The American Journal of Pathology, Vol. 194, No. 12, December 2024

ELSEVIER

The American Journal of PATHOLOGY

ajp.amjpathol.org

MACHINE LEARNING, COMPUTATIONAL PATHOLOGY, AND BIOPHYSICAL IMAGING

A Deep Learning Approach for the Identification of the Molecular Subtypes of Pancreatic Ductal Adenocarcinoma Based on Whole Slide Pathology Images

Check for updates

Pouya Ahmadvand,* Hossein Farahani,* David Farnell,[†‡] Amirali Darbandsari,[§] James Topham,[¶] Joanna Karasinska,[¶] Jessica Nelson,[‖] Julia Naso,[†] Steven J.M. Jones,[**] Daniel Renouf,[††] David F. Schaeffer,[†‡¶] and Ali Bashashati*[†]

From the School of Biomedical Engineering, and the Departments of Pathology and Laboratory Medicine,[†] Electrical and Computer Engineering,[§] and Medicine,[††] University of British Columbia, Vancouver, British Columbia; the Vancouver General Hospital, Vancouver,[‡] British Columbia; and the Pancreas Centre BC,[¶] and the Michael Smith Genome Sciences Center,[**] British Columbia Cancer Research Center,[‖] Vancouver, British Columbia, Canada*

Accepted for publication August 19, 2024.

Address correspondence to Ali Bashashati, Ph.D., or David Schaeffer, M.D., Biomedical Research Center (BRC), 2222 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada. E-mail: ali.bashashati@ubc.ca or david.schaeffer@vch.ca.

Delayed diagnosis and treatment resistance result in high pancreatic ductal adenocarcinoma (PDAC) mortality rates. Identifying molecular subtypes can improve treatment, but current methods are costly and time-consuming. In this study, deep learning models were used to identify histologic features that classify PDAC molecular subtypes based on routine hematoxylin-eosin–stained histopathologic slides. A total of 97 histopathology slides associated with resectable PDAC from The Cancer Genome Atlas project were used to train a deep learning model and test the performance on 44 needle biopsy material (110 slides) from a local annotated patient cohort. The model achieved balanced accuracy of 96.19% and 83.03% in identifying the classical and basal subtypes of PDAC in The Cancer Genome Atlas and the local cohort, respectively. This study provides a promising method to cost-effectively and rapidly classify PDAC molecular subtypes based on routine hematoxylin-eosin–stained slides, potentially leading to more effective clinical management of this disease. *(Am J Pathol 2024, 194: 2302–2312; https://doi.org/10.1016/j.ajpath.2024.08.006)*

Pancreatic ductal adenocarcinomas (PDACs) have recently surpassed breast cancer to become the third leading cause of cancer death in Canada and the United States (*https://www.cancer.net/cancer-types/pancreatic-cancer/statistics* and *http://cancer.ca/Canadian-Cancer-Statistics-2019-EN*, both last accessed October 23, 2023). Surgery can cure approximately one-fifth of PDAC cases, if detected early. In spite of surgical intervention, the 5-year survival rate remains at 20%.[1,2] Approximately 80% of patients have metastatic disease at diagnosis, and most of these patients succumb to the disease within a year[3] because of a lack of clinically relevant biomarkers that can predict treatment response, late disease presentation, chemoresistance, and limited therapeutic options remain major challenges in the clinical management of PDAC.

The discovery of molecular subtypes of PDAC[4–6] has facilitated efforts to identify clinically relevant biomarker signatures to stratify patients into different prognostic subgroups. However, agreement on suggested subgroups for clinical trials in PDAC has been challenging, with each proposed subtyping schema using independent PDAC cohorts, leading to uncertainty in robustness, generalizability, and utility value.

Rashid et al[7] conducted a systematic analysis and introduced a robust single-sample classifier [Purity Independent Subtyping of Tumors (PurIST)] to predict subtypes in patients with

Supported by Canadian Institutes of Health Research number 418734 (A.B.); the Natural Sciences and Engineering Research Council of Canada RGPIN-2019-04896 (A.B.); the Michael Smith Foundation for Health Research (A.B.); the BC (British Columbia) Cancer Foundation (A.B. and D.F.S.); and the VGH (Vancouver General Hospital) and UBC (University of British Columbia) Hospital Foundation (A.B. and D.F.S.).

P.A., H.F., and D.F. contributed equally to this work.

Copyright © 2024 American Society for Investigative Pathology. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (*http://creativecommons.org/licenses/by-nc-nd/4.0*).
https://doi.org/10.1016/j.ajpath.2024.08.006

PDAC. PurIST eliminates the need for next-generation sequencing technology by being compatible with NanoString platforms. The classifier was trained using k-top scoring pairs and demonstrated robust performance compared with consensus clustering results. The PurIST model identified significant differences in overall survival between basal-like and classical samples and showed basal-like samples were not responsive to combination of chemotherapy drugs folinic acid, fluorouracil, irinotecan, and oxaliplatin (FOLFIRINOX) treatment. These findings suggest potential clinical decision-making applications for the proposed model.

