

INTERNATIONAL JOURNAL OF MEDICAL SCIENCES

ARTIFICIAL INTELLIGENCE-ASSISTED DIAGNOSTIC ACCURACY IN EARLY-STAGE PANCREATIC CANCER DETECTION: A COMPARATIVE ANALYSIS

TAKAHASHI Hiroshi¹

Department of Gastroenterology and Hepatology, Tokyo Medical University, Tokyo, Japan.

Email: h.takahashi@tokyomed.ac.jp

SANTOS Maria Elena dos²

Institute of Oncological Research, University of São Paulo, São Paulo, Brazil.

MOLEFE Thabo³

Division of Radiology and Imaging Sciences, University of Cape Town, Cape Town, South Africa.

ABSTRACT

Background: Pancreatic ductal adenocarcinoma (PDAC) remains one of the most lethal malignancies, with five-year survival rates below 10% largely due to late diagnosis. Artificial intelligence (AI) and machine learning algorithms have shown promise in enhancing early detection capabilities across various imaging modalities.

Objective: To evaluate the diagnostic accuracy of deep learning-based AI algorithms in detecting early-stage pancreatic cancer from computed tomography (CT) and magnetic resonance imaging (MRI) scans compared to conventional radiological assessment.

Methods: This multicenter retrospective study analyzed 2,847 imaging studies (1,623 CT scans and 1,224 MRI scans) from patients across three continents between January 2019 and December 2023. A convolutional neural network (CNN) was trained on 70% of the dataset and validated on the remaining 30%. Performance metrics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUROC) were calculated and compared with interpretations from experienced radiologists.

Results: The AI algorithm demonstrated a sensitivity of 91.3% (95% CI: 88.7-93.5%) and specificity of 94.8% (95% CI: 93.1-96.2%) for detecting early-stage PDAC, compared to 84.2% (95% CI: 81.1-87.0%) and 89.6% (95% CI: 87.5-91.4%) respectively for radiologist interpretation ($p < 0.001$). The AUROC for AI was 0.961 versus 0.893 for conventional reading. AI algorithms identified 127 additional true positive cases that were initially missed on routine radiological review, representing a 15.1% improvement in early detection rates.

Conclusion: AI-assisted diagnostic tools significantly enhance the detection of early-stage pancreatic cancer, offering potential for improved patient outcomes through earlier therapeutic intervention. Integration of these algorithms into clinical workflows warrants further prospective evaluation.

Keywords: Pancreatic cancer, artificial intelligence, machine learning, diagnostic imaging, early detection, convolutional neural network

Introduction

Pancreatic ductal adenocarcinoma represents a formidable challenge in modern oncology, constituting the fourth leading cause of cancer-related mortality in developed nations despite accounting for only 3% of all cancer diagnoses. The stark prognosis associated with PDAC—characterized by a median survival of less than six months for metastatic disease—stems primarily from the tumor's propensity for silent progression during early stages when curative resection remains feasible. Approximately 80-85% of patients present with locally advanced or metastatic disease at diagnosis, rendering them ineligible for potentially curative surgical intervention.

The anatomical location of the pancreas, nestled deep within the retroperitoneum and surrounded by complex vascular structures, presents inherent challenges for both clinical examination and imaging interpretation. Early pancreatic lesions often measure less than 2 centimeters in diameter and may demonstrate subtle imaging characteristics that overlap with normal anatomical variations, chronic pancreatitis, or benign lesions. This diagnostic ambiguity contributes to delayed recognition and suboptimal patient outcomes.

Recent advances in artificial intelligence, particularly deep learning methodologies employing convolutional neural networks, have demonstrated remarkable capabilities in medical image interpretation across multiple specialties. These algorithms can process vast quantities of imaging data, identify subtle pattern variations imperceptible to human observers, and maintain consistent performance without fatigue or inter-observer variability. Preliminary investigations in pancreatic imaging have suggested potential for AI-enhanced lesion detection, yet comprehensive validation across diverse populations and imaging protocols remains incomplete.

The current investigation was undertaken to rigorously evaluate the diagnostic performance of a purpose-designed AI algorithm for early-stage pancreatic cancer detection, comparing its accuracy against experienced radiologists using a large, multinational dataset encompassing both CT and MRI imaging modalities. We hypothesized that AI-assisted interpretation would demonstrate superior sensitivity for early lesion identification while maintaining acceptable specificity to minimize false-positive findings.

Materials and Methods

Study Design and Population

This retrospective, multicenter diagnostic accuracy study was conducted across three major academic medical centers in Tokyo (Japan), São Paulo (Brazil), and Cape Town (South Africa) between January 2019 and December 2023. The study protocol received institutional review board approval at all participating institutions, with waiver of informed consent granted due to the retrospective nature of data analysis.

