JHWCR JOURNAL OF HEALTH, WELLNESS AND COMMUNITY RESEARCH 3007-0570

https://doi.org/10.61919/p8f43057

Original Article

Radiologist-AI Double Reading for Incidental Oncology Findings: Detecting Overlooked Metastatic Clues in Routine Chest and Abdomen CT

Usman Akram[1], Ahmad Waqar[2], Ayesha Waheed[3], Amna Abbas[4], Fareeha F Khan[5], Shaikh Khalid Muhammad[6]

[1] MPhil Student, University of Lahore, Lahore, Pakistan
[2] Student, Beaconhouse Potohar Campus, Islamabad, Pakistan
[3] Radiology Training Completed, Hayatabad Medical Complex, Peshawar, Pakistan
[4] MBBS, CMH Institute of Medical Sciences; House Job Completed, General Hospital, Lahore, Pakistan
[5] Senior Medical Officer, Intensive Care Unit, The Kidney Center Hospital, Pakistan ORCID: 0009-0004-3391-1326
[6] MBBS, FCPS Medicine, Professor of Medicine, CMC Teaching Hospital, Pakistan

***Corresponding author: Usman Akram, usmanakram69900@gmail.com**

"Cite this Article" Received: 02 April 2026; **Accepted**: 05 May 2026; **Published**: 19 May 2026

Author Contributions: Concept: UA; Design: AW; Data Collection: AYW and AA; Analysis: FFK; Drafting: SKM. **Ethical Approval** was obtained by the Respective Institution. **Informed Consent**: Written informed consent was obtained from all participants; **Conflict of Interest**: The authors declare no conflict of interest. **Funding**: No external funding; **Data Availability**: Available from the corresponding author on reasonable request; **Acknowledgments**: N/A.

ABSTRACT

Background: Routine contrast-enhanced chest and abdomen CT examinations performed for non-oncologic indications may contain subtle incidental oncologic findings that are missed or insufficiently emphasized during focused reporting. Objective: To evaluate whether a radiologist–AI double-reading quality assurance pathway can identify overlooked incidental oncologic findings on routine non-oncology CT and determine their clinical significance. Methods: This retrospective clinical correlation study included 2,486 eligible contrast-enhanced chest, abdomen, or combined chest–abdomen CT examinations from 3,214 retrieved studies at a tertiary-care hospital in Islamabad, Pakistan. De-identified scans were screened using a sensitivity-oriented AI triage system for suspicious skeletal, nodal, and visceral abnormalities. AI-flagged examinations underwent independent radiologist re-review, consensus adjudication, comparison with initial reports, and clinical validation through histopathology, follow-up imaging progression, or strong clinic-oncologic correlation. Results: AI flagged 412 of 2,486 examinations (16.6%). Radiologist re-review confirmed suspicious incidental oncologic findings in 187 flagged cases (45.4%). Of these, 56 cases (29.9%) were discordant with the initial report, representing 2.3% of the total cohort. Skeletal lesions were the most frequent overlooked abnormality (24/56; 42.9%), followed by nodal disease (17/56; 30.4%) and visceral lesions (15/56; 26.8%). Clinical validation confirmed malignancy-related disease in 38 of 56 discordant cases (67.9%). Conclusion: A radiologist–AI double-reading pathway selectively identified clinically meaningful overlooked oncologic findings while limiting secondary review to a minority of routine CT examinations. Keywords: Artificial intelligence; computed tomography; incidental findings; metastasis; radiology quality assurance; double reading; oncology.

INTRODUCTION

Computed tomography of the chest and abdomen is one of the most frequently used cross-sectional imaging tools in acute, inpatient, and general medical practice, where examinations are commonly requested to answer focused non-oncologic questions such as infection, trauma, abdominal pain, vascular pathology, or postoperative complications (1). Despite this focused clinical framing, these examinations cover large anatomical regions that may contain clinically important incidental oncologic abnormalities, including subtle lytic bone lesions, borderline or morphologically suspicious lymph nodes, small hepatic deposits, adrenal lesions, pulmonary nodules, or early peritoneal disease (2). Such findings may represent the first radiologic evidence of an undiagnosed malignancy or early metastatic

spread, and delayed recognition can affect staging, referral pathways, treatment eligibility, and patient outcomes. Diagnostic discrepancy in radiology is a recognized problem, and double reading has been shown to add value across imaging settings, although its routine use is constrained by workload, staffing limitations, and turnaround-time pressures (3).

The risk of overlooking incidental metastatic disease is especially relevant in routine non-oncology CT because the radiologist's attention is often directed toward the immediate clinical question rather than a systematic oncologic survey. Human factors such as fatigue, interruptions, satisfaction-of-search, perceptual bias, and high reporting volume can increase the likelihood that subtle lesions outside the primary diagnostic target are not reported or are described without sufficient clinical urgency (4,5). Skeletal metastases are particularly vulnerable to under-recognition because bones are included throughout the scanned field of view, metastatic appearances may be lytic, sclerotic, mixed, or subtle, and benign mimics such as degenerative change, trauma, or insufficiency fractures can reduce diagnostic confidence. Previous evidence suggests that deep learning support can improve radiologists' detection of bone metastases on CT, supporting the concept that computational tools may be useful as an adjunctive second reader for visually subtle but clinically significant lesions (6). Similar challenges apply to nodal and visceral metastatic disease, where small-volume abnormalities may be difficult to distinguish from benign or indeterminate findings, particularly when CT protocols were not optimized for oncologic staging.