Despite this, the clinical assessment of PDAC molecular subtypes typically requires labor-intensive and costly assays with extended turnaround times, constraining their practical use. As an alternative, histopathology provides a convenient visual readout of disease biology and provides an inexpensive view of the aggregate effects of molecular alterations on cancer cells and may serve as a surrogate marker for molecular subtypes by examining tumor tissue on routinely prepared hematoxylin-eosin (H&E)–stained slides[8] from both resection specimens as well as biopsies of metastatic disease. Two recent studies have explained some relationship between gland formation as well as neutrophil infiltration and the molecular subtypes of PDAC.[8,9]

With advances in artificial intelligence in the past decade and the increased capacity to digitize histopathology slides, modern machine learning methods have been successful in the analysis of histopathology images, despite the requirement for a large amount of labeled data.[10–13] These models have demonstrated their potential for a wide range of applications, including cancer diagnosis,[13–16] prognostication, and predicting genomic and molecular subtypes.

Within the context of pancreatic cancer, Nimgaonkar et al[17] described a machine learning–generated histologic signature that associates with outcomes following adjuvant gemcitabine, whereas this signature does not correlate with the previously developed transcriptomic classification systems.[17] Furthermore, in another series, Saillard et al[18] proposed a deep learning framework to identify the molecular subtypes of PDAC from histopathology images associated with formalin-fixed, paraffin-embedded tissue in resected and needle biopsies and reported an area under the curve ranging from 0.81 to 0.86.

The current study presents the performance of several deep learning–based models to identify the molecular subtypes of PDAC. Although the proposed model was trained on formalin-fixed, paraffin-embedded tissue, this study showed, for the first time, the utility of the proposed models in needle biopsies of frozen sections of PDAC with positive predictive values of 1.0 and a negative predictive value of 0.85.

Materials and Methods

Data Acquisition and Cohort Construction

This study used whole section slides associated with two separate cohorts of patients with PDAC along with their PurIST molecular subtypes. More specifically, the first cohort contained the histopathology slides from The Cancer Genome Atlas project [under The Cancer Genome Atlas–Pancreatic Adenocarcinoma (TCGA-PAAD) project],[19] which was used for training and validation of the proposed model. This data set was composed of 97 whole slide images (scanned at a 40× objective magnification) from 95 patients, and the molecular subtypes were obtained using the PurIST algorithm. The breakdown of the subtypes is shown in Table 1: 19 basal-like slides (18 patients) and 78 classical slides (77 patients). A board-certified pathologist (D.F.) blinded to molecular subtypes annotated the tumor and benign regions of all the slides in this data set. This data set was obtained from the Genomic Data Commons portal (*https://portal.gdc.cancer.gov*, last accessed October 23, 2023).

The external data set consisted of 110 whole slide images, obtained from 44 patients from the authors' own institution. All these cases were from patients with nonresectable PDAC who were enrolled in the Personalized Oncogenomics (NCT02155621; *https://clinicaltrials.gov/study/NCT02155621*) and/or PanGen (NCT02869802; *https://clinicaltrials.gov/study/NCT02869802*) trials at British Columbia Cancer. These slides comprised 33 basal-like and 77 classical PDACs and were used for testing the proposed histopathology-based molecular classifier. The data set included frozen section needle biopsy tissue samples from different organs, including the liver and pancreas. These slides were not annotated by any pathologist and were solely used for testing the classifier. All experiments were conducted in accordance with the Declaration of Helsinki and the International Ethical Guidelines for Biomedical Research Involving Human Subjects. Anonymized archival tissue samples were retrieved and digitized after approval by the institutional Research Ethics Board (REB number H18-03646).

Quality Control and Color Normalization

HistoQC[20] was used to identify and exclude areas of the slide images that contained artifacts (eg, blurry regions, pen marks, and white background). Afterwards, to ensure consistency in the color distribution of both data sets, a color normalization strategy was used, as described in the study by Boschman et al.[21] Three reference images from the training data set that represented a broad range of colors were selected. Each image was then randomly normalized with one of the reference images using the methods of Reinhard et al,[22] Vahadane et al,[23] or Macenko et al.[24] This approach was adopted as a singular reference image or color normalization method and may not be optimal for all data sets or tasks. By using a combination of normalization methods, the images become diverse enough to train a generalizable model, while also being similar enough to overcome any potential domain shift caused by differences in color profiles across data sets. This strategy aims to improve the model's generalizability and reduce the impact of color variations between data sets.