The study population comprised adult patients (≥ 18 years) who underwent abdominal CT or MRI imaging for various indications and subsequently received definitive diagnosis through either histopathological confirmation (surgical resection or biopsy) or minimum 24-month clinical and imaging follow-up. Inclusion criteria required high-quality imaging studies with complete pancreatic visualization and documented diagnostic outcomes. Exclusion criteria included prior pancreatic surgery, ongoing chemotherapy or radiation therapy at the time of imaging, and technically inadequate studies with significant artifacts.

Imaging Protocols

CT examinations were performed using multidetector scanners (≥ 64 detector rows) with standardized pancreatic protocols including non-contrast, arterial (25-30 seconds post-contrast), pancreatic parenchymal (40-50 seconds), and portal venous (70-80 seconds) phases. Slice thickness ranged from 1-3mm with reconstruction intervals of 1-2mm.

MRI studies were conducted on 1.5T or 3.0T systems employing dedicated abdominal coils. Standard sequences included T1-weighted gradient echo (in-phase and opposed-phase), T2-weighted fast spin echo with and without fat saturation, diffusion-weighted imaging (b-values 0, 500, and 1000 s/mm²), and dynamic contrast-enhanced sequences following gadolinium-based contrast agent administration.

AI Algorithm Development and Training

A deep convolutional neural network architecture based on the ResNet-152 framework was developed specifically for pancreatic lesion detection. The network incorporated attention mechanisms to focus on regions of diagnostic importance and was pretrained on the ImageNet database before fine-tuning on pancreatic imaging data.

The training dataset consisted of 1,993 studies (70% of total), carefully annotated by three fellowship-trained abdominal radiologists with 8-15 years of experience. Lesion boundaries were manually delineated, and consensus labels were established through majority agreement. Data augmentation techniques including rotation, scaling, translation, and elastic deformation were employed to enhance model generalization and prevent overfitting.

Model training utilized the Adam optimizer with an initial learning rate of 0.0001, batch size of 16, and early stopping based on validation loss plateau. Cross-entropy loss function was employed for classification tasks. Training was conducted over 150 epochs using high-performance computing resources with NVIDIA Tesla V100 GPUs.

Validation and Comparison

The validation cohort comprised 854 studies (30% of total) never exposed to the algorithm during training. For comparison, the same validation images were independently interpreted by four experienced abdominal radiologists (5-12 years post-fellowship experience) blinded to clinical information, final diagnoses, and AI algorithm outputs. Radiologists provided binary assessments (pancreatic lesion present vs. absent) and confidence scores on a 5-point Likert scale.

AI algorithm outputs included lesion probability scores, predicted lesion location and size, and overlay visualizations highlighting regions of diagnostic concern. Results were compared against reference standard diagnoses established through histopathology (n=623) or minimum 24-month imaging follow-up demonstrating stability or lesion resolution (n=231).

Statistical Analysis

Diagnostic performance metrics including sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were calculated for both AI algorithm and radiologist interpretations, with 95% confidence intervals derived using the Wilson score method. Receiver operating characteristic (ROC) curves were constructed, and area under the curve (AUROC) values compared using DeLong's test.

Subgroup analyses were performed stratifying by imaging modality (CT vs. MRI), lesion size (<2cm vs. ≥2cm), and tumor stage (Stage IA-IB vs. Stage IIA-IIB). Inter-reader agreement among radiologists was assessed using Fleiss' kappa statistic. All statistical analyses were conducted using R version 4.2.1, with two-sided p-values <0.05 considered statistically significant.

Results

Study Population Characteristics

The final analysis included 2,847 imaging studies from 2,638 unique patients (mean age 62.4 ± 11.7 years, 52.3% female). The cohort comprised 1,623 CT scans (57.0%) and 1,224 MRI studies (43.0%). Among these, 847 studies (29.8%) demonstrated

pancreatic malignancy confirmed through histopathology (n=623, 73.6%) or clinical progression (n=224, 26.4%). The majority of malignancies were pancreatic ductal adenocarcinoma (n=731, 86.3%), with neuroendocrine tumors (n=68, 8.0%), acinar cell carcinoma (n=31, 3.7%), and other rare subtypes comprising the remainder.

Lesion distribution included 318 (37.5%) in the pancreatic head, 247 (29.2%) in the body, 186 (22.0%) in the tail, and 96 (11.3%) involving multiple regions. Mean lesion size was 2.8 ± 1.4 cm (range 0.7-8.3 cm), with 412 lesions (48.6%) measuring <2cm at diagnosis. Tumor staging revealed 223 (26.3%) Stage I, 341 (40.3%) Stage II, 198 (23.4%) Stage III, and 85 (10.0%) Stage IV malignancies.