Artificial intelligence offers a potential strategy to bridge the gap between universal double reading, which is often operationally unrealistic, and purely single-reader reporting, which may leave some clinically meaningful incidental findings undetected. In this context, AI is best conceptualized not as a replacement for radiologists but as a sensitivity-oriented triage layer that identifies examinations requiring targeted human re-review. This approach is particularly attractive for quality assurance because it concentrates radiologist effort on a selected subset of routine CT examinations rather than requiring systematic second reading of all studies. Recent developments in AI-assisted imaging interpretation and post-report review have highlighted the feasibility of using computational tools to support discrepancy detection, but clinical implementation requires careful validation, radiologist adjudication, and integration with follow-up systems rather than reliance on AI outputs alone (7,8).

A further challenge is that detection alone does not guarantee clinical benefit. Incidental oncologic findings must be clearly communicated, assigned appropriate urgency, and linked to a reliable follow-up pathway. Prior work on actionable incidental findings has shown that missed follow-up can occur even when abnormalities are identified, emphasizing the need for structured reporting language, clinician notification, and longitudinal tracking systems to close the loop between radiologic detection and patient management (9,10). This issue is particularly important in high-volume tertiary-care settings in low- and middle-income countries, where patients may move between institutions, outpatient continuity may be fragmented, and delays in oncology referral can be amplified by access, cost, and system-level barriers. In such environments, a radiologist–AI quality assurance pathway may be clinically useful only if it combines accurate case triage with expert confirmation and follow-up governance.

The population of interest in the present study comprises adult patients undergoing routine contrast-enhanced chest, abdominal, or combined chest–abdomen CT for non-oncologic indications in a tertiary-care hospital setting. The intervention is a radiologist–AI double-reading quality assurance pathway in which AI flags potentially suspicious skeletal, nodal, or visceral abnormalities for targeted radiologist re-evaluation. The comparator is the original routine radiology report generated during standard clinical care, and the primary outcome is the prevalence and clinical significance of discordant incidental oncologic findings validated through pathology, follow-up imaging progression, or strong clinic-oncologic correlation. Accordingly, this study was designed to determine whether AI-assisted selective secondary review can identify overlooked metastatic clues in routine non-oncology CT examinations

and whether these findings represent clinically meaningful disease rather than incidental, non-actionable abnormalities.

MATERIALS AND METHODS

This retrospective re-evaluation and clinical correlation study was conducted at a tertiary-care hospital in Islamabad, Pakistan, to assess whether a radiologist–AI double-reading quality assurance pathway could identify incidental oncologic findings that were not meaningfully reported in routine contrast-enhanced CT examinations performed for non-oncology indications. The study population consisted of adult patients who underwent contrast-enhanced CT of the chest, abdomen, or combined chest–abdomen during routine clinical care. The design was selected to reflect real-world radiology workflow, in which examinations are commonly requested for focused clinical questions such as suspected infection, trauma, acute abdominal pain, vascular pathology, postoperative complications, or other non-malignancy-related presentations, while still covering anatomical regions where metastatic disease may be visible.

Eligible examinations were contrast-enhanced CT studies of the chest, abdomen, or combined chest–abdomen performed for non-oncologic primary indications. Eligibility was determined from the imaging request, clinical indication, and radiology report context. Examinations were excluded if they were performed for known cancer staging, treatment-response assessment, oncologic surveillance, or follow-up of a previously documented malignant lesion. Studies were also excluded if image quality was inadequate for reliable review because of severe motion artifact, incomplete anatomical coverage, missing essential image series, or technical limitations that prevented meaningful assessment of skeletal, nodal, or visceral structures. When more than one eligible CT examination was available for the same patient during the study period, only the earliest eligible examination was included to avoid over-representation of patients with repeated imaging.

CT examinations were retrieved from the institutional picture archiving and communication system, and relevant clinical information was extracted from available medical records. Data collected for each case included patient age, sex, CT region, contrast-enhancement status, scan indication, original radiology report, AI-flag category, radiologist re-review classification, discrepancy status, follow-up evidence, and final clinical validation outcome. Before AI processing and research review, all imaging and clinical data were de-identified using a study-code system. Direct patient identifiers were removed from the research dataset, and the linkage file required for clinical correlation was stored separately with restricted access.

The target findings were predefined as incidental oncologic abnormalities suspicious for metastatic or malignancy-related disease within the scanned field of view. Skeletal findings included lytic, sclerotic, mixed, permeative, destructive, or cortical-breaching lesions involving structures such as the vertebrae, ribs, sternum, pelvis, or proximal femora. Nodal findings included lymph nodes with suspicious size, morphology, necrosis, clustering, or anatomical distribution suggestive of malignant involvement. Visceral findings included suspicious hepatic lesions, adrenal lesions, pulmonary nodules in a metastatic pattern, peritoneal or omental nodularity, or other solid-organ abnormalities with imaging characteristics concerning for malignancy. Indeterminate benign-appearing lesions were not classified as clinically meaningful oncologic findings unless subsequent evidence supported malignant behavior.

