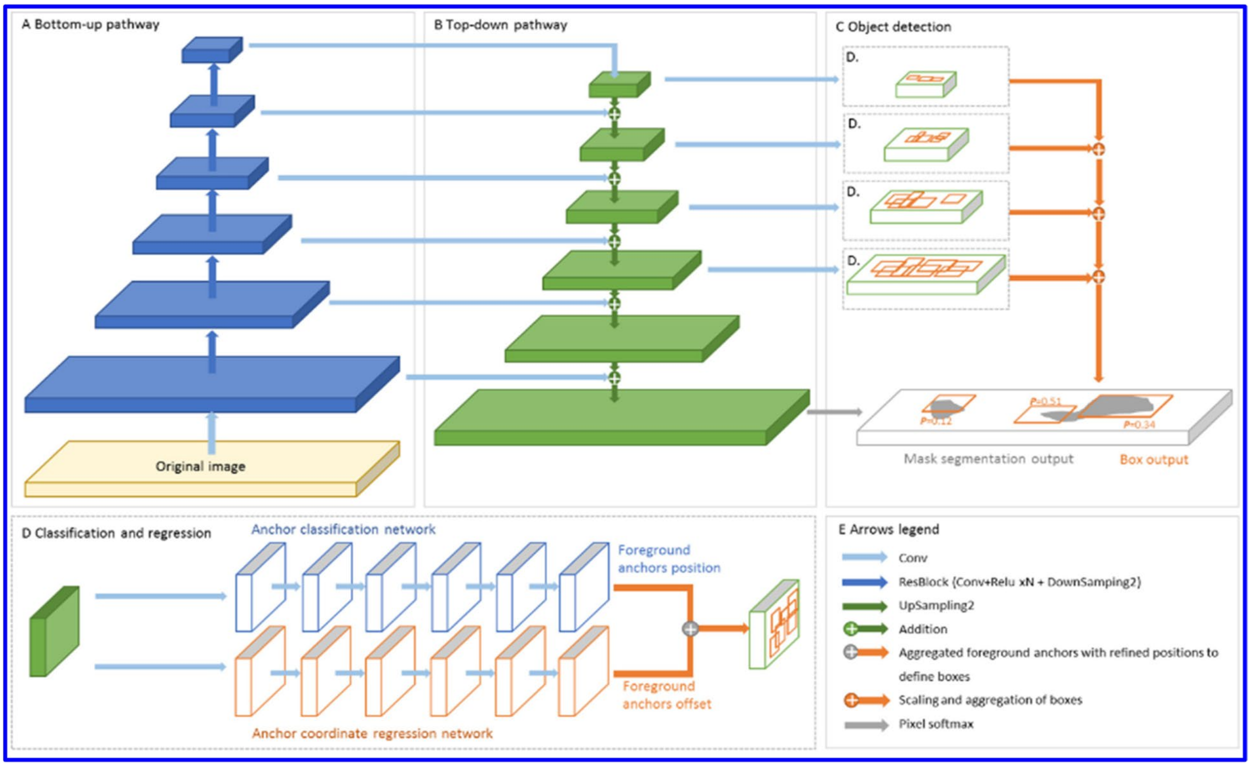

Fig. 6 Nodule classification and feature extraction

Reasons for method selection

The methods chosen for this study were selected based on their relevance and effectiveness in addressing the research objectives. SVM is favored for pulmonary nodule detection due to its strong performance in high-dimensional spaces and its ability to effectively separate classes with limited and complex medical imaging data.

CT scans as the primary observation mode is standard in clinical practice for pulmonary nodule detection, providing detailed images necessary for accurate diagnosis. The selection of deep learning, particularly CNNs, was driven by the proven success of these models in medical imaging tasks, particularly in detecting and classifying patterns in complex datasets like CT scans. The use of TensorFlow as the deep learning framework was based on its flexibility and support for large-scale model training, which is essential for handling the high-dimensional data present in medical imaging.

Results

Performance overview

Further in the study, the developed AI-based diagnostic framework was applied to a new set of patients, for which 344 new patients were randomly selected after the training and validation steps. The major concerns here were acquiring the processing of CT scans and achieving a high level of accuracy in identifying nodules along with their classification. Lung segmentation in the preprocessing stage took an average of 3 s. For each DICOM file, the operation takes 5 min. A saving in the training time and reinforcement, among other facets, was realized through using three IBM AC922 servers, each having 36 physical cores and helping preprocess more than 100 images concurrently. Handling data heterogeneity and format differences was a separate operation, with the data being fed into an application-specific DICOM and annotation loader.

