Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 15 [3] February 2026. 50-54

Online ISSN 2277-1808
Journal's URL: http://www.bepls.com
CODEN: BEPLAD

ORIGINAL ARTICLE **OPEN ACCESS**

Development of automated AI algorithms for detecting liver and pulmonary lesions, renal masses, or mesenteric injuries on CT imaging

Nida Nawaz[1], Ali Assad[2], Muhammad Usman Farooq Baig[3], Roshna Qamar[4], Omna Younus[5], Muhammad Mubeen Ul Haq[6]
[1] Senior Registrar, Radiology, Rai Foundation Medical College, Pakistan.
[2] Consultant Radiologist, Chughtai Healthcare, Pakistan.
[3] Senior Registrar, Shalamar Medical and Dental College, Shalamar Hospital, Lahore, Pakistan.
[4] Assistant Professor, Radiology, Akhtar Saeed Medical and Dental College, Lahore, Pakistan.
[5] Senior Medical Officer, Akram Medical Complex, Lahore, Pakistan.
[6] Consultant Radiologist, Chughtai Healthcare.
Corresponding author: **Nida Nawaz 0009-0007-6731-1599**

ABSTRACT
The timely and accurate detection of critical abnormalities on computed tomography (CT) imaging—such as liver and pulmonary lesions, renal masses, and mesenteric injuries—is essential for clinical decision-making in emergency and oncology settings. Traditional human-based interpretation is reliable but time-intensive and subject to inter-reader variability. Artificial intelligence (AI) offers a promising approach to automate detection and prioritization of actionable findings on CT scans. The objective of this study was to develop and validate a set of automated deep learning algorithms for identifying liver lesions, pulmonary nodules, renal masses, and mesenteric injuries on CT imaging. A multicenter retrospective dataset comprising 10,000 contrast-enhanced CT studies with expert annotations was used to train convolutional neural network (CNN) models. Performance was evaluated using a held-out test set of 2,000 studies. The liver lesion detection model achieved an area under the receiver operating characteristic curve (AUROC) of 0.95 and sensitivity of 92.4% at 1.5 false positives per scan. Pulmonary lesion detection yielded an AUROC of 0.96 with 94.1% sensitivity at 1.2 false positives per scan. Renal mass detection achieved an AUROC of 0.93 and 90.7% sensitivity, while mesenteric injury detection demonstrated an AUROC of 0.91 with 88.3% sensitivity. In conclusion, the developed AI algorithms demonstrated high accuracy and potential for automated triage and clinical support in CT interpretation, offering rapid and sensitive detection of critical abnormalities.
Keywords: *Artificial intelligence, Computed tomography, Lesion detection, Deep learning, Automated diagnosis*

Received 29.12.2025 Revised 12.01.2026 Accepted 04.02.2026

INTRODUCTION

Computed tomography (CT) imaging is a cornerstone modality in modern clinical diagnostics, particularly for the detection of abdominal and thoracic pathology including liver lesions, pulmonary nodules, renal masses, and mesenteric injuries [1]. Accurate and rapid interpretation of CT scans is critical to patient outcomes, yet radiologist workload and global shortages of trained readers pose significant challenges [2]. Artificial intelligence (AI), particularly deep learning techniques such as convolutional neural networks (CNNs), has emerged as a viable method to learn complex image patterns and assist in automated detection of radiologic abnormalities [3]. Early studies have demonstrated that AI can approximate radiologist performance for specific tasks, such as pulmonary nodule detection on chest CT or liver lesion segmentation on contrast-enhanced studies [4].

Despite progress, most previously reported AI systems focus on single organ targets or limited datasets, and few encompass multiple critical lesion types in a unified framework [5]. Moreover, many models are developed using homogeneous datasets that may not generalize across scanners, protocols, and patient populations encountered in real-world practice [6].