All de-identified CT examinations were screened retrospectively using an AI system configured as a sensitivity-oriented triage tool for suspicious skeletal, nodal, and visceral abnormalities. The AI output generated case-level suspicion flags and localization prompts to guide human review. The AI assessment was performed without access to subsequent clinical outcomes, pathology results, or follow-up imaging findings. The purpose of AI screening was not to establish a final diagnosis but to identify examinations requiring targeted radiologist re-evaluation within a quality assurance workflow.

AI-flagged examinations were independently re-reviewed by two radiologists using the complete CT image set and AI localization prompts. During the first stage of review, radiologists assessed whether the flagged abnormality represented a suspicious incidental oncologic finding and assigned a confidence score using a four-point scale: 1 indicating unlikely malignant, 2 indicating indeterminate, 3 indicating suspicious, and 4 indicating highly suspicious. Findings with a confidence score of ≥3 were considered radiologist-confirmed suspicious abnormalities. Reader disagreement was resolved by consensus review with a senior radiologist, and the consensus decision was used as the final radiologic classification.

After radiologist adjudication, confirmed suspicious findings were compared with the corresponding routine radiology reports to determine discrepancy status. A finding was classified as concordant when it had been clearly described or appropriately implied in the initial report. It was classified as partially concordant when the abnormality had been mentioned but was underestimated, incompletely characterized, or not linked to an appropriate follow-up recommendation. It was classified as discordant when the abnormality was absent from the report or was not framed as actionable despite imaging features suspicious for malignancy. Discordant findings represented potentially overlooked incidental oncologic findings.

Clinical validation was performed for discordant findings using a hierarchical framework based on available follow-up evidence. A discordant finding was considered validated as malignancy-related when there was histopathologic or cytologic confirmation, unequivocal progression on follow-up imaging consistent with metastatic disease, or strong clinic-oncologic correlation documented in the medical record. Imaging progression was defined by interval growth of the lesion, development of additional lesions in a metastatic distribution, or evolution of imaging features consistent with malignant disease. Cases without malignant confirmation during follow-up were classified as nonvalidated, stable, or insufficiently supported depending on the available evidence.

The primary outcome was the prevalence of potentially overlooked incidental oncologic findings among routine non-oncology CT examinations, calculated as the number of discordant radiologist-confirmed suspicious cases divided by the total number of eligible CT examinations. The secondary outcome was the clinical validation rate, calculated as the proportion of discordant cases subsequently confirmed as metastatic or malignancy-related disease. Additional outcomes included the anatomical distribution of overlooked findings, AI flag yield, proportion of AI-flagged cases confirmed by radiologist review, time from index CT to cancer confirmation, and inter-reader agreement for suspicious versus non-suspicious classification.

Bias was addressed through retrospective de-identification, predefined eligibility criteria, standardized target-lesion categories, independent dual radiologist review, consensus adjudication, and separation of AI screening from clinical outcome validation. To reduce incorporation bias, AI screening was performed without access to follow-up outcomes. To reduce reviewer variability, radiologists used a standardized confidence scale and discrepancy classification system. Potential confounding related to scan type, anatomical region, clinical indication, and workflow complexity was assessed through exploratory analysis of factors associated with discordant findings.

The sample size was determined by the number of eligible CT examinations available within the institutional retrospective sampling frame after application of inclusion and exclusion criteria. Because the study was designed as a quality assurance re-evaluation of routine imaging practice, all eligible examinations within the study frame were included to maximize precision of prevalence estimates and improve representation of routine CT workflow.

Quantitative data were summarized using descriptive statistics. Continuous variables were reported as medians with interquartile ranges when non-normally distributed, and categorical variables were reported as frequencies and percentages. The prevalence of AI-flagged examinations, radiologist-confirmed suspicious findings, discordant findings, and clinically validated malignancy-related findings

was calculated using the eligible CT cohort as the denominator where appropriate. Category-specific yields were calculated for skeletal, nodal, and visceral abnormalities. Inter-reader agreement for binary suspicious versus non-suspicious classification was assessed using Cohen's kappa. Comparisons between groups were performed using chi-square or Fisher's exact tests for categorical variables and Mann–Whitney U tests for non-normally distributed continuous variables. Exploratory logistic regression was used to evaluate factors associated with discordant findings, including scan region, lesion category, clinical indication, and workflow-related variables where available. Statistical significance was assessed using a two-sided threshold of $p<0.05$.

Ethical approval was obtained from the institutional ethics committee, and the requirement for informed consent was waived because the study used retrospectively collected, de-identified imaging and clinical data. Patient confidentiality was maintained throughout data extraction, AI processing, radiologist review, and outcome correlation. Results were analyzed and reported in aggregate, and no directly identifiable patient information was included. Data integrity was supported through coded case identifiers, restricted access to the linkage file, standardized extraction fields, independent radiologist review, consensus adjudication, and preservation of the audit trail linking AI flags, radiologist classifications, discrepancy status, and validation outcomes.

