EDITORIAL

Risk stratification of pancreatic cysts: a convoluted path to finding the needle in the haystack

Improved resolution and quality of abdominal cross-sectional imaging and its widespread clinical use has been associated with an increase in the incidental detection of pancreatic cystic lesions.[1,2] Intraductal papillary mucinous neoplasms (IPMNs) are the most common type of incidentally detected pancreatic cysts, and although they are precancerous, most of them will never progress to cancer.[3-5] Current risk stratification guidelines for the management of presumed IPMNs aim to identify cysts with advanced neoplasia: high-grade dysplasia or invasive cancer, by use of a combination of clinical and imaging characteristics.[3,4] The only definitive treatment available currently for IPMNs is surgical resection, and in clinical practice fewer than half of surgically resected IPMNs harbor advanced neoplasia. Although surgery at the stage of high-grade dysplasia is associated with improved survival, identifying this small target remains a challenge.[6] Accurate detection of advanced neoplasia in IPMNs has been the focus of several recent studies exploring novel biomarkers and artificial intelligence (AI)−aided imaging modalities.[7]

EUS is commonly used in clinical practice for risk stratification of pancreatic cysts and is usually reserved for cysts with worrisome features identified on cross-sectional imaging. EUS allows for cytologic evaluation of cyst fluid or an intracystic nodule when present. Cytologic analysis is informative when the result is positive for advanced neoplasia, but it lacks sensitivity, making a negative test result clinically low value. Needle-based confocal laser endomicroscopy (nCLE) is a novel EUS-based approach that has the potential to improve the detection of advanced neoplasia in IPMNs. It is based on real-time imaging of cyst epithelium. An antegrade EUS locates the IPMN, and a probe is passed through the EUS needle to the interior of the cyst. After intravenous fluorescein is administered for improved resolution, the probe then uses a low-power laser to acquire high magnification and resolution images with the aim to evaluate epithelium, vasculature, and cellularity.[8] Proof-of-concept studies have been completed that demonstrate the feasibility of this technique in differentiating different histologic subtypes of pancreatic cystic lesions[9] and for the detection of advanced neoplasia with reasonable accuracy.[10]

0016-5107/$36.00
https://doi.org/10.1016/j.gie.2021.03.012

In this issue of *Gastrointestinal Endoscopy*, Machicado et al[11] present a post-hoc analysis of their single-center prospective study (INDEX study NCT0251648) and aimed to derive convolutional neural network computer-aided diagnosis (CNN-CAD) algorithms for risk stratification of IPMNs. They compared 2 CNN-CAD algorithms for the detection of advanced neoplasia in IPMNs using pathologic dysplasia grade as a criterion standard and concluded that both algorithms can accurately risk stratify IPMNs, thereby providing early data on the feasibility of this novel diagnostic approach.

In their study, EUS-nCLE videos from 35 consecutive individuals with IPMNs and confirmed histopathologic features were edited to eliminate noncontributory images and annotated by blinded observers to yield 15,027 video frames. Two CNN-CAD algorithms were then created by use of the known dysplasia grade as the "ground truth." One was a holistic-based model that extracted its own nCLE features using only the ground truth, whereas the other was a segmentation model that used the annotated regions of interest, papillary thickness, and darkness: nCLE features with the highest diagnostic accuracy and interobserver agreement in a previous study by the authors.[10] The reported outcomes for both CNN-CAD models included sensitivity, specificity, and accuracy of differentiating advanced neoplasia from low-grade dysplasia. Both models were compared with the sensitivity, specificity, and accuracy of the American Gastroenterology Association surgical criteria[4] (56%, 82%, 69%) and Fukuoka international consensus guidelines "high-risk stigmata"[3] (56%, 94%, 74%). Both CNN-CAD models demonstrated comparable specificity, the authors reported improved

> The world we live in is changing rapidly, and the healthcare of tomorrow is anticipated to be very different from that of today. Although the clinical application of AI in healthcare is at a nascent stage, the medical literature is already colonized with numerous early-stage studies promising transformation and forecasting an unforeseen disruption that will overcome limitations in our current understanding of diagnosis and treatment.

sensitivity and accuracy compared with the reported guidelines, although statistical significance was limited by the small sample size.