Liver lesions and pulmonary nodules are frequently encountered in both oncology surveillance and emergency imaging, whereas renal masses and mesenteric injuries demand rapid identification in trauma

and acute care settings [7]. A comprehensive AI framework capable of reliably detecting these diverse findings would significantly enhance diagnostic workflows.
Therefore, this study aimed to develop and rigorously validate automated AI algorithms for detecting liver and pulmonary lesions, renal masses, and mesenteric injuries on CT imaging using a large, multicenter annotated dataset, evaluating algorithm performance against an expert reference standard.

MATERIAL AND METHODS

Study Design and Data Collection

A multicenter retrospective diagnostic accuracy study was conducted at Rai Foundation Medical College using de-identified contrast-enhanced CT scans collected from three tertiary care centers between January 2023 and December 2024.

Ethical Approval

The study was approved by the Institutional Ethical Review Committees of all participating centers under reference numbers ERC/AI-RAD/2018/127 with a waiver of informed consent due to retrospective data use.

Dataset and Annotation

A total of **12,000 contrast-enhanced CT studies** of the thorax and abdomen were retrospectively collected from multiple tertiary-care centers. Studies included arterial and portal venous phases where available. Prior to AI development, all scans underwent **standard CT interpretation** by radiologists as part of routine clinical reporting.
For AI training and validation, **expert annotation** was performed using a structured consensus protocol. Three board-certified radiologists (≥5 years' experience) independently reviewed each study, labeling:

- Liver lesions (benign and malignant)
- Pulmonary nodules ≥3 mm
- Renal masses
- Mesenteric injuries (including mesenteric hematoma, active bleeding, and vascular disruption)

Discrepancies were resolved through consensus review. These annotations served as the reference standard.

AI Model Development

Convolutional neural network architectures with 3D volumetric processing capabilities were developed separately for each detection task. Data were split into training (70%), validation (10%), and testing (20%) sets, ensuring balanced representation of all target findings.

Reference Standard

The reference standard comprised consensus annotations from at least three board-certified radiologists with ≥5 years of CT interpretation experience.

AI MODEL DEVELOPMENT AND 3D IMAGE PROCESSING

Separate three-dimensional convolutional neural network (3D CNN) architectures were developed for each target abnormality. Entire CT volumes were processed rather than individual slices, allowing the models to learn spatial and contextual relationships across axial, coronal, and sagittal planes.

- CT volumes were resampled to uniform voxel spacing
- Intensity normalization was applied
- The 3D CNN analyzed volumetric patches to detect lesions
- Output probability maps were generated and overlaid on CT images to highlight suspicious regions

This approach enabled improved detection of subtle findings, particularly small pulmonary nodules and mesenteric injuries that may be missed on 2D slice-based review.
Data were divided into training (70%), validation (10%), and testing (20%) sets with balanced representation of all lesion types.

Performance metrics and justification

Model performance was evaluated using:

- **AUROC**, to assess overall discriminative ability
- **Sensitivity**, prioritized to minimize missed critical findings
- **Specificity**, to assess false-positive control
- **False positives per scan**, reflecting clinical usability

Thresholds were selected to optimize sensitivity while maintaining acceptable false-positive rates, consistent with AI triage applications in radiology.
Performance metrics were calculated on the independent test set. 95% confidence intervals were estimated using bootstrapped resampling. AI performance was compared with baseline radiologist interpretation using McNemar's test for paired proportions, with statistical significance set at $p < 0.05$.

RESULTS

Table 1. Model Detection Performance Metrics on Test Set

Detection Task	AUROC	Sensitivity (%)	Specificity (%)	FP/Scan
Liver Lesions	0.95 (0.94–0.96)	92.4	89.1	1.5
Pulmonary Lesions	0.96 (0.95–0.97)	94.1	90.3	1.2
Renal Masses	0.93 (0.92–0.95)	90.7	88.0	1.7
Mesenteric Injuries	0.91 (0.89–0.93)	88.3	87.2	2.1

All algorithms demonstrated excellent discrimination with AUROCs ≥0.91. Pulmonary and liver lesion models achieved the highest sensitivity.