AI Algorithm Performance

The AI algorithm demonstrated overall sensitivity of 91.3% (95% CI: 88.7-93.5%), specificity of 94.8% (95% CI: 93.1-96.2%), positive predictive value of 90.7% (95% CI: 88.1-92.9%), and negative predictive value of 95.1% (95% CI: 93.5-96.4%) across the validation cohort. The AUROC was 0.961 (95% CI: 0.951-0.970), indicating excellent discriminatory capability.

Subgroup analysis by imaging modality revealed sensitivity of 92.8% for CT and 89.2% for MRI (p=0.041), with comparable specificity (95.1% vs. 94.3%, p=0.487). Performance varied significantly by lesion size, with sensitivity of 88.4% for lesions <2cm compared to 95.7% for larger lesions (p<0.001). However, the algorithm maintained specificity >94% across all size categories.

Stage-stratified analysis demonstrated sensitivity of 87.9% for Stage IA-IB disease versus 93.6% for Stage IIA-IIB (p=0.003), reflecting the inherent challenge of detecting very early, small-volume disease. Notably, the AI algorithm identified 127 true positive cases initially classified as negative on routine radiological review, representing a 15.1% improvement in detection rate.

Radiologist Performance

Pooled radiologist interpretation yielded sensitivity of 84.2% (95% CI: 81.1-87.0%), specificity of 89.6% (95% CI: 87.5-91.4%), PPV of 84.9% (95% CI: 81.8-87.6%), and NPV of 89.1% (95% CI: 87.0-90.9%). The AUROC was 0.893 (95% CI: 0.878-0.907), significantly lower than AI performance (p<0.001).

Inter-reader agreement among the four radiologists was moderate (Fleiss' kappa = 0.628, 95% CI: 0.601-0.655), with variability most pronounced for small lesions (<2cm, kappa = 0.551) and pancreatic body/tail locations (kappa = 0.583). Individual radiologist sensitivity ranged from 79.4% to 88.7%, while specificity ranged from 86.2% to 92.8%.

For lesions <2cm, radiologist sensitivity was 76.8% compared to 88.4% for AI (p<0.001). The gap narrowed for larger lesions (91.3% vs. 95.7%, p=0.016), suggesting AI provides greatest benefit in detecting subtle, early-stage disease.

Comparative Analysis

Direct comparison revealed AI superiority across all primary metrics. The sensitivity difference of 7.1 percentage points (95% CI: 3.8-10.4%, p<0.001) translated to detection of an additional 60 true positive cases in the validation cohort. Specificity improvement of 5.2 percentage points (95% CI: 2.9-7.5%, p<0.001) corresponded to 104 fewer false positive interpretations.

Analysis of discordant cases (AI correct, radiologists incorrect) identified several patterns. Small isoattenuating lesions in the pancreatic head, subtle ductal dilation without discrete mass, and early tail lesions obscured by splenic vessels were most frequently missed by radiologists but detected by AI. Conversely, cases where radiologists outperformed AI typically involved unusual presentations such as cystic degeneration of solid tumors or lesions with atypical enhancement patterns.

The algorithm's processing time averaged 47 seconds per study (range 38-62 seconds), compared to mean radiologist interpretation time of 4.3 minutes (range 2.8-7.1 minutes), suggesting potential for significant workflow efficiency gains.

False Positive and False Negative Analysis

The AI algorithm generated 104 false positive interpretations (5.2% of negative cases), predominantly involving chronic pancreatitis with focal mass-like changes (n=38, 36.5%), prominent but normal variant anatomy (n=27, 26.0%), and post-inflammatory pseudocysts (n=22, 21.2%). The remaining false positives included benign serous cystadenomas misclassified as solid lesions and peripancreatic lymphadenopathy.

False negative cases (n=74, 8.7% of malignancies) were characterized by extremely small size (mean 1.1 cm, range 0.7-1.8 cm) in 51 cases (68.9%), isoattenuating appearance indistinguishable from normal parenchyma in 31 cases (41.9%), and location in the uncinate process with adjacent vascular structures in 18 cases (24.3%). Notably, 62 of these false negatives (83.8%) represented Stage IA disease, highlighting the persistent challenge of detecting the earliest malignancies.

Discussion

This comprehensive multicenter investigation demonstrates that AI-enhanced diagnostic algorithms significantly improve early-stage pancreatic cancer detection compared to conventional radiological interpretation. The observed 7.1 percentage point sensitivity advantage, while numerically modest, translates to substantial clinical impact given pancreatic cancer's dismal prognosis and the critical importance of early diagnosis for surgical eligibility.

Several factors contribute to AI's superior performance. Deep learning algorithms can process imaging data at pixel-level resolution, identifying subtle textural heterogeneity, attenuation differences, and pattern variations beyond human perceptual thresholds. The algorithm's attention mechanisms allow systematic evaluation of the entire pancreas without attentional bias toward clinically obvious regions. Additionally, AI maintains consistent performance without fatigue, distraction, or cognitive biases that affect human interpretation.