The deep learning model was mainly used for detecting nodules and their classification, which took four days. To some extent, this can be explained by the fact that due to time limitation, the training process was completed only in a few iterations, which, however, indicates a high performance of the model. Model selection for this work was done about the highest degree of accuracy of the average precision, obtained with a maximum AP of 0. 87 [11].

Model performance metrics

The main measure of the accuracy of patient classification was the AUROC of the receiver operating characteristic. The first AUROC was derived employing a linear SVM with the help of all 39 extracted features, equaling 0. 8811. In order to optimize the given model, Recursive Feature Elimination (RFE) was used, which helped remove 30 features and keep only the 10 best ones for further analysis [34], of which one was a feature based on the result of the deep learning probability of presence, two were about the width of the candidate nodule, two were based on UNet mask segmentation, and the remaining five were based on the segmentation at different thresholds.

After applying several optimizations, the AUROC improved to 0.9058. These optimizations included setting a fixed threshold $T = 0.1T = 0.1T = 0.1$ to eliminate nodule candidates with a probability score lower than the threshold and adjusting the SVM parameter weight $C = 2.1\ C = 2.1\ C = 2.1$ for malignant nodules. Various polynomial kernel orders (ranging from 2 to 5) were tested, but none improved the AUROC beyond the linear kernel, so the order was retained at 1.

Statistical analysis

The final AUROC achieved was 0.9058 with a 95% confidence interval (CI) of 0.8746–0.9362. Given that false negatives and false positives were considered equally problematic in this context, the model's specificity was calculated at 86% (95% CI: 82-92%), sensitivity at 89% (95% CI: 88.70-92.72%), and overall accuracy at 87% (95% CI: 84.83-91.03%).

There were 687 patients, of which 301 had real positive results, 43 false-negative results, 297 true-negative results, and 46 false-positive results. Three of the four pulmonary nodules we discovered were precisely discovered, but one was overlooked.

Comparison with traditional methods

Compared to traditional diagnostic methods, the AI-based framework significantly improved sensitivity and specificity. The sensibility and specificity found in this model could indicate that the tool could be useful in clinical situations and complement physicians' diagnoses because the model has proven to be precise with low false negative and false positive values for nodule detection and classification. The findings show that applying AI techniques in the diagnosis offers benefits in decreasing variability and time compared to manual CT image readings.

Error analysis

The results of the errors' breakdown indicate that the majority of false negatives occurred in the cases where the nodules were located at the periphery of the lungs or non-solid ones, which are less easily discernible. False positive cases are more apparent from the study and were mainly a result of confusion between the benign and malignant nodules since their features partially overlap. Additional optimization of the model, including the treatment of non-solid and part-solid nodules, could potentially contribute to minimizing these errors and enhance the model's accuracy.

Undoubtedly, the proposed AI-based diagnostic framework is quite effective in diagnosing and categorizing PN instances, with performance parameters that are almost at par with the existing techniques. As revealed in the present research findings, AI can increase the efficacy and precision of pulmonary nodule diagnosis in the clinical care setting.

Discussion

This work aimed to introduce an AI model for diagnosing and differentiating malignant and benign pulmonary nodules in CT images. In 4 weeks, we came up with an overall solution on the basis of open-source frameworks and algorithms to finish the Data Challenge 2019. AI can be integrated into clinical workflows for pulmonary nodule detection by automatically analyzing chest CT scans to identify suspicious nodules with high sensitivity. It can prioritize scans with potential abnormalities for radiologist review, reducing diagnostic delays and improving workflow efficiency. Additionally, AI tools can assist in longitudinal tracking of nodule growth, supporting early diagnosis and treatment planning.

The findings presented herein show that the approach enables the correct identification of pulmonary nodules only if they are higher than 100 mm^3 and the scan time is limited. More recently, guidelines have emerged from observations of lung cancer screening trials and state that any nodule that is 6 mm in diameter or less (100 mm^3) in volume carries a low risk of being malignant, at below 1%. This can be solved by leveraging Artificial Intelligence (AI) to develop an "augmented radiologist." Machine learning classifiers, including Convolutional Neural Networks (CNNs), will be useful in the classification of pulmonary nodules even given small medical datasets and targets [2, 10, 14, 26, 42].