Table 2. Comparative Radiologist Performance on Subset (n=500)

Task	Radiologist Sensitivity (%)	AI Sensitivity (%)	p-value
Liver Lesions	90.8	92.4	0.14
Pulmonary Lesions	93.5	94.1	0.28
Renal Masses	89.2	90.7	0.22
Mesenteric Injuries	87.1	88.3	0.31

AI sensitivity was statistically comparable to expert radiologists across tasks.

Table 3. Subgroup Performance by Lesion Size

Lesion Size	Sensitivity (%)	Pulmonary	Liver
<6 mm	85.2	80.3	—
6–10 mm	90.4	88.7	—
>10 mm	96.8	95.5	—

The results indicate that the developed AI algorithms performed consistently well across all evaluated detection tasks on CT imaging, demonstrating strong diagnostic capability and clinical applicability.

Overall model performance showed excellent discriminative ability, with AUROC values ranging from 0.91 to 0.96. This indicates that the algorithms were effective in distinguishing true abnormalities from normal findings across a wide range of thoracic and abdominal pathologies. Liver and pulmonary lesion detection achieved the highest AUROC values, reflecting the robustness of the models for these more common and well-defined imaging targets.

Sensitivity values exceeded 88% for all detection tasks, suggesting that the majority of clinically relevant findings were correctly identified by the AI models. Pulmonary lesion detection showed the highest sensitivity at 94.1%, followed closely by liver lesion detection at 92.4%. These findings highlight the ability of the algorithms to detect subtle abnormalities that may be challenging during routine CT interpretation, particularly in time-sensitive clinical environments.

Specificity remained high across all tasks, ranging from 87.2% to 90.3%, indicating effective control of false-positive findings. The false positives per scan were relatively low, with pulmonary and liver lesion models generating fewer than two false alerts per scan on average. This balance between high sensitivity and acceptable false-positive rates supports the feasibility of integrating these algorithms into clinical workflows without significantly increasing radiologist workload.

When compared with expert radiologist performance on a subset of cases, AI sensitivity was slightly higher across all tasks, although these differences were not statistically significant. This demonstrates that AI performance was comparable to that of experienced radiologists rather than superior, reinforcing the role of AI as a supportive tool rather than a replacement for human expertise. The absence of statistically significant differences suggests that AI-assisted interpretation could provide consistent performance across readers and institutions.

Subgroup analysis by lesion size revealed that detection sensitivity increased with lesion size for both pulmonary and liver lesions. Smaller lesions less than 6 mm were detected with lower sensitivity compared to larger lesions, which is consistent with known limitations of CT imaging and both human and algorithmic interpretation. However, sensitivity improved markedly for lesions larger than 10 mm, indicating strong performance for clinically significant findings that typically warrant further evaluation or intervention.

Taken together, these results demonstrate that three-dimensional AI models can reliably analyze volumetric CT data and achieve performance comparable to expert radiologists. The findings support the use of AI as an adjunctive tool to enhance detection consistency, reduce missed findings, and improve efficiency in CT interpretation, particularly in high-volume or resource-limited clinical settings.

Smaller lesions (<6 mm) were detected with slightly lower sensitivity, consistent with known limitations.

DISCUSSION

This study developed and validated a suite of automated AI algorithms for detecting liver lesions, pulmonary nodules, renal masses, and mesenteric injuries on CT imaging, demonstrating strong performance across multiple clinically relevant tasks. The high AUROC values and sensitivity profiles reflect robust detection capabilities that were statistically comparable to those of experienced radiologists [9].

AI-assisted detection showed particular strength in identifying both liver and pulmonary lesions, aligning with earlier reports highlighting the effectiveness of deep learning for thoraco-abdominal anomaly detection [10]. Renal mass and mesenteric injury detection also achieved strong performance, though with marginally higher false positives per scan, likely reflecting the greater heterogeneity of appearance and lower prevalence within the dataset [11].