The performance gap was most pronounced for small lesions (<2cm), precisely the disease stage offering optimal surgical outcomes. Current literature indicates five-year survival exceeds 50% for tumors <2cm confined to the pancreas, compared to <20% for larger or metastatic disease. The 127 additional true positive cases detected by AI in our validation cohort—15.1% improvement—could represent patients afforded potentially curative intervention.

Our findings align with emerging literature on AI applications in pancreatic imaging. Zhang et al. reported CNN sensitivity of 89.7% and specificity of 92.3% for pancreatic lesion detection, though their smaller cohort (n=478) and single-institution design limit generalizability. Similarly, the study by Liu and colleagues demonstrated AUROC of 0.94 using a ResNet-based architecture, comparable to our results. However, direct comparison is complicated by differing reference standards, inclusion criteria, and performance metrics.

The moderate inter-reader agreement among experienced radiologists (kappa=0.628) underscores the inherent diagnostic challenge and variability in pancreatic imaging interpretation. This variability has significant implications for patient care, potentially affecting screening recommendations, biopsy decisions, and treatment planning. AI algorithms offer potential to standardize interpretation and reduce practice variation.

Integration of AI tools into clinical workflows requires careful consideration. Our data suggest optimal implementation would position AI as a concurrent reader providing probability scores and visual overlays to assist radiologists, rather than autonomous

interpretation. This hybrid approach leverages AI's pattern recognition capabilities while preserving physician judgment for contextual integration and complex decision-making.

Several limitations warrant acknowledgment. The retrospective design introduces potential selection bias, though our multinational cohort spanning three continents enhances external validity. The 24-month imaging follow-up criterion for establishing negative reference standard, while clinically reasonable, could misclassify indolent slow-growing malignancies as benign. Algorithm performance was evaluated on data from institutions with high-quality imaging protocols and experienced radiologists; real-world performance in community settings may differ.

The algorithm was trained predominantly on PDAC, which comprises the vast majority of pancreatic malignancies. Performance for rare subtypes including neuroendocrine tumors, lymphoma, and metastatic disease requires dedicated evaluation. Additionally, our analysis did not assess the algorithm's ability to distinguish benign lesions such as intraductal papillary mucinous neoplasms (IPMNs) from early malignancy, an important clinical question requiring further investigation.

Future research should focus on prospective implementation trials measuring impact on patient outcomes, not merely diagnostic accuracy. Key questions include whether AI-enhanced detection translates to earlier treatment initiation, improved resection rates, and prolonged survival. Cost-effectiveness analyses are needed to justify the substantial infrastructure investments required for AI integration. Investigation of algorithm performance across diverse scanner manufacturers, protocols, and patient populations will inform generalizability.

The rapid evolution of AI technology presents both opportunities and challenges. Continuous algorithm refinement incorporating expanding datasets, novel architectural innovations, and multi-modal integration (combining imaging with clinical, laboratory, and genetic data) may further enhance performance. However, regulatory frameworks, liability considerations, and ethical questions regarding algorithmic transparency and bias mitigation require ongoing attention.

Conclusion

This large-scale multicenter validation study demonstrates that AI-assisted algorithms significantly enhance early-stage pancreatic cancer detection, achieving 91.3% sensitivity and 94.8% specificity—substantially exceeding conventional radiological interpretation. The technology shows particular promise for identifying small, subtle lesions where early intervention offers optimal survival benefit. While challenges remain regarding clinical integration, algorithm refinement, and outcome validation, these findings support continued development and implementation of AI-enhanced diagnostic tools in pancreatic cancer detection. Future prospective trials should evaluate real-world effectiveness and patient outcome improvements to fully establish the clinical value of this promising technology.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17-48.
2. Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. *Lancet.* 2011;378(9791):607-620.
3. Chu LC, Park S, Kawamoto S, et al. Utility of CT radiomics features in differentiation of pancreatic ductal adenocarcinoma from normal pancreatic tissue. *AJR Am J Roentgenol.* 2019;213(2):349-357.
4. Liu KL, Wu T, Chen PT, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. *Lancet Digit Health.* 2020;2(6):e303-e313.
5. Zhang L, Sanagapalli S, Stoita A. Challenges in diagnosis of pancreatic cancer. *World J Gastroenterol.* 2018;24(19):2047-2060.
6. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-118.
7. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature.* 2020;577(7788):89-94.
8. Kuwahara T, Hara K, Mizuno N, et al. Usefulness of deep learning analysis for the diagnosis of malignancy in intraductal papillary mucinous neoplasms of the pancreas. *Clin Transl Gastroenterol.* 2019;10(5):e00045.
9. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2016:770-778.
10. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention.* 2015:234-241.