Table 1 The Details of the Two PDAC Data Sets

Data set	Unit	Subtype	
		Basal	Classical
Training and validation data set (TCGA-PAAD)	Sample type	Whole section	
	Fixation procedure	Formalin fixed	
	Magnification	40.0	
	Patients	18	77
	Slides	19	78
	Benign annotations	25	50
	Tumor annotations	129	231
	256 × 256-Pixel benign patches at 5× magnification	990	4670
	256 × 256-Pixel tumor patches at 5× magnification	6080	18,313
External validation data set (POG-PANGEN)	Sample type	Needle biopsy	
	Fixation procedure	Fresh frozen	
	Magnification	40.0	
	Patients	9	35
	Slides	33	77
	256 × 256-Pixel tumor patches at 5× magnification	8138	14,695

These data sets were used for training the machine learning models to classify patients into prognostic subgroups based on the Purity Independent Subtyping of Tumors (PurIST) algorithm.

PAAD, pancreatic adenocarcinoma; PANGEN, Prospectively Defining Metastatic Pancreatic Ductal Adenocarcinoma Subtypes by Comprehensive Genomic Analysis; PDAC, pancreatic ductal adenocarcinoma; POG, Personalized Oncogenomics; TCGA, The Cancer Genome Atlas.

Deep Learning Models for Classification

A two-stage classification strategy was proposed to identify the PurIST molecular subtype of PDAC (Figure 1). In the first stage, a deep learning–based model was trained to identify regions of the slide images that contains tumor. In the second stage, another deep learning–based classifier was trained to determine the PurIST subtype of a given patient based on the identified tumor regions. Each classifier is trained separately.

At the first step, 2048 × 2048-pixel patches on 40× magnification were extracted from the intersection of generated HistoQC masks and the annotated regions of the train and validation data set slides. Then, the acquired high-resolution patches were resized to 256 × 256 pixel to provide patches on 5× magnification. After the patch extraction, 24,393 tumor patches and 5660 benign patches were extracted from the annotated regions (Table 1).

In a threefold cross-validation scheme, the tumor and benign patches (5500 patches each) were randomly selected and grouped into three subgroups: training (66%), validation (17%), and testing (17%). Then, a deep learning classifier was trained to differentiate between the benign and tumor regions at the patch level. More specifically, Resnet34[25] model pretrained on ImageNet[26] was used as the backbone model for this classifier. The last fully connected layer was replaced with two nodes. During fine-tuning, data augmentation techniques, such as rotating, flipping, and cropping, were applied on the patches. The model was then trained for five epochs with a batch size of eight, a learning rate of 1×10^{-5}, and a weight decay of 5×10^{-4}. An RTX3090 GPU with 24 GB GRAM served as the powerhouse for both training and testing purposes. Table 2 and Supplemental Table 1 show the performance of the tumor versus benign patch classifier.

To classify patients based on their PurIST molecular subtypes, a 10-fold cross-validation scheme was applied to ensure fair and robust evaluation at the patient level. Each subset was then further divided into training (approximately 57%), validation (approximately 16%), and testing (approximately 27%) splits (Supplemental Table 2). Four deep learning architectures [namely Vanilla,[21,27] Iterative Draw and Rank Sampling (IDaRS),[28] and attention-based models, including Deep Multiple Instance Learning (DeepMIL)[29] and Variability-aware Multiple Instance Learning (VarMIL)],[30] were used to train classifiers for stratifying patients into PurIST molecular subtypes. A brief overview of these models is provided below.

Vanilla

This approach is a widely used method in digital pathology for classifying cancer subtypes in patients. It simplifies weakly supervised classification tasks into fully supervised problems. In this approach, each patch is assumed to have the same label as its originating slide. The extracted patches from annotated regions are fed into the model, and its weights are updated until it accurately predicts the patch labels. A simple majority vote is used in Vanilla to aggregate results, with all patches carrying equal weight.

Although the Vanilla approach is simple and effective for classifying cancer subtypes, it has limitations, such as its inability to capture spatial relationships between patches. Additionally, the simple majority voting