RESULTS

During the study period, 3,214 contrast-enhanced CT examinations of the chest, abdomen, or combined chest–abdomen were retrieved from the institutional PACS. After applying eligibility criteria, 2,486 CT examinations from 2,486 unique patients were included in the final analysis. The median patient age was 52 years, with an interquartile range of 41–63 years. Males accounted for 1,392 patients, representing 56.0% of the cohort, while females accounted for 1,094 patients, representing 44.0%. Combined chest–abdomen CT was the most frequent scan type, comprising 1,030 examinations, or 41.4% of the final cohort, followed by chest CT only in 812 cases, or 32.7%, and abdomen CT only in 644 cases, or 25.9%. The most frequent clinical indication was suspected infection or sepsis workup, which accounted for 706 examinations, or 28.4%, followed by acute abdominal pain in 599 cases, or 24.1%, trauma assessment in 390 cases, or 15.7%, vascular evaluation in 338 cases, or 13.6%, and miscellaneous non-oncologic indications in 453 cases, or 18.2%. Overall, the radiologist–AI pathway concentrated secondary review on 16.6% of routine non-oncology CT examinations while identifying 56 potentially overlooked incidental oncologic findings, of which 38 were clinically validated as malignancy-related. This corresponds to approximately 1 overlooked suspicious oncologic finding per 44 routine CT examinations and 1 clinically validated malignancy-related finding per 65 routine CT examinations in the final cohort. The diagnostic yield was highest when AI triage was followed by expert radiologist adjudication and clinical correlation, demonstrating that the workflow functioned as a selective quality assurance pathway rather than a standalone diagnostic system.

Table 1. Baseline Characteristics of the Final CT Cohort

Characteristic	Value	Percentage / Summary
Total CT examinations retrieved	3,214	100.0%
Final eligible CT examinations	2,486	77.3% of retrieved examinations
Excluded examinations	728	22.7% of retrieved examinations
Unique patients included	2,486	100.0% of final cohort
Median age, years	52	IQR: 41–63
Male sex	1,392	56.0%
Female sex	1,094	44.0%
Chest CT only	812	32.7%
Abdomen CT only	644	25.9%
Combined chest–abdomen CT	1,030	41.4%
Infection/sepsis workup	706	28.4%
Acute abdominal pain	599	24.1%
Trauma assessment	390	15.7%
Vascular evaluation	338	13.6%
Miscellaneous non-oncologic indications	453	18.2%

AI screening flagged 412 of the 2,486 included CT examinations, corresponding to 16.6% of the final cohort. Among flagged examinations, skeletal abnormalities were the most frequent AI flag category, identified in 196 cases, representing 7.9% of all eligible CT examinations and 47.6% of AI-flagged cases. Suspected nodal disease was flagged in 143 cases, corresponding to 5.8% of the total cohort and 34.7% of flagged cases. Suspected visceral metastases were flagged in 101 cases, corresponding to 4.1% of the total cohort and 24.5% of flagged cases. Because some examinations contained more than one category of AI flag, the sum of category-specific flags exceeded the total number of AI-flagged examinations.

Table 2. AI Screening Output by Abnormality Category

AI Screening Output	Number of Cases	% of Total Cohort, n=2,486	% of AI-Flagged Cases, n=412
Any AI-flagged abnormality	412	16.6%	100.0%
Skeletal lesion flag	196	7.9%	47.6%
Suspected nodal disease flag	143	5.8%	34.7%
Suspected visceral metastasis flag	101	4.1%	24.5%

Following independent radiologist re-evaluation and consensus adjudication, 187 of the 412 AI-flagged examinations were confirmed to contain at least one suspicious or highly suspicious incidental oncologic finding, giving a radiologist-confirmed yield of 45.4% among AI-flagged cases and 7.5% among all eligible CT examinations. When radiologist-confirmed suspicious cases were compared with the initial radiology reports, 82 cases, or 43.9%, were concordant; 49 cases, or 26.2%, were partially concordant; and 56 cases, or 29.9%, were discordant. The 56 discordant cases represented 2.3% of the total eligible CT cohort and were considered potentially overlooked incidental oncologic findings. Inter-reader agreement for binary suspicious versus non-suspicious classification was substantial, with Cohen's κ of 0.74 before consensus and 0.81 after consensus discussion.

Table 3. Radiologist Re-Evaluation, Discrepancy Classification, and Agreement

Outcome	Number of Cases	Denominator	Percentage / Statistic
AI-flagged examinations	412	2,486	16.6%
Radiologist-confirmed suspicious findings	187	412	45.4%
Radiologist-confirmed suspicious findings	187	2,486	7.5%
Concordant findings	82	187	43.9%
Partially concordant findings	49	187	26.2%
Discordant findings	56	187	29.9%
Discordant findings	56	2,486	2.3%
Cohen's κ before consensus	—	—	0.74
Cohen's κ after consensus	—	—	0.81

Among the 56 discordant cases, skeletal lesions were the most common overlooked abnormality, accounting for 24 cases, or 42.9% of all discordant findings. Lymph node abnormalities were the second most frequent category, observed in 17 cases, or 30.4%. Visceral abnormalities accounted for the remaining 15 cases, or 26.8%, including 9 liver lesions, representing 16.1%, and 6 other visceral sites, representing 10.7%. The predominance of skeletal findings indicates that osseous structures visible on routine CT, particularly regions such as the thoracic spine, ribs, and pelvis, represented an important source of missed or underemphasized oncologic abnormalities.

