Europe=4, Asia=5) with experience in nCLE (median case volume of 50) and blinded to the histopathological diagnosis, participated in the study. Observers reviewed the same set of nCLE videos in different sequences between two phases separated by a 2-week washout period. Observers were first queried on PCL classification (mucinous vs. non-mucinous) and then on specific PCL diagnosis. Solid pseudopapillary neoplasm (SPN) and cystic neuroendocrine tumors (NET) have similar nCLE imaging features and were grouped together. Fleiss's kappa (k) for IOA and Cohen's k for IOR were interpreted using Landis and Koch interpretation.

Results: Observers reviewed 76 unique nCLE videos in subjects with confirmed histopathology (IPMN=37; mucinous cystic neoplasm (MCN)=14; serous cystadenoma (SCA)=9; pseudocyst=3; NET/SPN=13), most of which were rated as good quality (65%). The IOA (k=0.82) and IOR (k=0.82) were *almost perfect* to differentiate mucinous vs. non-mucinous PCLs (***Table 1***). For the diagnosis of individual cyst types, IOA was greatest for SCA (*almost perfect*; k=0.85), followed by IPMN (*substantial*, k=0.72), and NET/SPN (substantial, k=0.73). The IOA was *moderate* for MCN (k=0.47), *and* pseudocyst (k=0.57). Compared to histopathology, observers differentiated mucinous vs. non-mucinous PCLs with high diagnostic indices (sensitivity (SN)=95%, specificity (SP)=94%, and accuracy (Acc)=95%). Among non-mucinous PCLs, diagnostic indices were higher for SCA (SN=95%, SP=99%, Acc=98%) compared to NET/SPN (SN=81%, SP=99%, Acc=96%), and pseudocyst (SN=87%, SP=96%, Acc=96%). The interpretation of nCLE images represented by diagnostic accuracy of differentiating mucinous from non-mucinous PCLs (correlation coefficient R=0.152, p=0.62) was not dependent on individual EUS-nCLE case volume (***Figure 1***).

Conclusion: Compared to histopathology, independent blinded observers achieved high diagnostic accuracy, and excellent agreement and reliability for EUS-nCLE imaging guided diagnosis and differentiation of PCLs.

	IOA, kappa (95% CI) ϕ	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
Mucinous (IPMN/MCN) vs. non-mucinous PCLs	0.82 (0.77, 0.87)	95.2 (93.3, 96.6)	94.2 (91.1, 96.2)	94.8 (93.3, 96.1)
Differentiation within mucinous PCLs				
IPMN	0.72 (0.64, 0.79)	84.4 (80.9, 87.4)	87.9 (84.9, 90.5)	86.2 (83.9, 88.2)
MCN	0.47 (0.35, 0.59)	57.1 (49.9, 64.1)	90.2 (87.9, 92.1)	84.1 (81.7, 86.3)
Differentiation within non-mucinous PCLs				
Cystic-NET or SPN^	0.73 (0.64, 0.83)	80.5 (73.9, 85.7)	98.9 (97.9, 99.4)	95.8 (94.3, 96.8)
Pseudocyst	0.57 (0.36, 0.79)	87.2 (73.3, 94.4)	96.2 (94.8, 97.3)	95.9 (94.4, 96.9)
SCA	0.85 (0.76, 0.95)	94.9 (89.3, 97.6)	98.6 (97.6, 99.2)	98.2 (97.1, 98.8)

* Results of phase 2 are presented in the abstract and table, since results for phase 1 and 2 were comparable.
ϕ Landis and Koch interpretation for k values: <0, *no agreement*; 0-0.20, *slight*; 0.21-0.40, *fair*; 0.41-0.60, *moderate*; 0.61-0.80, *substantial*; 0.81-1, *almost perfect agreement*.
^ Since pseudopapillary neoplasm (SPN) and cystic-neuroendocrine tumor (cystic-NET) demonstrate similar nCLE imaging features, both cysts were combined as one diagnostic category for analysis.

IPMN: intraductal papillary mucinous neoplasm; MCN: mucinous cystic neoplasm; NET: neuroendocrine tumor; SPN: solid pseudopapillary tumor; SCA: serous cystadenoma; CI: confidence interval.

Table 1: Interobserver agreement (IOA) and diagnostic characteristics of EUS-guided needle-based confocal laser endomicroscopy (nCLE) for different PCLs.*

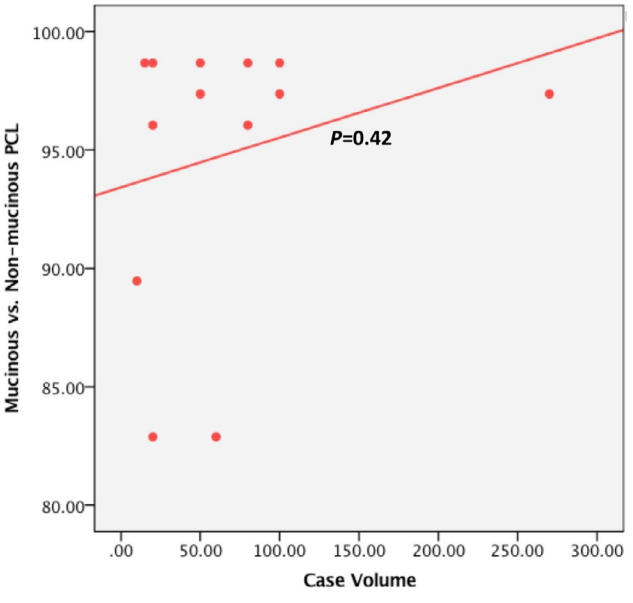

Figure 1. Correlation between individual observer EUS-guided needle-based confocal laser endomicroscopy (nCLE) case volume and diagnostic accuracy in nCLE image guided differentiation (mucinous vs. non-mucinous) of pancreatic cystic lesions (PCLs).