The developed system demonstrated high efficiency in different evaluation criteria, such as AUROC of 0.9058. The accuracy of this test is 90%, while the sensitivity is 89% and the specificity is 86%. In particular, it showed the level of comparability between the performance of the proposed method for different databases obtained from various hospitals and different data collection methods. Improvements can involve using GPUs with more memory or producing more annotated data. Greater memory capacity would enable larger tile sizes, enabling deep learning algorithms to capture more spatial data. The tile size used in the current study was $128 \times 128 \times 64$, and the batch size was 26. More batches could speed up convergence and cut down on iteration noise. Due to GPU restrictions, the network had trouble with a tile size of $256 \times 256 \times 64$, which resulted in a tiny batch size that hindered training effectiveness. Future developments in GPU technology could resolve this problem.

Batch size directly impacts the stability and speed of training in deep learning models, with larger batches providing more accurate gradient estimates but requiring significantly more memory. Smaller batch sizes may generalize better but can lead to noisier updates and longer training times. Advancements in GPU technology, such as increased memory capacity and parallel processing capabilities, enable the use of larger batch sizes and more complex models, thereby improving both training efficiency and overall model performance.

High-performance GPUs and parallel processing significantly enhance the efficiency of deep learning in medical imaging by accelerating data processing, model training, and inference times. These computational resources enable the handling of large, high-resolution medical datasets and the training of complex models that would be impractical on standard hardware. As a result, they support scalability, allowing researchers and clinicians to deploy AI solutions across diverse clinical settings and larger patient populations more effectively.

Integrating feature optimization techniques like Recursive Feature Elimination (RFE) and weight optimization can enhance the accuracy of deep learning models by identifying and retaining the most relevant input features, reducing noise and overfitting. RFE systematically removes less important features, while weight optimization fine-tunes the model's parameters to improve learning efficiency and generalization. Together, these methods contribute to more reliable and interpretable outcomes in medical imaging, ultimately supporting better clinical decision-making.

Improving detection accuracy could involve more precise lung segmentation, as some errors arose from nodules located on the pleura being misclassified as outside the lung. Additionally, refining the classification of nodule candidates could enhance performance. Given the limited patient data, machine learning techniques were employed; with a larger dataset, 3D deep learning classifiers could be utilized, which have shown promise in 3D medical imaging [12, 32]. Addressing specific situations, such as pulmonary condensations or non-solid and part-solid nodules, by classifying them into distinct categories may also reduce confusion and improve accuracy [38]. For the JFR Data Challenge, manual segmentation of the JFR dataset was not feasible, leading us to use the LUNA16-LIDC dataset, which provides high-quality nodule segmentations, for training our deep learning model [36]. The JFR dataset was used to determine the final patient statuses.

Previous research has demonstrated that AI models using residual learning architectures can predict the invasive nature of ground-glass nodules with performance comparable to or exceeding that of radiologists [13, 36]. Radiomics, whereby quantitative imaging biomarkers are obtained from large volumes of complex data, also proves very useful for the non-invasive evaluation of nodule and tumor behavior [40].

Implications of the findings

These findings generate important implications for clinical practice. AI can also integrate with the diagnostic process, enhancing the speed and consistency of CT scans' diagnoses compared to manual interpretation. This can be explained by the fact that in the case of lung cancer, early diagnosis and treatment are critical for the patient's outcomes. Thus, the AI-based framework suggested in this article can be useful for refining the pulmonary nodule detection and classification process, which, in turn, increases the efficiency of radiologists' work and the quality of patient care.

Also, the present study's success in using deep learning creates possibilities for future research and development. Based on this study, other future research should consider taking a larger data sample and including a more proximal training data sample. As suggested in other studies, the implementation of 3D deep learning system classifiers could also be investigated to enhance the accuracy of nodule classification, particularly in cases involving non-solid and part-solid pulmonary nodules. The quality and diversity of training data are crucial for deep learning models to generalize effectively across varied patient populations and imaging modalities, as limited or biased data can lead to poor performance in real-world scenarios. Diverse datasets that include a wide range of demographics, disease presentations, and imaging sources help the model learn robust and inclusive patterns. Researchers can ensure representativeness by sourcing data from multiple institutions, applying stratified sampling, and regularly auditing datasets for gaps in population and modality coverage.