IPMNs are frequently multifocal. A diagnostic modality for detection of advanced neoplasia that requires cyst puncture will ideally need to be paired with an algorithm for determining the "target" cyst and avoid the need to puncture multiple cysts, thereby minimizing procedure-related risks such as bleeding and pancreatitis. Moreover, for IPMNs involving the main pancreatic duct, the advanced neoplasia is frequently in the main duct, not in the epithelium located in a branch-duct cyst, and hence is not always amenable to interrogation with the use of cyst puncture.[12] Finally, spontaneous denudation of cyst lining epithelium has been described in both mucinous cystic neoplasms and IPMNs and may limit the application of EUS-nCLE.[13] These are key considerations for future applications of this innovative approach to risk stratification of IPMNs. Molecular markers assayed in secretin-stimulated pancreatic juice collected from the duodenum have been studied as a minimally invasive approach for the detection of pancreatic cancer.[14] One could envision a future state in which pancreatic juice biomarkers are used in combination with novel EUS imaging techniques such as the one described here for risk stratification of IPMNs. Further refinement of the knowledge gained in this study to develop EUS-based imaging techniques that avoid the need for cyst puncture may have a more acceptable risk profile for IPMN surveillance.

A widely acceptable method for determining sample size for an AI-based image classification is currently not available, and AI-based and machine learning–based image classification algorithms in small sample sets, such as those described in this study, remain limited in scope to disease classification, not diagnosis. However, the current study provides proof of concept and is hypothesis generating for studies exploring larger datasets with an aim to develop AI tools that can be incorporated as diagnostic tests in clinical practice. A key limitation in the field is the absence of high-quality EUS-nCLE video data that are paired with surgical pathologic results in IPMNs, and the fact that it took the authors several years to assemble this small sample set of high-quality videos linked to pathologic data further highlights the limitation.

Data preparation is a key element in AI-based approaches. In this study, videos were edited to remove noncontributory images such as image blur due to rapid probe movement, artifacts, absence of discernible patterns, and static images that reduced the mean duration of nCLE video per patient from 6.72 minutes to 35.8 seconds. Eight nCLE-naïve observers performed the identification and demarcation of IPMN papillae for the segmentation model. An important aspect of using multiple observers for image segmentation is to report reproducibility and/or interobserver variability specific to the segmentation task. Dice similarity coefficient (DSC) is a commonly used metric to report reproducibility in segmentation tasks.[15] Allocating common variability training sets to all observers and comparing the DSCs of the contours for each is critically important for the interpretation of downstream results, and an interobserver variability of the DSC of <10% is considered reasonable.[16] Adoption of this mechanism would eliminate propagation of the errors resulting from variability and bias for including several observers into model performance. Classification and prediction models developed on the basis of known features provide interpretability of the model outcomes. However, caution should be used while selecting the top features using dimensionality reduction techniques that the model uses for classification or prediction. Importantly, features that are dependent or correlated with each other should be eliminated because that would favorably improve model outcomes due to association.

The authors used all the 15,027 frames after preliminary video editing to develop and test the CNN models in a 5-fold cross-validation scheme. During each fold experiment, the authors divided the dataset into training (60%), validation (20%), and testing (20%), without reserving an independent test set. Although the authors indicate that a small final test, if reserved, may cause high variance, the model performance might have been affected by testing the same frames that were used for training. The use of an independent set of images to test the algorithm, which has not been seen by the model during training, leads to refinement of the methodologic approach. To avoid bias in the dataset, race, age, and gender are important considerations for generalizing the findings from this work and for clinical translation to the use of a CNN-CAD for diagnosing advanced neoplasia in IPMNs with EUS-nCLE.

The authors should be commended for their pioneering work combining AI and EUS video data to address a clinically relevant knowledge gap in risk stratification of IPMNs. The results provide a potential approach to overcoming the interobserver disagreement that has been described as a limitation of nCLE qualitative assessments.[10]

The world we live in is changing rapidly, and the healthcare of tomorrow is anticipated to be very different from that of today. Although the clinical application of AI in healthcare is at a nascent stage, the medical literature is already colonized with numerous early-stage studies promising transformation and forecasting an unforeseen disruption that will overcome limitations in our current understanding of diagnosis and treatment. This inherent excitement associated with novel technology needs to be intelligently curated by master clinicians in partnership with data science experts. As we develop models applicable to patient care, it will be critically important to envision and plan for iterative improvements, optimize applications that minimize burnout, and seamlessly integrate these models into existing workflows. It is an exciting time for investigators in the field of early detection of cancer; the path is convoluted, but we are inching closer to finding the needle in the haystack.

