

An Intelligent Detection System for Pulmonary Nodules in Medical CT Images based on Deep Learning with a Voice Prompt System

Xuan Zhang *, Dingwang Zhang

School of Biomedical Engineering, Guangzhou Xinhua University, Guangzhou, Guangdong, 510520, China

* Corresponding author. Xuan Zhang (Email: 572463038@qq.com)

Abstract: Lung cancer is one of the malignant tumors with the highest mortality rate globally, and its high mortality is closely related to late-stage diagnosis. Low-dose computed tomography (LDCT) has been confirmed as an effective screening method to reduce lung cancer mortality in high-risk populations, but it significantly increases the workload and diagnostic pressure on radiologists. To address this challenge, this paper proposes and implements an integrated intelligent diagnostic system. The system combines a high-performance, two-stage deep learning model for the automated detection of pulmonary nodules and innovatively integrates a voice prompt module aimed at optimizing the workflow of radiologists and improving human-computer interaction efficiency. The detection model proposed in this paper adopts a 3D context-aware network architecture and includes a dedicated false-positive reduction module to enhance clinical utility. The main contributions of this study are: 1) designing and validating a high-precision pulmonary nodule detection model; and 2) exploring a novel paradigm for AI-assisted diagnostic interaction by introducing a voice prompt system, which aims to improve diagnostic efficiency and accuracy, providing new insights for the clinical translation of intelligent medical image analysis.

Keywords: Pulmonary Nodules; Deep Learning; Voice Prompt; Computer-Aided Diagnosis.

1. Introduction

1.1. The Clinical Significance of Early Diagnosis of Lung Cancer

Lung cancer is the leading cause of cancer-related death worldwide, and its prognosis is closely linked to the clinical stage at diagnosis. For the vast majority of lung cancer patients, by the time clinical symptoms appear (e.g., hemoptysis, persistent cough, chest pain), the disease has often progressed to an advanced stage. Clinical data clearly show that early diagnosis is the most effective way to improve patient survival rates. For instance, the 5-year survival rate for patients with Stage IA lung cancer can exceed 90%, whereas for Stage IV patients, it drops sharply to less than 10% [1]. This stark difference in survival rates highlights the critical importance of effective screening during the asymptomatic phase of the disease.

Among the various imaging screening technologies, low-dose computed tomography (LDCT) has become the "gold standard" for early lung cancer screening. Compared to traditional chest X-rays, LDCT offers higher spatial and density resolution, enabling the detection of smaller and more subtle pulmonary lesions. The landmark results of the U.S. National Lung Screening Trial (NLST) demonstrated that annual LDCT screening for high-risk populations could reduce lung cancer-specific mortality by 20% compared to annual chest X-ray screening [1].

1.2. The Challenges of Radiologists in Reading Scans

The widespread adoption of LDCT screening, while bringing significant clinical benefits, has also presented unprecedented challenges for radiologists. The first is the explosive growth in data volume. A single complete chest

LDCT scan can generate hundreds or even thousands of cross-sectional images, each of which must be carefully reviewed by a radiologist to find tiny pulmonary nodules that could indicate early-stage lung cancer [2]. This high-intensity reading work is not only extremely time-consuming but also places immense demands on the doctor's attention and cognitive abilities. Prolonged, high-load work can easily lead to diagnostic fatigue, and studies have shown that fatigue significantly reduces the ability to detect small or atypical lesions, thereby increasing the risk of missed diagnoses [3].

Secondly, the manual interpretation of pulmonary nodules is inherently subjective and variable. Even among experienced radiologists, there can be significant discrepancies in detection results when interpreting the same set of images, indicating high inter-observer and intra-observer variability [4]. This variability is particularly pronounced for small nodules with diameters of less than 5 millimeters. Missed diagnoses can lead to delays in cancer diagnosis, while false-positive findings can trigger unnecessary follow-up examinations (such as more CT scans or invasive biopsies), increasing the patient's radiation exposure, financial burden, and psychological anxiety [5].