Sensitivity decreased for smaller pulmonary lesions (<6 mm), consistent with previous research indicating that ultra-small nodules challenge both automated and human observers due to subtle contrast and morphological features [12]. Nonetheless, the models maintained clinically acceptable detection rates, suggesting practical utility for prioritization and triage.

The multicenter dataset and rigorous reference standard strengthened the study's validity, addressing common limitations such as overfitting to a single institution's imaging protocols [13]. The comparative analysis further supports the potential for AI deployment as an adjunct tool rather than a replacement for expert interpretation.

A key novelty of this work lies in the simultaneous development of multiple models optimized for varied anatomic targets within CT imaging, enabling comprehensive automated screening across organs of high clinical concern. This contrasts with prior studies that focus on isolated tasks, and underscores the feasibility of integrated AI assistance in diverse diagnostic contexts [14].

Limitations include the retrospective nature of the dataset and potential selection bias in annotated cases. Prospective validation in real-time clinical workflows is warranted to confirm generalizability and impact on patient outcomes.

CONCLUSION

The developed AI algorithms demonstrated rapid, sensitive, and clinically relevant detection of liver lesions, pulmonary nodules, renal masses, and mesenteric injuries on CT imaging, offering a novel automated solution to support diagnostic workflows and enhance radiologic efficiency without compromising accuracy.

ACKNOWLEDGEMENTS

The authors thank the radiology departments, annotators, and data engineers at participating centers for their contributions to this project.

ETHICS STATEMENT

The study was conducted with approval from Institutional Ethical Review Committees under reference numbers ERC/AI-RAD/2018/127.

INFORMED CONSENT

The requirement for informed consent was waived due to the use of de-identified retrospective imaging data.

COMPETING INTERESTS

The authors declare no competing interests.

FINANCIAL DISCLOSURE

This study did not receive external funding.

REFERENCES

1. Kalender WA. (2011). Computed Tomography: Fundamentals, System Technology, Image Quality, Applications. 3rd ed. Wiley-VCH.
2. Brady AP. (2017). Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging. 8(1):171-182.
3. Litjens G, Kooi T, Bejnordi BE, et al. (2017). A survey on deep learning in medical image analysis. Med Image Anal. 42:60-88.
4. Ardila D, Kiraly AP, Bharadwaj S, et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 25(6):954-961.

5. Tajbakhsh N, Shin JY, Gurudu SR, et al. (2016). Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging. 35(5):1299-1312.
6. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med. 15(11):e1002683.
7. Lakhman Y, Yang J, Karcich J, et al. (2020). CT of trauma patients: body imaging and beyond. Radiol Clin North Am. 58(2):405-426.
8. Rajpurkar P, Irvin J, Ball RL, et al. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 15(11):e1002686.
9. Zhu X, Zhang B, Jiang A, et al. (2021). Deep learning algorithms for automated detection of liver lesions in CT: a systematic review. Insights Imaging. 12(1):130.
10. Yasaka K, Akai H, Abe O, et al. (2018). Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology. 286(3):887-896.
11. Nam JG, Park S, Hwang EJ, et al. (2020). Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. Radiology. 295(1):201-211.
12. Oakden-Rayner L. (2020). Exploring large-scale public medical image datasets. Acad Radiol. 27(1):106-112.
13. Lakhani P, Sundaram B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. 284(2):574-582.
14. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 17(1):195.

CITATION OF THIS ARTICLE

Nida N, Ali A, Muhammad Usman Farooq B, Roshna Q, Omna Y, Muhammad Mubeen Ul H . Development of automated AI algorithms for detecting liver and pulmonary lesions, renal masses, or mesenteric injuries on CT imaging. Bull. Env. Pharmacol. Life Sci., Vol 15 [3] February 2026. 50-54