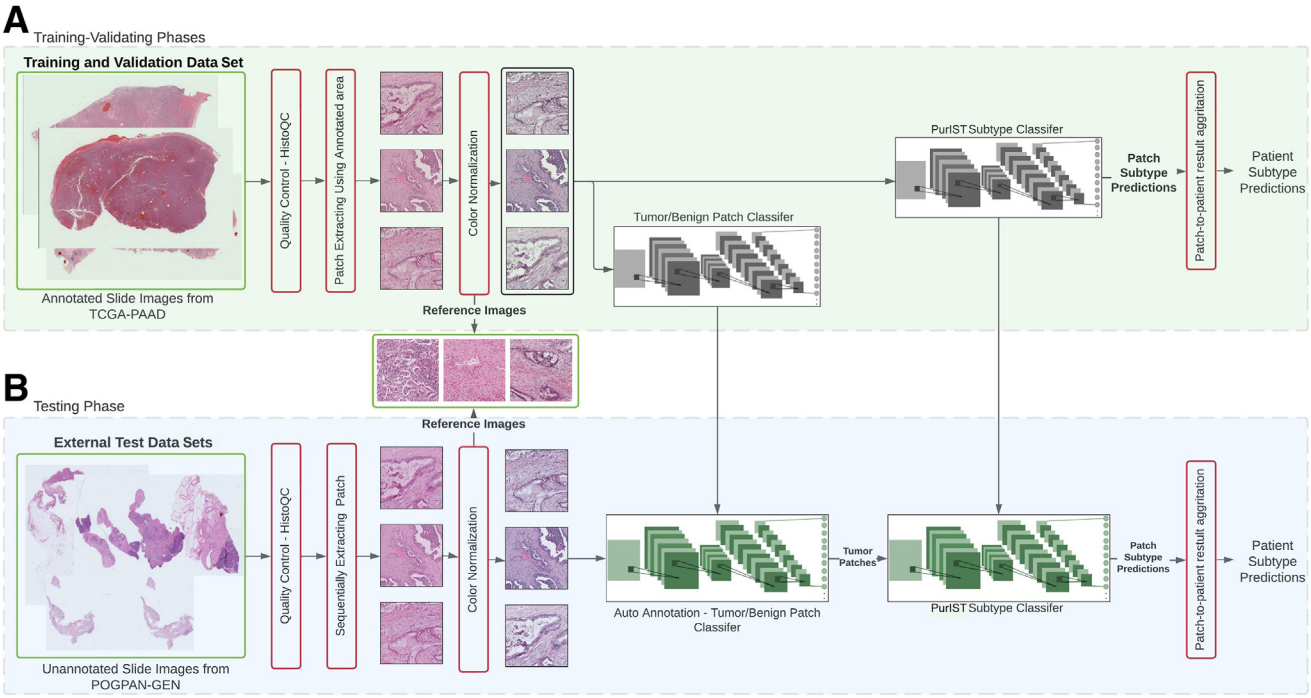

Figure 1 The proposed pipeline for training, validating, and testing a model aimed at classifying pancreatic ductal adenocarcinoma slides into Purity Independent Subtyping of Tumors (PurIST) molecular subtypes. This pipeline consists of two primary sections. The green section represents the necessary steps for training and validating the model (**A**), whereas the blue section outlines the steps for deploying and using the trained model to classify unannotated slides (**B**). During the training phase, annotated patches from the training set are used to train two classifiers separately. POG, Personalized Oncogenomics; TCGA, The Cancer Genome Atlas.

technique may not work well in cases where the tumor area is small or unevenly distributed in the slide. Despite these limitations, the Vanilla approach remains a popular and valuable tool in digital pathology for identifying cancer subtypes in patients.

IDaRS

Unlike the Vanilla approach, which assumes all patches extracted from the tissue slide areas are cancerous and passed to the model during the training phase, the IDaRS approach introduces a selection procedure to include only informative patches during the training phase. The model is trained in a fully supervised manner, and slide labels are assigned to corresponding extracted patches. The IDaRS model features a rank and sampling module responsible for selecting k top patches that the model has the highest confidence in predicting their labels, along with r random patches from the slide. This selection process enables the model to learn more robust features and prevent overfitting.

Attention-Based Models

The attention mechanism, introduced in 2014, guides the model to focus on specific image areas to perform a machine-learning task (eg, detecting an object in the image). Several studies have recently used attention mechanisms to focus on discriminative and informative patches in histopathology images. In these models, attention layers and slide representation act as result aggregators and determine slide labels. Attention layer weights specify which patches are more informative in

Table 2 The Tumor/Benign Classifier Performance of the Three Different Splitting Schemes

Split no.	Benign accuracy, %	Tumor accuracy , %	κ Value	F1 score	Area under the curve	Average accuracy, %
1	92.07	90.79	0.8288	0.9144	0.9757	91.43
2	85.88	94.97	0.8176	0.9087	0.9687	90.42
3	79.79	91.27	0.7190	0.8594	0.9410	85.53
Average, means ± SD, %	85.91 ± 6.14	92.34 ± 2.29	78.85 ± 6.04	89.42 ± 3.02	96.18 ± 1.84	89.13 ± 3.16

Table 3 The Average of the Models' Statistics on the 10-Fold Cross-Validation Splits