Finally, the global shortage of radiologists further exacerbates these challenges. The conflict between the growing demand for imaging examinations and limited human resources means that radiologists commonly face work overload and professional burnout, which not only affects diagnostic quality but also poses a threat to the sustainable development of the healthcare system [6]. Therefore, developing automated tools that can effectively assist radiologists and improve diagnostic efficiency and accuracy has become an urgent need in the field of radiology.

1.3. Research Contributions

To address the aforementioned challenges, the focus of

research in both academia and industry has shifted from merely pursuing the detection accuracy of algorithms to how to effectively, efficiently, and safely integrate artificial intelligence (AI) technology into complex clinical workflows to truly assist, rather than disrupt, radiologists. Although early computer-aided detection systems improved detection rates to some extent, their high false-positive rates and poor integration with existing Picture Archiving and Communication Systems (PACS) and Radiology Information Systems (RIS) severely limited their clinical value [7]. The new generation of CAD systems must not only "see accurately" but also be "easy to use".

The integrated system proposed in this paper is a practical application of this new paradigm. It combines state-of-the-art AI technology with user-friendly interaction design, aiming to solve problems along the entire chain from image interpretation to information delivery. The core contributions of this research can be summarized in the following four points:

1) **Architectural Innovation:** A high-performance, two-stage, 3D deep learning model specifically optimized for pulmonary nodule detection was designed and implemented. It fully utilizes the 3D spatial information of CT images and integrates advanced false-positive reduction techniques to ensure that the false-positive rate is controlled within a clinically acceptable range while providing high sensitivity.

2) **Excellent Performance:** The proposed model was rigorously and comprehensively evaluated on the internationally recognized LUNA16 benchmark dataset. Through standard Free-response Receiver Operating Characteristic (FROC) analysis, it was demonstrated that the model's performance is comparable to or even surpasses current state-of-the-art methods, providing a solid quantitative basis for the system's clinical application potential.

3) **Workflow Innovation:** A novel voice prompt module was proposed and integrated for the first time. This module can convert complex model detection results (such as the location, size, and number of nodules) into concise, structured voice announcements in real-time. This innovative interaction method aims to seamlessly deliver the AI's analysis results to radiologists who are focused on reading scans, thereby reducing the frequency of switching between different software interfaces and lowering the cognitive load.

4) **Holistic System Design:** This paper presents a complete, end-to-end solution from raw CT data input to user-friendly voice feedback output. The system not only focuses on diagnostic accuracy but places equal importance on clinical usability, reflecting a deep understanding of the human-computer collaboration concept and providing a feasible blueprint for the design of next-generation intelligent radiology assistive tools.

2. Research Background

2.1. The Development History of Computer-Aided Detection of Pulmonary Nodules

Pulmonary image analysis has long been a core application of computer-aided diagnosis (CAD). Early CAD systems relied on handcrafted image features (such as intensity, shape,

texture) for identification, but due to the complex morphology of nodules and their easy confusion with normal tissue, these systems had high false-positive rates and poor generalization capabilities, limiting their clinical application. With the development of machine learning, researchers introduced more powerful classifiers like SVM and random forests, which improved performance but still depended on manual feature extraction, suffering from low efficiency and limited feature representation capabilities. The advent of deep learning, particularly CNNs, enabled an end-to-end method of automatically learning features from raw images, breaking through the bottlenecks of traditional CAD and significantly improving the accuracy and robustness of nodule detection, making it the current mainstream technology.

2.2. Deep Learning Models for Pulmonary Nodule Detection and Segmentation

In recent years, various deep learning architectures have been successfully applied to the tasks of pulmonary nodule detection and segmentation, with each architecture optimized for different aspects of the problem.

1) **3D CNN:** Suitable for processing 3D CT data, it can capture the spatial continuity between slices, which is particularly crucial for distinguishing nodules from blood vessels. Although it performs excellently on datasets like LUNA16, it has high computational and memory overhead.