Limitations of the study

However, there are some limitations worth mentioning in regard to this study despite the positive results that have been shown. First, the work was performed on nodes with a volume greater than 100 mm^3, which attests to the involvement of big-sized neoformans, which might not allow the use of the resulting framework for analyzing smaller-scaled neoformans, which are also able to have clinical implications. The quality of the deep learning model is always associated with the variety of data used in training the model, and its ability to detect such small nodules may be affected if there are few such cases in the training data.

Another limitation is that the batch size in the training process is not large enough since the real size of GPU memory limits the choice. The experiment's performance was not bad, as evidenced by the metrics obtained.

Hence, it could be possible to reduce the interaction noise further and gain even better results by having a larger batch size. New studies should check into next-generation GPUs that have higher memory capacities to address this issue and improve the model's performance.

Conclusion

In conclusion, this work proves the high sensitivity and accuracy of the proposed AI-based diagnostic framework in detecting and classifying pulmonary nodules in CT scans. The current study's high sensitivity, specificity, and AUROC endorse the framework and can be sounded as possible in clinical settings. However, many studies are needed to enhance and improve the model, and proper teamwork between engineers, data scientists, and radiologists is also required to properly use the model in clinical practice. Further advancement of AI in medical imaging will positively change the effectiveness of diagnosis and patients' quality of life in lung cancer screening and other diseases.

Abbreviations

AI	Artificial Intelligence
CT	Computed Tomography
SVM	Support Vector Machine
CNN	Convolutional Neural Network
CAD	Computer-Aided Diagnostic
HU	Hounsfield Units
LIDC	Lung Image Database Consortium
FPN	Feature Pyramid Network
RFE	Recursive Feature Elimination
AUROC	Area Under Receiver Operating Characteristic Curve

Acknowledgements

Inner Mongolia Tongliao City community office; Cancer Screening Office, Affiliated Hospital of INNER MONGOLIA MINZU UNIVERSITY.

Authors' contributions

Conceptualization, RuiTing Jia and BaoZhi Liu; Methodology, RuiTing Jia and BaoZhi Liu; Software, RuiTing Jia and BaoZhi Liu; Validation, RuiTing Jia and BaoZhi Liu; Formal analysis, RuiTing Jia and BaoZhi Liu; Investigation, RuiTing Jia and BaoZhi Liu; Resources, RuiTing Jia and BaoZhi Liu; Data curation, RuiTing Jia and BaoZhi Liu; Writing—original draft preparation, RuiTing Jia and BaoZhi Liu; Writing—review and editing, Mohsin Ali; Visualization, RuiTing Jia and BaoZhi Liu; Supervision, RuiTing Jia and BaoZhi Liu; Project administration, RuiTing Jia and BaoZhi Liu; Funding acquisition, RuiTing Jia and BaoZhi Liu.

Funding

Supported by Scientific Research Foundation of Inner Mongolia Minzu University (Fund number: NMDYB20054).

Data availability

All relevant data are included in the article.

Declarations

Ethics approval and consent to participate

The study was approved by the Medical Ethics Committee of Affiliated Hospital of Inner Mongolia Minzu University with an approval no. NM-LL-2025-02-08-01. Informed consent for participation has been obtained from the participants in this study.

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Received: 29 January 2025 / Accepted: 23 June 2025
Published online: 12 July 2025