Table 4. Anatomical Distribution and Clinical Validation of Discordant Findings

Discordant Finding Category	Discordant Cases, n=56	% of Discordant Cases
Skeletal lesions	24	42.9%
Lymph nodes	17	30.4%
Liver lesions	9	16.1%
Other visceral sites	6	10.7%
All visceral lesions	15	26.8%
Total discordant findings	56	100.0%

Clinical validation confirmed malignancy-related disease in 38 of the 56 discordant cases, corresponding to a validation rate of 67.9%. Histopathology provided confirmation in 14 cases, representing 36.8% of validated cases and 25.0% of all discordant cases. Imaging progression confirmed malignant behavior in 19 cases, representing 50.0% of validated cases and 33.9% of all discordant cases. Strong clinical-oncologic correlation accounted for 5 validated cases, representing 13.2% of validated cases and 8.9% of all discordant cases. The remaining 18 discordant cases, or 32.1%, were not clinically validated as

malignancy-related within the available follow-up framework. Among clinically validated cases, the median time from index CT to cancer confirmation was 4.2 months, with an interquartile range of 2.1–7.8 months.

Table 5. Clinical Validation Pathway and Diagnostic Yield of the Radiologist–AI QA Workflow

Outcome / Validation Pathway	Number of Cases	Denominator	Percentage / Summary
Total eligible CT examinations	2,486	2,486	100.0%
AI-flagged examinations	412	2,486	16.6%
Radiologist-confirmed suspicious findings	187	412	45.4%
Discordant potentially overlooked findings	56	2,486	2.3%
Clinically validated malignancy-related findings	38	56	67.9%
Clinically validated malignancy-related findings	38	2,486	1.5%
Pathology-confirmed cases	14	38	36.8%
Imaging progression-confirmed cases	19	38	50.0%
Strong clinical-correlation cases	5	38	13.2%
Nonvalidated discordant cases	18	56	32.1%
Median time from index CT to cancer confirmation	—	38	4.2 months, IQR 2.1–7.8
Positive predictive value for radiologist-confirmed suspicious findings	187	412	45.4%
Positive predictive value for clinically validated malignancy-related findings	63	412	15.3%

Exploratory analysis identified combined chest–abdomen CT examinations, skeletal involvement outside the primary clinical focus, and scans performed during high-volume emergency shifts as factors associated with higher odds of discordance. Combined chest–abdomen CT was associated with an odds ratio of 1.9 for overlooked findings, skeletal involvement outside the main clinical focus was associated with an odds ratio of 2.3, and high-volume emergency shift imaging was associated with an odds ratio of 1.7. These associations suggest that missed or underemphasized oncologic findings clustered in examinations with wider anatomical coverage, peripheral skeletal abnormalities, and workflow conditions likely to increase interpretive complexity.

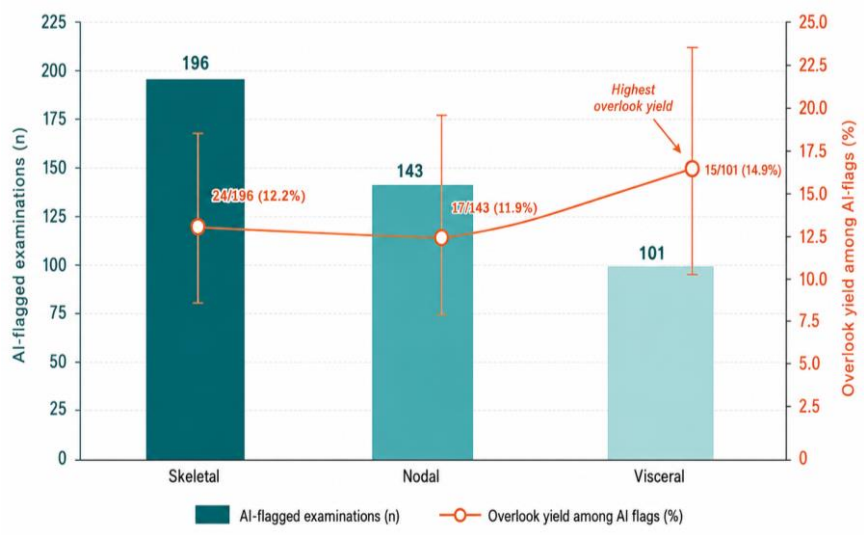

Figure 1. Category-Specific Overlook Yield Among AI-Flagged Routine CT Findings

This figure shows the relationship between AI-flag volume and the proportion of overlooked incidental oncologic findings across skeletal, nodal, and visceral lesion categories. Teal bars represent the number of AI-flagged examinations in each category, with skeletal findings producing the highest flag volume (196 cases), followed by nodal abnormalities (143 cases) and visceral lesions (101 cases). The coral line shows the overlook yield among AI-flagged cases, calculated as discordant findings divided by AI-flagged examinations. Visceral lesions demonstrated the highest overlook yield at 14.9% (15/101), followed by skeletal lesions at 12.2% (24/196) and nodal disease at 11.9% (17/143). Error bars represent

95% confidence intervals, highlighting category-level uncertainty while showing that clinically relevant overlooked findings occurred across all three lesion groups.