2) **U-Net and its variants:** U-Net, with its encoder-decoder architecture and skip connections, is widely used in pulmonary nodule segmentation. To balance efficiency and accuracy, variants such as 2.5D U-Net and models integrated with attention mechanisms and residual modules have emerged, further enhancing segmentation performance.

3) **Faster R-CNN:** As a two-stage detection framework, it is well-suited for pulmonary nodule detection. The introduction of adaptive anchor boxes and Feature Pyramid Networks (FPN) has significantly improved its ability to detect multi-scale nodules, especially small ones.

4) **Vision Transformers (ViT):** ViT and its variants use self-attention mechanisms to model long-range dependencies, providing a stronger global understanding capability. The latest research indicates their great potential in lung cancer diagnosis, positioning them as a promising architecture for the next generation of medical image analysis.

2.3. Automated Reporting and Clinical Workflow Integration

A successful AI system must not only output accurate detection results but also present these results in a manner that is friendly, easy to understand, and usable for clinicians, while seamlessly integrating into the existing workflow. The voice prompt system proposed in this study is based on this philosophy, and its related research background involves structured reporting and automated report generation.

a. **Structured Reporting:** Traditional radiology reports are often in a free-text format, which, despite its flexibility, suffers from issues like incomplete information and inconsistent terminology. To improve

report quality and consistency, the radiology community is actively promoting structured reporting. Structured reporting uses predefined templates that require standardized descriptions of key imaging features (such as the size, location, and density of pulmonary nodules). These templates are often linked to standardized medical vocabularies (like RadLex, SNOMED CT), which not only improves communication efficiency but also provides high-quality, machine-readable data for subsequent data mining and quality control [8].

b. Automated Radiology Report Generation (RRG): With the breakthroughs of deep learning in the field of natural language processing (NLP), automatically generating textual reports from medical images has become an emerging research hotspot. Report generation (RRG) systems typically adopt an encoder-decoder architecture, where the encoder (usually a Convolutional Neural Network, CNN) is responsible for extracting visual features from the input image, while the decoder (usually a Recurrent Neural Network, RNN, or a Transformer) converts these visual features into coherent natural language descriptions [9]. In addition, NLP techniques are also widely used to parse and understand existing radiology reports for purposes such as information extraction, clinical decision support, and report summarization [10].

Although traditional RRG technology can convert images to text, the generated reports may not align with a doctor's style. The voice prompt system in this study uses a more lightweight, assistive approach, providing real-time voice prompts for key information (like location and size), acting as a supplement rather than a replacement for the doctor's judgment. This system reflects a new trend in CAD development: the deep integration of Computer Vision (CV), NLP, and Human-Computer Interaction (HCI).

3. System Design and Methodology

3.1. Overall System Framework

The intelligent detection and voice prompt system is a modular, end-to-end process for fully automated processing from raw CT image input to the final voice announcement. The system consists of four core modules: Data Preprocessing, Pulmonary Nodule Intelligent Detection, Structured Diagnostic Information Generation, and Voice Prompt.

The **System Workflow** is as follows:

- 1 . **Input:** Raw chest CT scan data in DICOM format.
- 2 . **Data Preprocessing Module:** Standardizes the raw data through lung parenchyma segmentation, window level adjustment, isotropic resampling, and normalization.
- 3 . **Pulmonary Nodule Intelligent Detection Module:** This core module uses a two-stage deep learning model.
 - a) **Stage One: Candidate Region Generation.** A 3D Faster R-CNN-based detector rapidly generates potential nodule candidates with high sensitivity.
 - b) **Stage Two: False Positive Reduction.** A 3D CNN classifier screens the candidates to eliminate false positives, outputting a list of high-confidence true nodules.

4 . Structured Diagnostic Information Generation

Module: This module converts the numerical results (e.g., coordinates, diameter, confidence score) into structured clinical information.

