Radiology

REVIEWS AND COMMENTARY • EDITORIAL

Artificial Intelligence Outperforms Radiologists for Pancreatic Cancer Lymph Node Metastasis Prediction at CT

Linda C. Chu, MD • Elliot K. Fishman, MD

Dr Linda Chu is an associate professor and associate division director in the Diagnostic Division of the Department of Radiology at Johns Hopkins University. Her research interest focuses on artificial intelligence applications in pancreatic imaging. Dr Chu is the associate editor of podcasts for *Radiology*, an associate editor for *Journal of Computer Assisted Tomography*, and a co-investigator for a Lustgarten Foundation grant focused on pancreatic cancer detection.

Dr Elliot Fishman is a professor of radiology and division director in the Diagnostic Division of the Department of Radiology at Johns Hopkins University. His research interests include artificial intelligence and advanced 3D visualization techniques. Dr Fishman is an editorial board member for more than 35 journals and a past president of the Society of Body CT/MR. He is the principal investigator for a Lustgarten Foundation grant.

Pancreatic cancer is the third leading cause of cancer-related death in the United States, and pancreatic ductal adenocarcinoma (PDAC) accounts for 95% of all pancreatic cancers. Despite treatment advances in the past few decades, the 5-year overall survival rate remains approximately 10% (1).

Surgical resection is the only potentially curative treatment for patients with PDAC and, unfortunately, up to 80% of patients who undergo surgical resection have recurring cancer within 2 years (2). For patients with resectable disease with high-risk features of recurrence (eg, large tumor size, high carbohydrate antigen [CA]) 19–9 level, suspected lymph node [LN] metastasis), many centers offer neoadjuvant chemotherapy before surgery. Neoadjuvant chemotherapy can improve the likelihood of margin-negative resection, provide upfront treatment of occult micrometastases, and avoid unnecessary (and futile) surgery in patients with aggressive tumor types (1). Therefore, accurate preoperative evaluation of LN status is important in triaging patients who will most likely benefit from neoadjuvant chemotherapy and who should proceed to upfront surgery.

The prediction of LN metastasis from preoperative imaging is challenging. In a previous meta-analysis, preoperative CT-based prediction of extraregional LN metastases in pancreatic and periampullary cancer only showed a pooled sensitivity of 25% and positive predictive value of 28% (3). Recently, there is mounting evidence that radiomics features extracted from the PDAC tumor region can help predict the presence or absence of LN metastasis (4–7). Most of these studies required manual segmentation of the tumor boundaries, a laborious process that limited the study sample sizes ($n < 300$). There is also a lot of subjectivity in the segmentation of tumor boundaries due to the infiltrative growth pattern typical of PDAC, which may limit its reproducibility. Therefore, automated tumor segmentation is a prerequisite for pancreatic cancer artificial intelligence (AI) research studies with sufficient scale for adequate validation and eventual clinical acceptance.

In this issue of *Radiology*, Bian et al (8) report on the development of an automated AI algorithm for segmentation of tumor and LN and prediction of LN metastasis at CT in patients with PDAC, as well as a comparison of the performance of the AI model with a clinical model and radiomics model. The study included 734 patients with surgically resected PDAC, of whom 394 (54%) had PDAC without LN metastasis and 340 (46%) had PDAC with LN metastasis. Arterial and portal venous phase images were reviewed by two abdominal radiologists. Suspicious LNs were defined based on a short-axis diameter greater than 10 mm, heterogeneity, ill-defined borders, or involvement of surrounding organs or blood vessels. The clinical prediction model incorporated pathologic and radiologic features. The AI model used deep network automatic tumor and LN segmentation, and an additional deep network model used imaging features of both primary tumor and LNs to predict LN metastasis. The radiomics model was derived from an arterial radiomics score (Rad-score; based on five radiomics features from an initial 1688) that was extracted from automatically segmented tumor boundaries in combination with suspected LN metastasis based on CT features.

The AI model showed the best performance in discriminating between patients with PDAC with or without LN metastasis. In the validation set, the areas under the receiver operating characteristic curve (AUCs) were 0.92 for the AI model, 0.65 for suspicious LNs based on CT criteria, 0.77 for the clinical model, and 0.68 for the radiomics model. Furthermore, AI-predicted LN metastasis was an independent preoperative predictor for worse

From the Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Hal B168, 600 N Wolfe St, Baltimore, MD 21287. Received August 9, 2022; revision requested August 15; revision received August 17; accepted August 19. **Address correspondence to** L.C.C. (email: *lindachu@jhmi.edu*).

Conflicts of interest are listed at the end of this article.

See also the article by Bian et al in this issue.

overall survival in a multivariable Cox regression analysis (hazard ratio, 1.46; P = .004), which may be helpful in guiding patient management.

One of the distinctive aspects of this study is the integration of an automated segmentation model, allowing the segmented data to be analyzed with an independent deep network or radiomics LN prediction model. This automation made it feasible to analyze larger data sets, such that the sample size of the study (n = 734) was almost three times larger than previous studies that used manually segmented PDAC tumor boundaries to predict LN metastasis (4–7). With any automated segmentation pipeline, the segmentation accuracy of the regions of interest becomes critically important. If the segmentation accuracy is poor, then it becomes "garbage in, garbage out."