DISCUSSION

This retrospective clinical correlation study demonstrates that routine contrast-enhanced CT examinations performed for non-oncologic indications can contain clinically meaningful incidental oncologic findings that may be missed, underemphasized, or not translated into actionable recommendations during standard reporting. In a final cohort of 2,486 routine chest, abdomen, or combined chest–abdomen CT examinations, AI triage flagged 412 studies, representing 16.6% of the cohort, and radiologist re-evaluation confirmed suspicious incidental oncologic findings in 187 flagged examinations. Among these, 56 cases were discordant with the initial report, yielding a potentially overlooked oncologic finding rate of 2.3% across the total cohort. Although this percentage appears modest, its clinical relevance is substantial because 38 of the 56 discordant findings, or 67.9%, were subsequently validated as malignancy-related through histopathology, imaging progression, or strong clinical-oncologic correlation. These findings support the concept that AI-assisted selective double reading may help identify a concentrated subset of routine CT examinations in which expert re-review has meaningful diagnostic yield.

The observed rate of discordant potentially overlooked findings is consistent with the broader understanding that radiology discrepancies are shaped not only by individual interpretation but also by case complexity, perceptual limitations, workload, and workflow pressures. Prior evidence on double reading in diagnostic radiology has shown wide variation in discrepancy rates across settings, reflecting differences in modality, case mix, reader expertise, and review design (11). The present findings extend that principle to the specific setting of incidental oncology detection in routine CT, where the clinical indication often directs attention toward acute or non-malignant pathology while subtle metastatic clues remain outside the primary diagnostic target. This is particularly important because non-oncology CT examinations frequently cover the skeleton, lymphatic stations, liver, adrenal glands, lungs, and peritoneum, all of which may harbor early metastatic disease even when cancer is not suspected at the time of imaging (12).

Skeletal lesions formed the largest proportion of discordant findings, accounting for 24 of 56 overlooked abnormalities, or 42.9%. This distribution is clinically plausible because osseous structures are widely included in chest and abdominal CT fields of view but may receive less systematic attention when the scan is requested for infection, trauma, vascular assessment, or abdominal pain. Subtle lytic lesions, cortical breaches, permeative marrow changes, or mixed bone lesions may be difficult to distinguish from degenerative disease, fractures, or benign bone islands, particularly in high-volume reporting environments. Previous work has shown that deep learning–based assistance can improve radiologists' performance in detecting bone metastases on CT, suggesting that AI may be particularly useful where abnormalities are visually learnable but vulnerable to satisfaction-of-search and peripheral under-recognition (13,14). In this study, the predominance of skeletal misses indicates that routine reporting checklists should reinforce deliberate review of the thoracic spine, ribs, sternum, pelvis, and proximal femora, especially in combined chest–abdomen CT examinations.

Nodal disease represented the second largest category of discordant findings, with 17 cases, or 30.4%, among the 56 overlooked abnormalities. Unlike many skeletal lesions, nodal metastases are often interpretively complex because size alone is an imperfect criterion for malignancy. Small or borderline lymph nodes may still be suspicious when they show abnormal morphology, clustering, necrosis, asymmetric distribution, or location within a known metastatic drainage pathway. In routine non-oncology CT, these features may be overlooked or interpreted conservatively because the pre-test suspicion of malignancy is low. This finding highlights the importance of contextual radiologist judgment after AI triage: AI may identify candidate nodal abnormalities, but the final determination of

clinical relevance requires integration of nodal morphology, distribution, patient age, clinical presentation, and follow-up evidence (15,16).

Visceral lesions accounted for 15 of 56 discordant cases, including 9 liver lesions and 6 other visceral sites. Although visceral findings were less frequent than skeletal and nodal abnormalities, they remain clinically important because small hepatic deposits, adrenal metastases, lung nodules in a metastatic pattern, and early peritoneal disease can influence staging and referral decisions. Routine portal venous phase CT is not always optimized for characterization of small hepatic or adrenal lesions, and subtle peripheral lesions may be vulnerable to perceptual under-detection. This reinforces the need for structured reporting language when such lesions are identified, particularly when they are not definitively benign (17). Ambiguous phrases such as “indeterminate lesion, correlate clinically” are less useful than explicit statements describing the suspicious feature, likely differential diagnosis, and recommended follow-up modality or interval.