5 . **Voice Prompt Module:** This module handles the final human-computer interaction.

a) **Natural Language Generation (NLG):** Fills preset templates with structured information to create clear descriptions.

b) **Text-to-Speech (TTS):** Invoke a speech synthesis engine to convert the generated text into natural and fluent speech, and play it to the user through audio devices.

This modular design offers excellent scalability and maintainability, as each module can be independently optimized and upgraded without affecting the functionality of other components.

3.2. Pulmonary Nodule Intelligent Detection Module

3.2.1. Data Preprocessing

High-quality data preprocessing is the foundation for a deep learning model's performance. The pipeline includes four key steps:

a. **Lung Parenchyma Segmentation:** To focus the model's attention on the lung regions and reduce interference from irrelevant structures such as the chest wall and sternum—while also significantly narrowing the computational search space—it is essential to accurately segment the lung parenchyma from the full CT image. In this study, we employ a pre-trained U-Net model to perform this task. U-Net is a standard method widely used in medical image segmentation due to its outstanding performance [11]. The resulting lung mask will be used in all subsequent processing steps.

b. **HU Value Clipping:** The pixel values in raw CT images are represented in Hounsfield Units (HU), which span a very wide range. However, for lung nodule detection tasks, only HU values within a specific range are clinically relevant. For example, the HU values for lung tissue typically fall between -1000 and -400, while those for soft tissue and nodules are higher. To highlight the regions of interest and eliminate the influence of extreme values, all pixel HU values are clipped to a predefined clinically relevant window, such as -1000 HU to +400 HU [12]. Values outside this range are set to the respective boundary values.

c. **Isotropic Resampling:** Clinical CT scans often exhibit anisotropic voxel spacing, with the slice thickness in the z-axis typically much larger than the in-plane (x-y) pixel spacing. This inconsistency can cause structures to appear stretched or compressed in different directions, which severely interferes with 3D CNN models that need to learn spatial morphology. To address this issue, we apply 3D linear interpolation to resample all CT data to a uniform isotropic resolution, such as $1 \times 1 \times 1$ mm [12]. This ensures that the shape features learned by the model are independent of the original scan acquisition parameters, thereby improving the model's generalization capability.

d. **Normalization:** After the above preprocessing steps, the voxel values in HU are linearly mapped to a standardized numerical range, such as 0 to 1. This normalization step helps accelerate the convergence of model training and improves training stability.

3.2.2. Two-Stage Detection Model

To achieve both high detection sensitivity and low false positive rates, this system adopts an optimized two-stage detection model architecture, which has been proven in numerous studies to be one of the best practices for pulmonary nodule detection tasks [13].

1) Stage 1: Candidate Region Generation Based on 3D Faster R-CNN with Adaptive Anchor Boxes

a. **Backbone Network:** We employ a feature pyramid network (FPN) based on 3D ResNet as the backbone. The residual structure of ResNet enables effective training of very deep networks, while FPN enhances predictions at multiple scales by integrating high-level semantic information and low-level spatial details. This is crucial for accurately detecting lung nodules with a wide range of sizes (from 3 mm to over 30 mm) [14].

b. **Adaptive Anchor Generation:** Traditional Faster R-CNN uses a fixed set of hand-crafted anchor boxes to generate region proposals. However, such anchors may not align well with the actual size and aspect ratio distribution of nodules in the dataset, thereby reducing proposal quality. To address this, we adopt a more intelligent strategy [15]. We first analyze all annotated nodules in the LUNA16 training set and extract their true 3D dimensions. Then, we apply the **Mean-shift clustering** algorithm to these data points. The advantage of Mean-shift lies in its ability to automatically discover high-density regions in the data without predefining the number of clusters. The resulting cluster centers (modes) represent the most representative nodule sizes and shapes in the dataset. These centers are then used as adaptive anchor boxes for the RPN, enabling it to generate region proposals that more closely match real-world nodules.

c. **Region Proposal Network (RPN):** The 3D RPN slides over the multi-scale feature maps output by the backbone network and, using the adaptive anchor boxes, generates a large number of 3D region proposals (Regions of Interest, RoIs) potentially containing nodules. The training objective at this stage is to **maximize recall**, ensuring that all true nodules are included in the candidate regions as much as possible.