Data set	Model	Basal, %	Classical, %	Patients' κ value	Patients' F1 score	Patients' AUC	Average patients' accuracy, %
Testing split from the training and validation data set	Vanilla	100	92.37 ± 4.9	0.84 ± 0.09	0.92 ± 0.05	0.99 ± 0.01	96.19 ± 2.45
	IDaRS	96.67 ± 10.54	96.12 ± 3.34	0.89 ± 0.13	0.94 ± 0.06	0.97 ± 0.06	96.4 ± 5.98
	DeepMIL	94.17 ± 12.45	88.85 ± 16.35	0.77 ± 0.24	0.87 ± 0.14	0.92 ± 0.09	91.51 ± 9.05
	VarMIL	97.5 ± 7.91	97.49 ± 4.27	0.92 ± 0.1	0.96 ± 0.05	0.98 ± 0.04	97.5 ± 4.09

Data are means ± SD.

AUC, area under the curve; DeepMIL, Deep Multiple Instance Learning; IDaRS, Iterative Draw and Rank Sampling; VarMIL, Variability-aware Multiple Instance Learning.

predicting slide labels. Two attention-based models were used in this study, DeepMIL and VarMIL.

DeepMIL

This model uses visual attention and weakly supervised learning to classify. It uses a multilayer perceptron as the attention layer, assigning weights to extracted patches and computing the slide representation vector. The slide representation vector is then input for the final stage, where a classifier predicts the molecular subtype. DeepMIL is the first model to combine visual attention and weakly supervised learning for classifying histopathology data sets.

VarMIL

VarMIL improves slide representation accuracy by incorporating an attention-weighted variance module into the attention layer. By adding this module, the model accounts for variability within regions of interest, thus increasing classification performance. This approach can also help model tumor heterogeneity, or the diversity of cancer cells within a tumor. By accounting for the variability within regions of interest, the model can better understand the different types of cancer cells present in tissue samples and make more accurate predictions.

For the Vanilla model, VGG-19[31] with batch normalization,[32] pretrained on ImageNet, was deployed as the model backbone for feature extraction. The last layer of the model was replaced with a fully connected layer consisting of two neurons, corresponding to the number of cancer subtypes. The network was trained for five epochs with a learning rate of 1×10^{-5} and a weight decay of 5×10^{-4}. In the IDaRS network, Resnet50[25] was used as the backbone model. In this architecture, k top patches with the highest model's confidence and r random patches from each slide are selected for contributing to the training. The model trained by using $k = 45$ and $r = 5$ to sample 50 patches from each slide. The learning rate and other hyperparameters were adopted from Vanilla's settings. For the DeepMIL model, ResNet50 was used as the feature extractor and deployed a three-layer multilayer perceptron with two neurons in the last layer as the classifier to determine the cancer subtype. The models were trained for 10 epochs. Similarly, in the VarMIL model, ResNet50 extracted features and incorporated a three-layer multilayer perceptron with two neurons in the last layer as the classifier. The model was trained for 10 epochs.

The deep learning codebase developed for this study is publicly available (*https://github.com/AIMLab-UBC/PDAC-PathAI-MolecularClassifier*, last accessed August 16, 2024).

Results

The regions of the slide image containing tumors were initially identified through the training of a deep learning classifier. Table 2 shows a summary of this classifier's performance on the held-out split of the training and validation data set. On average, this classifier achieved mean area under the curve of 96.18% and accuracy of 89.13% in identifying the tumor regions in the slide images.

Classifiers were additionally trained for the identification of PurIST molecular subtypes based on the recognized tumor patches. The resultant patch-level labels were aggregated to establish the patient-level molecular subtype. Table 3 summarizes the average performance of the models across the 10 cross-validation splits on TCGA-PAAD data set. For more detailed results on the Vanilla, IDaRS, DeepMIL,[29] and VarMIL models, please refer to Supplemental Tables 3 to 6, which provide comprehensive statistics over the 10-fold cross-validation splits. The VarMIL model outperformed the other models, achieving an average patient accuracy of 97.5% ± 4.09% over the 10-fold cross-validation splits.

Subsequently, the tumor/normal classifier was used on the held-out external data set to detect tumor regions and then applied the deep learning molecular classifiers on the resulting tumor patches in the external data set (Table 4). The Vanilla and IDaRS models outperformed the attention-based models, with Vanilla achieving the best accuracy among all models, with an average patient accuracy of 83.03% ± 6.35%. The VarMIL model had the best performance on the held-out training data set but performed poorly on this data set. This might be because attention-based and semisupervised learning models require large data sets for effective training. For detailed statistics of the models' performances, see Supplemental Tables 7 to 10.