DISCLOSURE

Dr Majumder is listed as an inventor of intellectual property owned by Mayo Clinic and Exact Sciences Corporation and a potential recipient of royalties paid to Mayo Clinic. The other authors disclosed no financial relationships.

Jaime de la Fuente, MD
Shivaram P. Arunachalam, PhD, MS
Shounak Majumder, MD
Division of Gastroenterology and Hepatology
Mayo Clinic
Rochester, Minnesota, USA

Abbreviations: AI, artificial intelligence; CAD, computer-aided diagnosis; CNN, convolutional neural network; DSC, dice similarity coefficient; IPMNs, intraductal papillary mucinous neoplasms; nCLE, needle confocal laser endomicroscopy.

REFERENCES

1. Moris M, Bridges MD, Pooley RA, et al. Association between advances in high-resolution cross-section imaging technologies and increase in prevalence of pancreatic cysts from 2005 to 2014. Clin Gastroenterol Hepatol 2016;14:585-93.e3.
2. Smith-Bindman R, Miglioretti DL, Larson EB. Rising use of diagnostic medical imaging in a large integrated health system. Health Aff (Millwood) 2008;27:1491-502.
3. Tanaka M, Fernández-Del Castillo C, Kamisawa T, et al. Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas. Pancreatology 2017;17:738-53.
4. Vege SS, Ziring B, Jain R, et al. American Gastroenterological Association Institute guideline on the diagnosis and management of asymptomatic neoplastic pancreatic cysts. Gastroenterology 2015;148:819-22.
5. Valsangkar NP, Morales-Oyarvide V, Thayer SP, et al. 851 resected cystic tumors of the pancreas: a 33-year experience at the Massachusetts General Hospital. Surgery 2012;152:S4-12.
6. Majumder S, Philip NA, Singh Nagpal SJ, et al. High-grade dysplasia in resected main-duct intraductal papillary mucinous neoplasm (MD-IPMN) is associated with an increased risk of subsequent pancreatic cancer. Am J Gastroenterol 2019;114:524-9.
7. de la Fuente J, Majumder S. Molecular diagnostics and testing for pancreatic cysts. Curr Treat Options Gastroenterol. Epub 2020 Jan 27.
8. Chauhan SS, Abu Dayyeh BK, Bhat YM, et al. Confocal laser endomicroscopy. Gastrointest Endosc 2014;80:928-38.
9. Krishna SG, Hart PA, Malli A, et al. Endoscopic ultrasound-guided confocal laser endomicroscopy increases accuracy of differentiation of pancreatic cystic lesions. Clin Gastroenterol Hepatol 2020;18:432-40.e6.
10. Krishna SG, Hart PA, DeWitt JM, et al. EUS-guided confocal laser endomicroscopy: prediction of dysplasia in intraductal papillary mucinous neoplasms (with video). Gastrointest Endosc 2020;91:551-63.e5.
11. Machicado JD, Chao W-L, Carlyn DE, et al. High performance in risk stratification of intraductal papillary mucinous neoplasms by confocal laser endomicroscopy image analysis with convolutional neural networks (with video). Gastrointest Endosc 2021;94:78-87.
12. Majumder S, Graham RP, Cheek EH, et al. Mapping ductal dysplasia in surgically resected pancreatic specimens containing intraductal papillary mucinous neoplasms [abstract]. In: Joint 50th Anniversary Meeting of the American Pancreatic Association and Japan Pancreas Society, November 6–9, 2019, Maui, Hawaii. Pancreas 2019;48:1484-5.
13. Gómez V, Majumder S, Smyrk TC, et al. Pancreatic cyst epithelial denudation: a natural phenomenon in the absence of treatment. Gastrointest Endosc 2016;84:788-93.
14. Majumder S, Raimondo M, Taylor WR, et al. Methylated DNA in pancreatic juice distinguishes patients with pancreatic cancer from controls. Clin Gastroenterol Hepatol 2020;18:676-83.e3.
15. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. Acad Radiol 2004;11:178-89.
16. Murray SW, Stables RH, Hart G, et al. Defining the magnitude of measurement variability in the virtual histology analysis of acute coronary syndrome plaques. Eur Heart J Cardiovasc Imaging 2013;14:167-74.