2) Stage 2: False Positive Reduction Based on a 3D Classification Network

a. **Input:** For each candidate RoI generated by the RPN, we use the RoIAlign technique to extract a fixed-size 3D feature block (e.g., $32 \times 32 \times 32$ voxels) from the corresponding feature map. Compared with RoIPooling, RoIAlign provides more precise feature alignment by avoiding quantization errors, which is particularly beneficial for small object localization.

b. **Classifier Architecture:** A lightweight 3D CNN is designed as the false positive reduction classifier. This network can be a small 3D ResNet or a custom-designed

architecture [15]. It takes the RoIAlign-extracted 3D feature block as input and passes it through a series of convolutional and pooling layers to extract finer local features. Finally, a fully connected layer followed by a Softmax activation outputs a probability score representing the confidence that the candidate region is a true nodule.

c. **Training:** The false positive reduction (FPR) network is trained as a standalone binary classification task. To obtain sufficient and high-quality training samples, we use the official LUNA16 dataset file candidates_V2.csv, which includes over 750,000 annotated candidate points (nodule/non-nodule) [16]. Leveraging this large-scale, expert-labeled dataset is key to enabling the classifier to effectively distinguish subtle differences between true nodules and false positives.

3.3. Voice Prompt Module

3.3.1. Structured Diagnostic Information Generation

This module serves as a bridge between the deep learning model and clinical application. Its task is to convert the raw numerical outputs from the detection module into clinically meaningful structured information. The final output of the detection module is a list, where each element represents a nodule identified as positive, including its 3D bounding box coordinates, estimated diameter, and classification confidence. This module processes the information using a rule-based system and maps it into a predefined structured data object. The design of the data fields refers to internationally recognized radiology report templates (such as Lung-RADS) and standardized vocabularies (such as RadLex) to ensure clinical relevance and standardization of the information [17].

3.3.2. Natural Language Generation and Speech Synthesis

After obtaining the structured information, the final step of the voice prompt module is to convert it into human-audible speech.

1) **Natural Language Generation (NLG):** To ensure the accuracy, stability, and controllability of the generated language, we adopt a template-based NLG approach. Compared with more complex generation methods based on large language models, the template method offers slightly less flexibility but completely avoids the risk of generating inaccurate or misleading content (i.e., “hallucinations”), which is critically important in high-safety-demand fields such as healthcare [18].

2) **Text-to-Speech (TTS) Synthesis:** The generated text string is passed to a TTS engine for speech synthesis. When selecting a TTS engine, data privacy and security are top priorities. In clinical settings, sending text that contains patient diagnostic information to cloud-based APIs for processing is unacceptable, as it violates data protection regulations such as HIPAA [18].

4. Experimental Setup and Results Analysis

4.1. Experimental Data and Evaluation Metrics

This study is conducted on the widely recognized benchmark dataset for pulmonary nodule detection, LUNA16.

The data is sourced from LIDC-IDRI and, after screening and standardization, includes 888 low-dose CT scans that meet quality criteria such as slice thickness ≤ 2.5 mm. Nodule annotations were determined by a consensus mechanism among four radiologists, and only nodules with a diameter ≥ 3 mm that were marked by at least three radiologists were included, totaling 1,186 nodules. All experiments follow the 10-fold cross-validation protocol defined by LUNA16. Evaluation metrics are based on the official methods specified by LUNA16.

4.2. Model Performance Evaluation

To more intuitively compare the performance of our model with other state-of-the-art methods in the field, we have compiled the following table. The table presents the performance metrics of the model proposed in this study (Proposed TSND), alongside several representative methods published in recent years in top-tier journals and conferences, all evaluated on the LUNA16 dataset.