Table 4 The Average of the Models' Performance on the External Data Set

Data set	Model	Basal, %	Classical, %	Patients' κ value	Patients' F1 score	Patients' AUC	Average patients' accuracy, %
External test	Vanilla	88.75 ± 10.94	77.31 ± 10.32	0.549 ± 0.12	0.766 ± 0.07	0.867 ± 0.03	83.03 ± 6.35
	IDaRS	78.89 ± 15.22	80.74 ± 3.41	0.53 ± 0.12	0.76 ± 0.06	0.78 ± 0.09	79.82 ± 7.68
	DeepMIL	76.25 ± 20.79	61.72 ± 14.21	0.27 ± 0.03	0.6 ± 0.03	0.7 ± 0.04	68.99 ± 3.89
	VarMIL	82.5 ± 26.48	46.55 ± 25.25	0.19 ± 0.07	0.51 ± 0.09	0.68 ± 0.05	64.53 ± 3.01

Data are means ± SD.

AUC, area under the curve; DeepMIL, Deep Multiple Instance Learning; IDaRS, Iterative Draw and Rank Sampling; VarMIL, Variability-aware Multiple Instance Learning.

Table 5 provides detailed information about the positive and negative predictive values associated with the Vanilla model. The model achieved a sensitivity of 0.85 and a specificity of 1.0, meaning that it correctly identified 85% of the basal-like samples and all of the classical samples. The model's positive predictive value was 1.0, indicating that when it predicted a sample was positive for the basal-like PDAC subtype, it was correct 100% of the time. The negative predictive value was 0.85, indicating that when the model predicted a sample was negative for the basal-like PDAC subtype, it was correct 85% of the time. These results suggest that the Vanilla model is highly accurate in predicting the presence or absence of the basal-like PDAC subtype.

Visualization techniques were subsequently used to emphasize the identified and detected morphologic features of the Vanilla model. Grad-CAM activation maps[33] were generated for slides from two data sets—TCGA-PAAD and the external data sets—to illustrate the regions that contributed to subtype prediction. Figure 2 displays the Grad-CAM activation maps for each subtype and data set. The regions highlighted in red in the activation maps indicate the areas that were crucial in making the subtype prediction. On close inspection of the maps, it can be observed that gland-forming and non−gland-forming areas are present, identified as the morphologic features contributing to the classification of basal-like and classical subtypes, respectively, by the Vanilla model. This suggests that the model has effectively captured and exploited the morphologic characteristics that distinguish between these subtypes.

Table 5 Confusion Matrix of the Best Split of the Vanilla Model in Predicting PDAC Subtype

Variable	Basal-like (+)	Classical (−)	
Positive test results	TP	FP	PPV
	8	0	1.0
Negative test results	FN	TN	NPV
	4	22	0.85
	Sensitivity: 0.85	Specificity: 1.0	

Evaluated by TP, FP, FN, TN, PPV, NPV, sensitivity, and specificity.

FN, false negative; FP, false positive; NPV, negative predictive value; PDAC, pancreatic ductal adenocarcinoma; PPV, positive predictive value; TN, true negative; TP, true positive.

The same analysis was further conducted on the external data set. Figure 3 displays heat maps and Grad-CAMs of two biopsied tissue slides, overlaid for enhanced visualization. Following a comprehensive review by a board-certified pathologist, it was confirmed that the model picked up the same gland-forming and non−gland-forming morphologies for the classification of basal-like and classical molecular subtypes.

Subsequently, instances where the correct molecular subtypes were not predicted by the model were investigated to understand the underlying reasons for the model's error. Table 6 shows the cases in which the model misclassified the basal-like samples. All tissue samples were obtained from the liver, with the PurIST algorithm confidently classifying patients into the classical subtype. On close review of these cases, it was observed that challenges in precisely detecting tumor regions on certain slides were encountered by the tumor/benign classifier used in the analysis, as illustrated by the generated Grad-CAM activation maps. This issue is apparent in Figure 4, which presents an overlaid Grad-CAM activation map on one of the misclassified slides. In this instance, the classifier erroneously identified numerous patches as tumor tissue instead of background benign and, consequently, benign patches were misclassified as basal-like.

Discussion

Over the past 10 years, genomic research using sequencing technologies has shown wide diversity in genomic and transcriptomic profiles of PDAC, with some of these variants having therapeutic value.[34,35] For example, RNA sequencing-based molecular subtypes of metastatic PDAC correlate with differential response to first-line chemotherapy, where patients with basal-like tumors may benefit from GEMABR as opposed to FOLFIRINOX.[36] Despite these significant advancements, the widespread implementation of genomics-guided treatment for patients with PDAC is hindered by limited accessibility, high cost, and long test turnaround times. As a result, molecular profiling is typically only available to a subset of patients in large cancer centers.