79

MISSED DIAGNOSIS OF PANCREATIC DUCTAL ADENOCARCINOMA DETECTION USING DEEP CONVOLUTIONAL NEURAL NETWORK

Sanne A. Hoogenboom, Kamalakkannan Ravi, Megan M. Engels, Ismail Irmakci, Elif Keles, Candice W. Bolan, Michael B. Wallace, Ulas Bagci

Background: The clinical diagnosis of Pancreatic ductal adenocarcinoma (PDAC) is challenging. Patients often present with non-specific symptoms, and up to one-third of patients are misdiagnosed. Because patients with PDAC that are initially misdiagnosed are associated with a higher risk of advanced disease at the time of diagnosis, early detection is essential for improving patient outcomes. Our purpose in this study is to detect missed diagnosis of PDAC cases using deep learning and understand how to distinguish the missed cancer cases from controls.

Methods and data: Due to the problem's nature, we have used imbalanced in-house and public data sets. We collected 184 abdominal 3D CT scans with 32,968 slices. The sample axial view of abdominal CT images is shown in Figure 1. The first set of 102 images were a mix of contrast-enhanced and non-contrast images from Mayo Clinic. Of the 102 images, 20 are missed PDAC samples, and the remaining are control images. An expert radiologist annotated all the control images, and histopathology confirmed the PDAC images. The second source of data is publicly available 82 abdominal control contrast-enhanced CT scans acquired by the National Institute of Health (NIH). Thus, the dataset is highly imbalanced, with 164 control and 20 missed PDAC images. Our train and validation set ratio is 80:20. We developed a convolutional neural network (CNN) model with ResNeSt-50 architecture and AlexNet for baseline. We use a class-dependent loss noted as label-distribution-aware margin (LDAM) and deferred re-weighting (DRW) to handle the class imbalance. The input is the axial slice of the abdominal CT scans, and the output is a binary classification of control and missed pancreatic cancer (PDAC).

Results: To evaluate this two-class imbalance problem, we use a weighted F1-score to account for the minority (PDAC) class along with accuracy. Table 1 shows the 4-fold cross-validation results of the missed pancreatic cancer detection with two different CNN models and three separate loss functions. We observe that the ResNeSt improved the accuracy by 0.64% and weighted F1-score by 2% from baseline AlexNet. The LDAM-DRW loss function improves the accuracy with ResNeSt, unlike with AlexNet, where it performs the same for all loss functions due to its model capacity. Further, the ResNeSt-LDAM-DRW setup improves the weighted F1-score compared to LDAM and binary cross-entropy (BCE) loss in conjunction with AlexNet and ResNeSt.

Conclusions: Our approach yields a notable performance improvement compared to the baseline AlexNet. It is the first work in the literature to investigate the missed diagnosis of PDAC, to the best of our knowledge. Furthermore, comprehensive evaluations and explanations for this missed cancer diagnosis detection would help understand the clinical problem, and we leave this as a direction for future work.

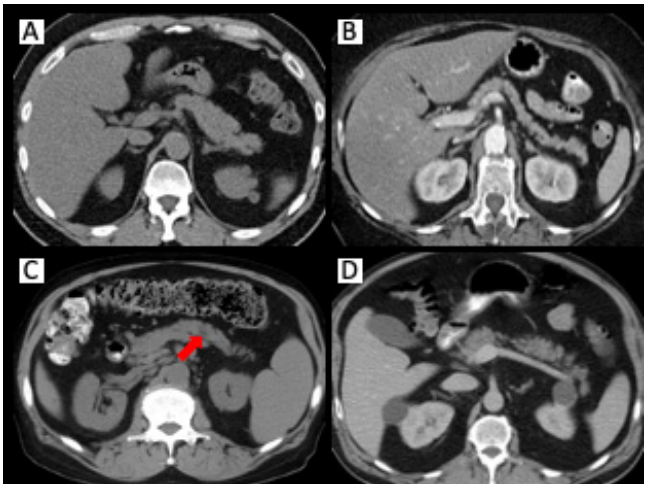

Figure 1: Sample axial view of the pancreas in CT images. (A) Control CT image without contrast. (B) Control CT image with contrast. (C) Missed PDAC CT image without contrast and the lesion is present in the pancreatic body. (D) Missed PDAC CT image with contrast.

Model	Loss function	Accuracy (%)	Weighted F1-score (%)
ResNeSt	BCE	93.40	91.05
	LDAM	93.37	90.99
	LDAM-DRW	93.39	91.28
AlexNet	BCE	92.75	89.28
	LDAM	92.75	89.28
	LDAM-DRW	92.75	89.28

Table 1: Accuracy and weighted F1-score over 4-folds cross-validation using ResNeSt and AlexNet with different loss functions: binary cross-entropy (BCE) loss, label-distribution-aware margin (LDAM) loss, and its combination with deferred re-weighting (DRW).