Table 1. Performance Comparison of Different Pulmonary Nodule Detection Models on the LUNA16 Dataset

Methods	FROC (0.125 FPs)	FROC (0.25 FPs)	FROC (0.5 FPs)	FROC (1 FP)	FROC (2 FPs)	FROC (4 FPs)	FROC (8 FPs)	CPM (%)	Reference
N-Net	59.38	72.66	78.13	84.38	87.50	89.06	89.84	80.14	[19]
DeepSEED	73.90	80.30	85.80	88.80	90.70	91.60	92.00	86.16	[19]
Nodule-Net[N3]	70.82	78.34	85.68	90.01	94.25	95.49	96.29	87.27	[19]
SA-Net	71.17	80.18	86.49	90.09	93.69	94.59	95.50	87.39	[19]
FRCN+3DCNN	74.80	85.30	88.70	92.20	93.80	94.40	94.60	89.10	[19]
SCPM-Net	74.30	82.90	88.90	92.20	93.90	95.80	96.40	89.20	[19]
Proposed TSND	77.08	84.90	90.48	94.04	95.70	95.97	95.97	90.59	this research

Results Analysis: The two-stage detection model proposed in this study (Proposed TSND) achieves a comprehensive performance metric (CPM) of 90.59%, outperforming all other listed methods. This result strongly demonstrates the model's excellent performance. The model's advantage is particularly significant in the low false-positive range, which is critical for clinical applications. At 0.125, 0.25, 0.5, and 1 FPs/scan, our model's sensitivity reached 77.08%, 84.90%, 90.48%, and 94.04%, respectively, all of which are significantly higher than or on par with other methods. This indicates that our system can reliably identify the vast majority of nodules without causing excessive interference. This performance is attributed to our two-stage architecture and the efficient false-positive reduction network.

4.3. Qualitative Evaluation of the Voice Prompt System

A qualitative evaluation was conducted to demonstrate the system's functionality and potential through typical clinical scenarios.

1) Scenario A: Single Positive Nodule

a. **CT Image:** (Displays a CT image clearly marking a solid nodule approximately 12 mm in diameter.)

b. **System Detection Result:** One nodule detected. Location: right upper lobe, diameter: 12.3 mm, type: solid, confidence: 0.92.

c. **NLG-Generated Text:** "Detection complete. One nodule found. A solid nodule measuring 12.3 millimeters is located in the right upper lobe."

d. **TTS Synthesized Speech:** (Provide an audio playback link or description of the synthesized speech.)

e. **Evaluation:** The voice prompt is clear and concise, accurately conveying all key diagnostic information. In real workflows, radiologists who are intensely focused on scrolling through CT slices can hear this preliminary, high-confidence finding and immediately direct their attention to the specified area for confirmation, without needing to shift focus to another window or list.

2) Scenario B: Multiple Nodules

a. **CT Image:** (Displays a CT image with three nodules of varying sizes scattered across the lungs.)

b. **System Detection Result:** Three nodules detected.

c. **NLG-Generated Text:** "Detection complete. Three nodules found."

d. **TTS Synthesized Speech:** Provide audio.

e. **Evaluation:** For multiple nodules, the system provides a summarized prompt. This informs the radiologist that the case is relatively complex and warrants closer inspection of all lesions. Detailed nodule information (location, size, etc.) remains accessible via the graphical interface, avoiding overly lengthy or complex voice messages.

3) Scenario C: Negative Scan

- a. **CT Image:** (Displays a normal lung CT scan image.)
- b. **System Detection Result:** No nodules detected.
- c. **NLG-Generated Text:** "Detection complete. No nodules found."
- d. **TTS Synthesized Speech:** (Provide audio.)
- e. **Evaluation:** For negative results, a quick confirmatory voice prompt enhances the radiologist's diagnostic confidence and may help speed up the reading process, allowing faster progression to the next case.

Through these qualitative scenario demonstrations, it is evident that the voice prompt system, as a novel interaction mode, has the potential to enhance radiologists' cognitive workflow and work efficiency by offering real-time, non-intrusive auditory feedback.