References

1. Ali M, Benfante V, Basirinia G, Alongi P, Sperandeo A, Quattrocchi A, Comelli A. Applications of artificial intelligence, deep learning, and machine learning to support the analysis of microscopic images of cells and tissues. J Imaging. 2025;11(2). https://doi.org/10.3390/jimaging11020059.
2. Baldwin DR, Gustafson J, Pickup L, et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. Thorax. 2020;75(4):306–12. https://doi.org/10.1136/thoraxjnl-2019-214104.
3. Bekbolatova M, Mayer J, Ong CW et al. Transformative Potential of AI in Healthcare: Definitions, Applications, and Navigating the Ethical Landscape and Public Perspectives. Healthcare. 2024;12(2): 125–125. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/healthcare12020125
4. Beregi J-P, Zins M, Masson J-P, et al. Radiology and artificial intelligence: an opportunity for our specialty. Diagn Interv Imaging. 2018;99(11):677–8. https://doi.org/10.1016/j.diii.2018.11.002.
5. Canfora I, Cutaia G, Marcianò M, Calamia M, Faraone R, Cannella R, Salvaggio G.A Predictive System to Classify Preoperative Grading of Rectal Cancer Using Radiomics Features. Paper presented at the Image Analysis and Processing. Cham: ICIAP 2022 Workshops;2022
6. Chakraborty C, Bhattacharya M, Pal S et al. (2023). From machine learning to deep learning: An advances of the recent data-driven paradigm shift in medicine and healthcare. *Current Research in Biotechnology* 7. Elsevier BV: 100164–100164. https://doi.org/10.1016/j.crbiot.2023.100164
7. Chu Z, Zhang Y, Li W et al. Primary solid lung cancerous nodules with different sizes: computed tomography features and their variations. BMC Cancer. 20109;19(1). BioMed Central. https://doi.org/10.1186/s12885-019-6274-0
8. de Margerie-Mellon C, Chassagnon G. Artificial intelligence: A critical review of applications for lung nodule and lung cancer. Diagn Interv Imaging. 2023;104(1):11–7. https://doi.org/10.1016/j.diii.2022.11.007.
9. Demir F, Akbulut Y, Taşcı B, Demir K. Improving brain tumor classification performance with an effective approach based on new deep learning model named 3ACL from 3D MRI data. Biomed Signal Process Control. 2023;81:104424. https://doi.org/10.1016/j.bspc.2022.104424.
10. EL-Bana S, Al-Kabbany A, Sharkas M. A Two-Stage framework for automated malignant pulmonary nodule detection in CT scans. Diagnostics. 2020;10(3):131. https://doi.org/10.3390/diagnostics10030131.
11. Everingham M, Van Gool L, Williams CKI, et al. The Pascal visual object classes (VOC) challenge. Int J Comput Vision. 2009;88(2):303–38. https://doi.org/10.1007/s11263-009-0275-4.
12. Feng C, Elazab A, Yang P, et al. Deep learning framework for alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM. IEEE Access. 2019;7:63605–18. https://doi.org/10.1109/access.2019.2913847.
13. Gong J, Liu J, Hao W et al. (2019) A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images. Eur Radiol. 2019. https://doi.org/10.1007/s00330-019-06533-w
14. González Maldonado S, Delorme S, Hüsing A, et al. Evaluation of prediction models for identifying malignancy in pulmonary nodules detected via Low-Dose computed tomography. JAMA Netw Open. 2020;3(2):e1921221. https://doi.org/10.1001/jamanetworkopen.2019.21221.
15. Hendrix W, Hendrix N, Scholten ET, Mourits M, Trap-de Jong J, Schalekamp S, Jacobs C. Deep learning for the detection of benign and malignant pulmonary nodules in non-screening chest CT scans. Commun Med. 2023;3(1):156. https://doi.org/10.1038/s43856-023-00388-5.
16. International Early Lung Cancer Action Program Investigators, Henschke CI, Yankelevitz DF et al. (2006) Survival of patients with stage I lung cancer detected on CT screening. The New England J Med. 2006;355(17):1763–71. United States:. https://doi.org/10.1056/NEJMoa060476
17. Jaeger PF, Simon, Bickelhaupt S et al. Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection.