The aggressiveness of PDAC poses a formidable challenge when it comes to using genomic sequencing for patient care. The disease's rapid clinical deterioration demands swift action in identifying eligible individuals for targeted therapies and clinical

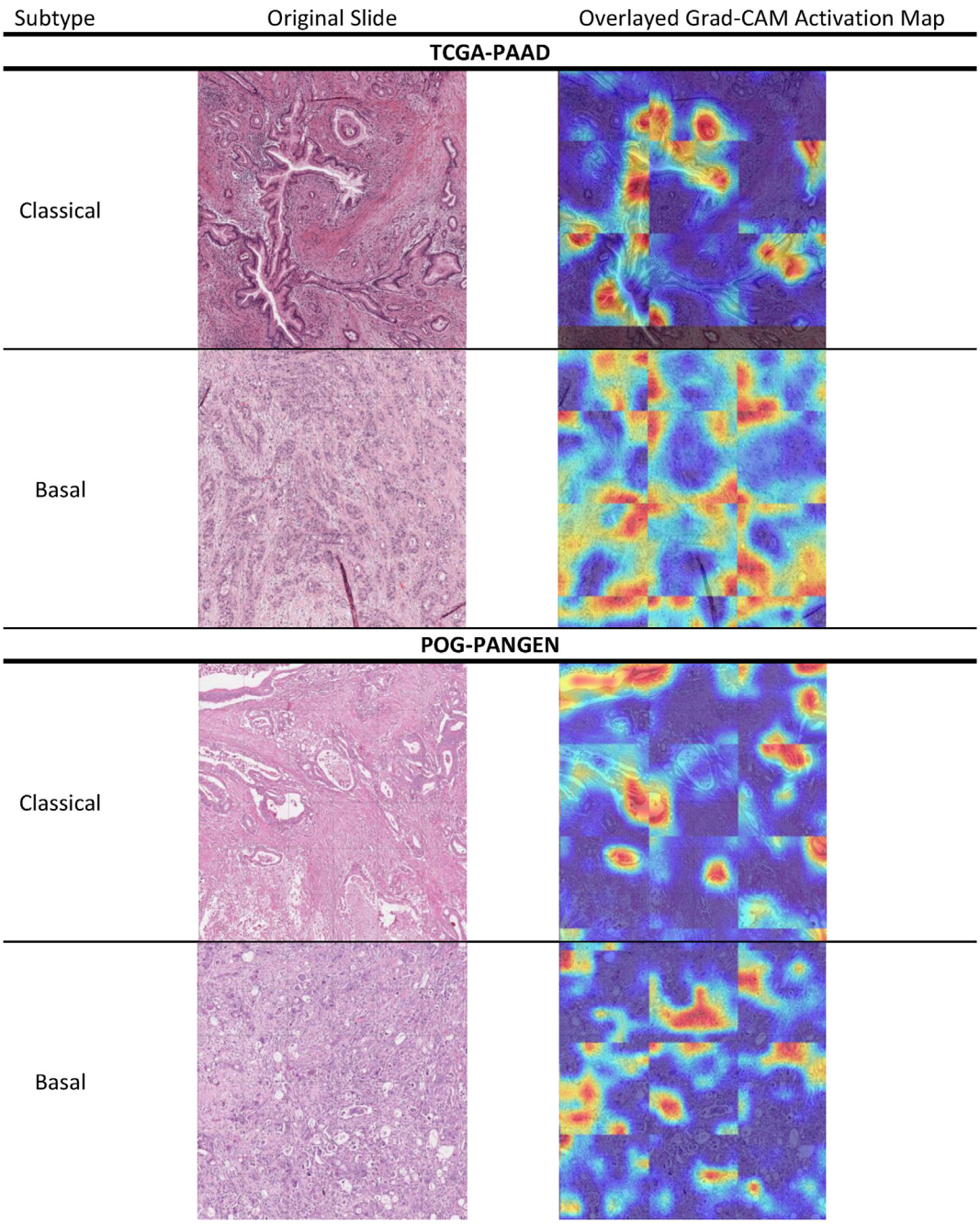

Figure 2 The Grad-CAM activation maps overlaid on the slides from the training and validation data set. For the classical case, the model has selected the areas that contain gland-forming components. PAAD, pancreatic adenocarcinoma; PANGEN, Prospectively Defining Metastatic Pancreatic Ductal Adenocarcinoma Subtypes by Comprehensive Genomic Analysis; POG, Personalized Oncogenomics; TCGA, The Cancer Genome Atlas.

trials. However, the current turnaround times for molecular profiling, which range from 19 to 52 days[36] from the time of biopsy, fall short of meeting these time-sensitive demands.