A key strength of the radiologist–AI pathway was its ability to concentrate secondary review on a manageable subset of examinations. Universal double reading of all routine CT examinations is difficult to sustain in busy tertiary-care radiology departments because it increases workload and may delay report turnaround. In contrast, AI triage selected 412 of 2,486 examinations for re-review, meaning that only 16.6% of routine scans required additional radiologist attention. Within this selected subset, 45.4% contained radiologist-confirmed suspicious findings, and 15.3% yielded clinically validated malignancy-related abnormalities. This supports the operational value of AI as a triage and quality assurance tool rather than a standalone diagnostic authority (18). The workflow is strongest when AI flags are interpreted by radiologists, adjudicated through consensus when needed, and linked to clinical validation rather than treated as automatic diagnoses.

The clinical validation rate is particularly important because it distinguishes meaningful missed disease from incidental image noise. In this study, 38 of 56 discordant cases were validated as malignancy-related, including 14 pathology-confirmed cases, 19 cases confirmed by imaging progression, and 5 cases supported by strong clinic-oncologic correlation. The median time from index CT to cancer confirmation was 4.2 months, with an interquartile range of 2.1–7.8 months, suggesting that a proportion of these abnormalities represented earlier visible signs of disease. However, because the study design was retrospective, these findings should be interpreted as evidence of missed diagnostic opportunity rather than proof that AI-assisted review improves survival, treatment eligibility, or long-term outcomes. Prospective implementation would be required to determine whether earlier detection through this pathway reduces time to oncology referral, changes stage at diagnosis, or improves patient-centered outcomes.

The findings also emphasize that incidental oncology detection is not only a perceptual problem but also a systems problem. Even when suspicious findings are identified, clinical benefit depends on whether they are communicated clearly, assigned appropriate urgency, and followed to completion. Prior work on actionable incidental findings has shown that failures in communication and follow-up can prevent radiologic findings from translating into patient benefit (19,20). Therefore, an AI-assisted second-look pathway should not end with lesion detection. It should include structured report language, clear follow-up recommendations, responsible clinician notification, and a tracking mechanism for unresolved suspicious findings. In resource-limited or fragmented care settings, a lightweight registry or radiology-led follow-up list may be sufficient to reduce loss to follow-up, provided that responsibility for action is clearly assigned.

The exploratory predictor analysis suggested that combined chest–abdomen CT examinations, skeletal involvement outside the primary clinical focus, and high-volume emergency shift imaging were associated with higher odds of discordance. These findings are plausible because wider anatomical coverage increases the number of structures requiring review, peripheral skeletal findings may fall outside the radiologist’s immediate attentional target, and high-volume shifts can intensify cognitive

load. Although these associations require cautious interpretation because the regression details were exploratory, they provide a useful practical signal: quality assurance workflows may be most valuable for anatomically extensive examinations and reporting contexts where workload pressure is highest. A targeted approach could prioritize combined chest–abdomen CT, emergency CT, and examinations with AI-flagged skeletal or nodal abnormalities for expedited secondary review.

This study has several limitations. First, the retrospective design limits causal inference and does not allow direct measurement of whether AI-assisted review improved clinical outcomes. Second, the study was conducted in a single tertiary-care setting, so the prevalence of overlooked findings and the yield of AI triage may differ across institutions with different patient populations, scanner protocols, radiologist staffing, cancer epidemiology, and follow-up systems. Third, clinical validation depended on available pathology, imaging follow-up, and medical record documentation; therefore, patients with incomplete follow-up may have led to underestimation or overestimation of the true malignancy-related yield. Fourth, the AI system was used at a sensitivity-oriented threshold, which is appropriate for quality assurance but may increase false-positive flags and radiologist workload. Finally, the study primarily evaluated a combined radiologist–AI pathway rather than independent AI diagnostic accuracy, so the results should be interpreted as evidence for workflow-level quality assurance rather than standalone AI performance.

Despite these limitations, the study provides a practical framework for improving incidental oncologic detection in routine CT practice. The data suggest that a radiologist–AI double-reading pathway can identify a measurable number of overlooked abnormalities while avoiding the burden of universal double reading. The greatest potential value lies in integrating AI triage with expert radiologist adjudication and closed-loop follow-up. For tertiary-care hospitals with high imaging volumes, this approach may provide a realistic quality assurance strategy: AI identifies examinations requiring attention, radiologists determine clinical relevance, and structured follow-up systems ensure that suspicious findings are not lost after reporting. Future prospective studies should evaluate time to diagnosis, follow-up completion, oncology referral intervals, workload impact, false-positive burden, cost-effectiveness, and patient outcomes to determine whether this workflow produces measurable clinical benefit beyond discrepancy detection.