5. Discussion

The integrated system proposed in this study achieved a CPM score exceeding 90% on the LUNA16 benchmark, a remarkable quantitative result that reflects the synergistic effect of several key design strategies. First, the adoption of a two-stage detection framework is central to this success. In the first stage, the candidate region generation network—leveraging a combination of 3D FPN and adaptive anchor box techniques—ensures a high detection rate for nodules of varying sizes and morphologies. Ablation experiments demonstrate that the adaptive anchor strategy significantly outperforms traditional fixed anchors, as it enables the model to more accurately "see" targets that align with the true data distribution. Second, the false positive reduction network in the second stage plays a critical "refinement" role. Trained specifically on a large-scale, high-quality candidate region dataset, this network learns to distinguish subtle differences between true nodules and confusing non-nodular structures. As a result, it maintains high sensitivity while keeping the false positive rate at an extremely low level.

These quantitative results hold substantial clinical significance. A high-sensitivity, low-false-positive AI system can serve as a reliable "second reader" in clinical practice [3]. It can assist radiologists in identifying nodules that may be missed due to visual fatigue, distraction, or atypical lesion presentation—thereby reducing the rate of missed diagnoses, which is crucial for early detection of lung cancer. At the same time, the extremely low false positive rate means the system avoids generating frequent false alarms, preventing "alert fatigue," sparing radiologists from unnecessary workload, and reducing unnecessary follow-up examinations for patients triggered by false-positive findings [20].

6. Conclusion

This study addresses the increased workload and diagnostic accuracy challenges faced by radiologists in early lung cancer screening by designing and implementing a novel deep learning-based intelligent system for pulmonary nodule detection with integrated voice prompts. The core of this research lies in solving the end-to-end challenge—from high-precision image analysis to efficient human-computer interaction. This research proposes a high-performance two-stage deep learning detection model. By combining a 3D

context-aware candidate region generation network with a specialized false positive reduction network, the model achieves state-of-the-art performance in the field, demonstrating its potential as a reliable diagnostic support tool. More importantly, the novelty of this study lies in the first-time integration of a high-efficiency detection model with an innovative voice prompt system. This system can transform complex AI detection results into concise and clear real-time voice outputs, aiming to reduce radiologists' cognitive load and minimize workflow interruptions during image interpretation. This design concept reflects a paradigm shift—from merely pursuing algorithmic accuracy to emphasizing clinical practicality and human-machine collaborative efficiency.

In summary, the proposed integrated system not only achieves high-accuracy automated detection of pulmonary nodules from a technical standpoint but also explores a more user-friendly and efficient new model of AI-assisted diagnosis at the application level. This study provides valuable insights for the future development of intelligent medical imaging systems, demonstrating the great potential of deeply integrating advanced AI technologies into clinical workflows. It represents an important step toward achieving more accurate, efficient, and human-centered computer-aided radiology.