The automated model detected the tumor in 100% (189 of 189) of patients, with a mean tumor size of less than 3 cm in the validation set; the distribution of tumor sizes and tumor stages was similar to what we would encounter clinically. The overlap between the automated and manual segmentation as measured by the median Dice-Sørensen coefficient, also known as the Dice similarity coefficient (DSC), was 0.68 for tumor and 0.59 for LN. This modest agreement between AI and manual segmentation raises concern that errors in the tumor and LN segmentation may propagate downstream, and the features extracted by the deep learning or radiomics model may not be truly representative of the underlying disease process. The authors previously investigated a radiomics model by using manually segmented PDAC tumor boundaries as input in 225 patients and reported an AUC of 0.81 in the validation cohort (7). This contrasts with the AUC of 0.68 in the current study that used automatic segmentation. Some of this reduction in model performance may be due to segmentation errors or differences in the study sample. Segmentation errors may be even more problematic for small structures (eg, LNs) because minor errors along the boundaries of smaller structures will have a greater impact on the DSC than similar errors along the boundaries of larger structures (eg, pancreas). The authors addressed this issue by demonstrating that the modest automated LN segmentation accuracy (DSC, 0.59) was comparable or even superior to radiologists (median interobserver DSC, 0.24; median intraobserver DSC, 0.49) and represented what was currently achievable.

In addition, the Bian et al (8) study incorporated imaging features of the LNs in the prediction model, whereas previous studies relied only on imaging features of the primary tumor. It can be challenging to correlate the exact LNs on radiologic and pathologic images retrospectively. At the authors' institution, pathologists reported the number of positive LNs and specific locations. Also, the authors were able to map out the location of the positive LNs into 18 LN stations in the peripancreatic region based on the detailed pathology reports to improve the accuracy of LN annotations. Using DSCs greater than or equal to 0.5 as a true-positive of LN detection, the true-positive rate and positive predictive value of the automated model were 80.1% and 40.5%, respectively. We speculate that the combination of tumor and LN features should improve the prediction of LN metastasis. The number of LN metastases and the number of positive LNs–to–sampled LNs (LN ratio) are important prognostic indicators (1). What remains controversial is whether the involvement of specific LN groups has any prognostic significance.

This study had several limitations. First, it was a retrospective single-center study with potential for bias. The promising results should be validated in prospective multicenter trials, which can be facilitated by the automated workflow. Second, the authors did not include CA 19–9 levels in the prediction model. Elevated CA 19–9 is an indicator of poor prognosis and could be readily incorporated into future AI prognostication studies. Third, segmentation errors from automatic segmentation could lead to downstream errors in the deep network or radiomics prediction models, as discussed earlier.

In conclusion, Bian et al (8) have demonstrated the feasibility of an automated pipeline. Future studies should refine the automated segmentation models to improve the quality of input data for the classification models. Similarly, a fully automated segmentation and CT-based prognostic marker model was recently developed that was predictive of overall survival in 1516 patients with resectable PDAC (9). These automated approaches can be applied to other challenging problems in pancreatic cancer imaging, such as initial staging and assessment of treatment response. Although current guidelines provide criteria for triaging patients into resectable, borderline, and locally advanced categories, there remains substantial subjectivity and variability among radiologists. Staging becomes even more difficult in the neoadjuvant setting in differentiating inflammation or fibrosis from viable tumors. AI has the potential to improve staging and our understanding of tumor biology, and to help select the patients most likely to benefit from surgery and tailor the most appropriate neoadjuvant or adjuvant therapy. In conjunction with current efforts focused on earlier detection of PDAC (10), AI has the potential to transform how we care for patients with pancreatic cancer and offers hope to patients with this dreadful disease.

Disclosures of conflicts of interest: L.C.C. Associate editor of podcasts, *Radiology*. **E.K.F.** No relevant relationships.

References

1. Grossberg AJ, Chu LC, Deig CR, et al. Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. CA Cancer J Clin 2020;70(5):375–403.
2. Groot VP, Rezaee N, Wu W, et al. Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. Ann Surg 2018;267(5):936–945.
3. Tseng DS, van Santvoort HC, Fegrachi S, et al. Diagnostic accuracy of CT in assessing extra-regional lymphadenopathy in pancreatic and periampullary cancer: a systematic review and meta-analysis. Surg Oncol 2014;23(4):229–235.
4. Bian Y, Guo S, Jiang H, et al. Relationship Between Radiomics and Risk of Lymph Node Metastasis in Pancreatic Ductal Adenocarcinoma. Pancreas 2019;48(9):1195–1203.
5. Li K, Yao Q, Xiao J, et al. Contrast-enhanced CT radiomics for predicting lymph node metastasis in pancreatic ductal adenocarcinoma: a pilot study. Cancer Imaging 2020;20(1):12.
6. Shi L, Wang L, Wu C, Wei Y, Zhang Y, Chen J. Preoperative Prediction of Lymph Node Metastasis of Pancreatic Ductal Adenocarcinoma Based on a Radiomics Nomogram of Dual-Parametric MRI Imaging. Front Oncol 2022;12:927077.

7. Bian Y, Guo S, Jiang H, et al. Radiomics nomogram for the preoperative prediction of lymph node metastasis in pancreatic ductal adenocarcinoma. Cancer Imaging 2022;22(1):4.
8. Bian Y, Zheng Z, Fang X, et al. Artificial Intelligence to Predict Lymph Node Metastasis at CT in Pancreatic Ductal Adenocarcinoma. Radiology 2022. https://doi.org/10.1148/radiol.220329. Published online September 6, 2022.
9. Yao J, Cao K, Hou Y, et al. Deep Learning for Fully Automated Prediction of Overall Survival in Patients Undergoing Resection for Pancreatic Cancer: A Retrospective Multicenter Study. Ann Surg 2022. 10.1097/SLA.0000000000005465. Published online July 4, 2022.
10. Mukherjee S, Patra A, Khasawneh H, et al. Radiomics-Based Machine-Learning Models Can Detect Pancreatic Cancer on Prediagnostic CTs at a Substantial Lead Time Prior to Clinical Diagnosis. Gastroenterology 2022. 10.1053/j.gastro.2022.06.066. Published online July 1, 2022.