PMLR. PMLR: 2020;171–183. Available at: https://proceedings.mlr.press/v116/jaeger20a . Accessed 30 Aug 2024.
18 Karalis VD. The integration of artificial intelligence into clinical practice. Appl Biosci. 2024;3(1):14–44. https://doi.org/10.3390/applbiosci3010002.
19 Khan AN, Al-Jahdali HH, Irion KL, et al. Solitary pulmonary nodule: A diagnostic algorithm in the light of current imaging technique. Avicenna J Med. 2011;1(2):39–51. https://doi.org/10.4103/2231-0770.90915.
20 Lanjewar MG, Panchbhai KG, Charanarur P. Lung cancer detection from CT scans using modified densenet with feature selection methods and ML classifiers. Expert Syst Appl. 2023;224:119961. https://doi.org/10.1016/j.eswa.2023.119961.
21 Lanjewar MG, Panchbhai KG, Charanarur P. Small size CNN-Based COVID-19 disease prediction system using CT scan images on PaaS cloud. Multimedia Tools Appl. 2024;83(21):60655–87. https://doi.org/10.1007/s11042-023-17884-4.
22 Lanjewar MG, Parab JS, Shaikh AY. Development of framework by combining CNN with KNN to detect alzheimer's disease using MRI images. Multimedia Tools Appl. 2023;82(8):12699–717. https://doi.org/10.1007/s11042-022-13935-4.
23 Lanjewar MG, Parab JS, Shaikh AY, Sequeira M. CNN with machine learning approaches using extratreesclassifier and MRMR feature selection techniques to detect liver diseases on cloud. Cluster Comput. 2023;26(6):3657–72. https://doi.org/10.1007/s10586-022-03752-7.
24 Lassau N, Théo Estienne P, de Vomecourt et al. Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI. *Diagnostic and interventional imaging* 100(4). Elsevier BV: 2019:199–209. https://doi.org/10.1016/j.diii.2019.02.001
25 Lassau N, Bousaid I, Chouzenoux E, et al. Three artificial intelligence data challenges based on CT and MRI. Diagn Interv Imaging. 2020;101(12):783–8. https://doi.org/10.1016/j.diii.2020.03.006.
26 Li D, Mikela Vilmun B, Frederik Carlsen J, et al. The performance of deep learning algorithms on automatic pulmonary nodule detection and classification tested on different datasets that are not derived from LIDC-IDRI: A systematic review. Diagnostics. 2019;9(4):207. https://doi.org/10.3390/diagnostics9040207.
27 Li W, Cao P, Zhao D, et al. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. Comput Math Methods Med. 2016;2016:1–7. https://doi.org/10.1155/2016/6215085.
28 Liao F, Liang M, Li Z, et al. Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky Noisy-OR network. IEEE Trans Neural Networks Learn Syst. 2019;30(11):3484–95. https://doi.org/10.1109/tnnls.2019.2892409.
29 Lin T-Y, Goyal P, Girshick R et al. Focal Loss for Dense Object Detection. 2017. Available at: https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html
30 Monkam P, Qi S, Ma H, et al. Detection and classification of pulmonary nodules using convolutional neural networks: A survey. IEEE Access. 2019;7:78075–91. https://doi.org/10.1109/access.2019.2920980.
31 Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol Surg. 2017;12(10):1799–808. https://doi.org/10.1007/s11548-017-1605-6.
32 Oh K, Chung Y-C, Kim KW, et al. Classification and visualization of alzheimer's disease using volumetric convolutional neural network and transfer learning. Sci Rep. 2019;9(1). https://doi.org/10.1038/s41598-019-54548-6.
33 Ramaswamy S, Truong K. Pulmonary Nodule Classification with Convolutional Neural Networks. 2016. Available at: http://vision.stanford.edu/teaching/cs231n/reports/2016/pdfs/324_Report.pdf . Accessed 30 Aug 2024.
34 Sanz H, Valim C, Vegas E, et al. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinformatics. 2018;19(1). https://doi.org/10.1186/s12859-018-2451-4.
35 Schreuder A, Scholten ET, van Ginneken B, et al. Artificial intelligence for detecting and characterizing pulmonary nodules in lung cancer CT screening: ready for practice? Translational Lung Cancer Res. 2021;10(5):2378–88. https://doi.org/10.21037/tlcr-2020-lcs-06.
36 Setio AAA, Traverso A, de Bel T et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical image analysis* 42. Netherlands: 2017; 1–13. https://doi.org/10.1016/j.media.2017.06.015
37 The National Lung Screening Trial Research Team. Reduced Lung-Cancer mortality with Low-Dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409. https://doi.org/10.1056/nejmoa1102873.
38 Tu X, Xie M, Gao J, et al. Automatic categorization and scoring of solid, Part-Solid and Non-Solid pulmonary nodules in CT images with convolutional neural network. Sci Rep. 2017;7(1):8533. https://doi.org/10.1038/s41598-017-08040-8.
39 Waymel Q, Badr S, Demondion X, et al. Impact of the rise of artificial intelligence in radiology: what do radiologists think? Diagn Interv Imaging. 2019;100(6):327–36. https://doi.org/10.1016/j.diii.2019.03.015.
40 Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Translational Lung Cancer Res. 2017;6(1):86–91. https://doi.org/10.21037/tlcr.2017.01.04.
41 Zhang G, Jiang S, Yang Z, et al. Automatic nodule detection for lung cancer in CT images: A review. Comput Biol Med. 2018;103:287–300. https://doi.org/10.1016/j.compbiomed.2018.10.033.
42 Zhao X, Liu L, Qi S, et al. Agile convolutional neural network for pulmonary nodule classification using CT images. Int J Comput Assist Radiol Surg. 2018;13(4):585–95. https://doi.org/10.1007/s11548-017-1696-0.

Publisher's Note