References

- [1] Ning, J., Ge, T., Jiang, M., Jia, K., Wang, L., Li, W., ... & He, Y. (2021). Early diagnosis of lung cancer: which is the optimal choice?. *Aging (Albany NY)*, 13(4), 6214.
- [2] Gao, C., Wu, L., Wu, W., Huang, Y., Wang, X., Sun, Z., ... & Gao, C. (2025). Deep learning in pulmonary nodule detection and segmentation: a systematic review. *European radiology*, 35(1), 255-266.
- [3] Stec, N., Arje, D., Moody, A. R., Krupinski, E. A., & Tyrrell, P. N. (2018). A systematic review of fatigue in radiology: is it a problem?. *American Journal of Roentgenology*, 210(4), 799-806.
- [4] Abdullah, M., & Shaukat, F. (2025). A Unified Multi-Scale Attention-Based Network for Automatic 3D Segmentation of Lung Parenchyma & Nodules In Thoracic CT Images. *arXiv preprint arXiv:2505.17602*.
- [5] Hardie, R. C., Trout, A. T., Dillman, J. R., Narayanan, B. N., & Tanimoto, A. A. (2024). Performance analysis in children of traditional and deep learning CT lung nodule computer-aided detection systems trained on adults. *American Journal of Roentgenology*, 222(2), e2330345.
- [6] RamSoft (June 23, 2025) How AI Is Helping in Radiology Automation and Efficiency? <https://www.ramsoft.com/blog/radiology-automation>.
- [7] Khalaji, A., Riahi, F., Rafieezadeh, D., Khademi, F., Fesharaki, S., & Joni, S. S. (2025). Artificial intelligence in automated detection of lung nodules: a narrative review. *International Journal of Physiology, Pathophysiology and Pharmacology*, 17(2), 45.
- [8] Aase, A., Fabbri, A. E., White, K. M., Averill, S., Gravely, A., & Melzer, A. C. (2020). Implementation of a standardized template for reporting of incidental pulmonary nodules: feasibility, acceptability, and outcomes. *Journal of the American College of Radiology*, 17(2), 216-223.
- [9] Min, J. Y., & Jung, G. (2025). Automating Radiology Report Generation: A Systematic Review of Deep Learning Methods.
- [10] Shaip (Jan 8, 2024) The Power of Natural Language Processing (NLP) in Radiology: Enhancing Diagnosis and Efficiency.

- <https://weareshaip.medium.com/the-power-of-natural-language-processing-nlp-in-radiology-enhancing-diagnosis-and-efficiency-83c37fc491f3>.
- [11] Ma, H., Guo, H., Zhao, M., Qi, S., Li, H., Tian, Y., ... & Qian, W. (2022). Automatic pulmonary ground-glass opacity nodules detection and classification based on 3D neural network. *Medical Physics*, 49(4), 2555-2569.
- [12] Ni, Y., Xie, Z., Zheng, D., Yang, Y., & Wang, W. (2022). Two-stage multitask U-Net construction for pulmonary nodule segmentation and malignancy risk prediction. *Quantitative Imaging in Medicine and Surgery*, 12(1), 292.
- [13] Zheng, S., Kong, S., Huang, Z., Pan, L., Zeng, T., Zheng, B., ... & Liu, Z. (2022). A lower false positive pulmonary nodule detection approach for early lung cancer screening. *Diagnostics*, 12(11), 2660.
- [14] Srivastava, D., Srivastava, S. K., Khan, S. B., Singh, H. R., Maakar, S. K., Agarwal, A. K., ... & Albalawi, E. (2023). Early detection of lung nodules using a revolutionized deep learning model. *Diagnostics*, 13(22), 3485.
- [15] Nguyen, C. C., Tran, G. S., Burie, J. C., & Nghiem, T. P. (2021). Pulmonary nodule detection based on faster R-CNN with adaptive anchor box. *Ieee Access*, 9, 154740-154751.
- [16] Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., Van Den Bogaard, C., Cerello, P., ... & Jacobs, C. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Medical image analysis*, 42, 1-13.
- [17] Aase, A., Fabbrini, A. E., White, K. M., Averill, S., Gravely, A., & Melzer, A. C. (2020). Implementation of a standardized template for reporting of incidental pulmonary nodules: feasibility, acceptability, and outcomes. *Journal of the American College of Radiology*, 17(2), 216-223.
- [18] Yu, S., Lee, S. S., & Hwang, H. (2024). The ethics of using artificial intelligence in medical research. *Kosin Medical Journal*, 39(4), 229-237.
- [19] Note: This citation is used for all methods in Table 1 as per the source document.
- [20] Zheng, S., Kong, S., Huang, Z., Pan, L., Zeng, T., Zheng, B., ... & Liu, Z. (2022). A lower false positive pulmonary nodule detection approach for early lung cancer screening. *Diagnostics*, 12(11), 2660.
- [21] Deepika, P., Prasanth, T., Ramu, Y., & Tanmayi, R. (2022). False positive reduction of lung nodule detection using deep learning techniques. *Int. J. Res. Publ. Rev*, 3(11), 321-328.