CONCLUSION

This retrospective clinical correlation study shows that routine contrast-enhanced chest and abdomen CT examinations performed for non-oncologic indications can contain clinically meaningful incidental oncologic findings that may be missed or insufficiently emphasized during standard reporting. In a cohort of 2,486 eligible CT examinations, the radiologist–AI quality assurance pathway limited secondary review to 412 AI-flagged studies, representing 16.6% of the cohort, while identifying 56 discordant potentially overlooked oncologic findings, of which 38 cases, or 67.9%, were clinically validated as malignancy-related through pathology, imaging progression, or strong clinical correlation. Skeletal lesions were the most frequent overlooked abnormality, followed by nodal and visceral findings, indicating that metastatic clues outside the primary clinical focus remain vulnerable to under-recognition in routine CT workflows. These findings support the use of AI as a selective triage layer rather than a standalone diagnostic tool, with radiologists retaining responsibility for final interpretation, discrepancy classification, and clinical framing. The study further highlights that effective incidental oncology detection requires not only improved lesion recognition but also structured reporting and closed-loop follow-up to ensure that suspicious findings lead to timely clinical action. Overall, a radiologist–AI double-reading pathway appears feasible for strengthening quality assurance in high-volume CT practice and may help reduce delayed recognition of metastatic disease when integrated with expert adjudication and reliable follow-up governance.

REFERENCES

1. Geijer H, Geijer M. Added value of double reading in diagnostic radiology, a systematic review. Insights Imaging. 2018;9:287–301. doi:10.1007/s13244-018-0599-0.

2. Lumbreras B, Donat L, Hernández-Aguado I. Incidental findings in imaging diagnostic tests: a systematic review. Br J Radiol. 2010;83(988):276–289. doi:10.1259/bjr/98067945.

3. O'Sullivan JW, Muntinga T, Grigg S, Ioannidis JPA. Prevalence and outcomes of incidental imaging findings: umbrella review. BMJ. 2018;361:k2387. doi:10.1136/bmj.k2387.

4. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: Fleischner Society 2017. Radiology. 2017;284(1):228–243. doi:10.1148/radiol.2017161659.

5. Noguchi T, Kido A, Hashimoto N, et al. Deep learning-based algorithm improved radiologists' performance in bone metastases detection on CT. Eur Radiol. 2022;32:5952–5964. doi:10.1007/s00330-022-08741-3.

6. Zhang Y, Liu X, Chen J, et al. A clinically applicable AI system for detection and diagnosis of bone metastases using CT scans. Nat Commun. 2025;16:9433. doi:10.1038/s41467-025-59433-7.

7. Fu L, Plecha D, Sommer F, et al. System-driven longitudinal follow-up of incidental imaging findings: an initiative with workflow changes. Br J Radiol. 2023;96:20220573. doi:10.1259/bjr.20220573.

8. Fu L, Prevedello LM, Reis M, et al. Implementation and evaluation of dictation macro templates to standardize radiology report recommendations. J Digit Imaging. 2025;38:1–10. doi:10.1007/s10278-024-01357-7.

9. Zarour A, Arpasi P, Song J, et al. The FIND Program: improving follow-up of incidental imaging findings. J Digit Imaging. 2023;36:1268–1276. doi:10.1007/s10278-023-00780-6.

10. Vey B, Gatidis S, Ruppert J, et al. Natural language processing to identify critical findings in radiology reports: a systematic evaluation. J Digit Imaging. 2023;36:360–370. doi:10.1007/s10278-022-00712-w.

11. Polites SF, Habermann EB, Chung AE, et al. A closed-loop system for incidental imaging findings: improving communication and follow-up. J Digit Imaging. 2023;36:1130–1138. doi:10.1007/s10278-022-00773-x.

12. Del Gaizo AJ, Allen BC. Actionable incidental findings on radiologic examinations: existing approaches and future directions. J Radiol Nurs. 2024;43(2):e1–e7. doi:10.1016/j.jradnu.2023.12.003.

13. Evans CS, Arthur R, Kane M, et al. Incidental radiology findings on CT studies in emergency department patients: systematic review and meta-analysis. Ann Emerg Med. 2022;80(3):243–256. doi:10.1016/j.annemergmed.2022.03.027.

14. Kroczek EK, Wieners G, Steffen I, et al. Non-traumatic incidental findings in patients undergoing whole-body CT at initial emergency admission. Emerg Med J. 2017;34(10):643–649. doi:10.1136/emermed-2015-205409.

15. Doughlin GK, Lee SW. Incidental findings in blunt trauma patients: prevalence, follow-up recommendations, and compliance. Emerg Radiol. 2017;24:425–431. doi:10.1007/s10140-017-1479-5.

16. Chapman S, Chadwick C. Expecting the unexpected: incidental findings at a level 1 trauma center. Emerg Radiol. 2023;30:593–602. doi:10.1007/s10140-023-02138-5.

17. Itri JN, Patel SH. Heuristics and cognitive error in medical imaging. AJR Am J Roentgenol. 2018;210(5):1097–1105. doi:10.2214/AJR.17.18907.

18. Bruno MA, Walker EA, Abujudeh HH. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging. 2015;6:171–182. doi:10.1007/s13244-016-0534-1.

19. Pesapane F, Gnocchi G, Quarrella C, et al. Errors in radiology: a standard review. J Clin Med. 2024;13(15):4306. doi:10.3390/jcm13154306.

20. Van Diepen S, van Lieshout EMM, Wijffels MME, van Walsum T. A deep learning-based approach to automated rib fracture detection and classification on CT. Int J Comput Assist Radiol Surg. 2025;20:1–12. doi:10.1007/s11548-025-03390-5.