In this study, the primary goal was to develop a deep learning–based framework to facilitate the identification of PDAC molecular subtypes from H&E-stained biopsy tissue slides. In contrast to molecular profiling and genomic sequencing, H&E staining is a cost-effective and widely available technique that is routinely performed in pathology laboratories for diagnostics and prognostication with fast turnaround times. Therefore, building artificial intelligence algorithms for the identification of molecular subtypes of PDAC from H&E-stained slide images provides a fast, inexpensive, and reliable solution for replacing the time-consuming, costly, and sometimes inaccessible molecular tests.

To accomplish the goal, a two-stage classification workflow was generated. In the first stage, a tumor/benign classifier was trained to identify the tumor regions within a slide image.

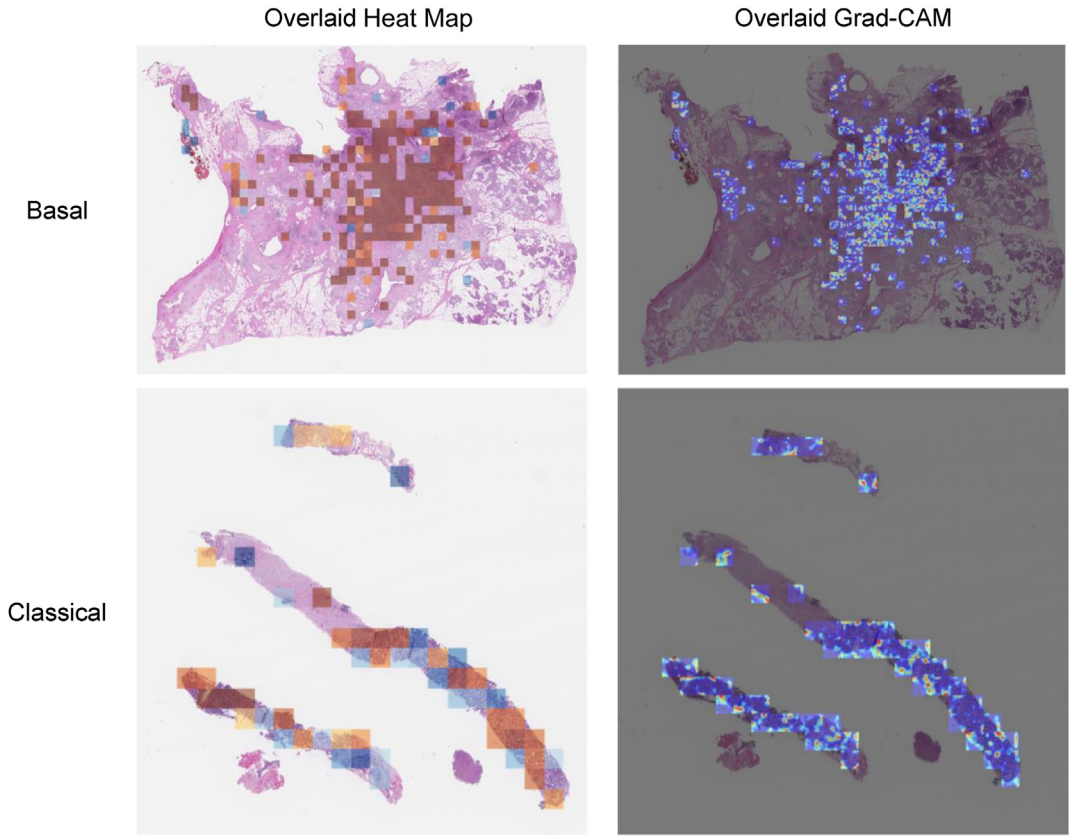

Figure 3 Overlaid heat map and activation maps on two biopsy slides. The heat map displays the regions of the slide that the model has identified as being relevant to the classification of the tissue subtype, whereas the Grad-CAM highlights the specific areas of the slide that the model has focused on to make its prediction.

Subsequently, four distinct deep learning models were trained, and their performance was assessed in a cross-validation setting, along with an external test data set of biopsy tissue from flash frozen samples. Biopsy tissue samples present unique challenges for image analysis and interpretation, as they are much smaller compared with whole-section tissue samples. Biopsy tissue samples may also contain only partial or incomplete structures, making it difficult to capture the full extent of the disease and its progression.

The proposed models achieved an average patient-level molecular subtype concordance of 83.03% $\pm$ 6.35% and an area under the curve of 0.867 $\pm$ 0.03 using a Vanilla deep learning model. This model outperformed implementations of more sophisticated strategies (namely, IDaRS, DeepMIL, and VarMIL). Balancing the subtypes during training helped prevent overfitting on any overrepresented classes. Additionally, color normalization was crucial for generating a model that can generalize across H&E-stained images processed at various centers. A color normalization approach that combines multiple normalization techniques and reference images was used to generate image sets that are sufficiently similar yet diverse enough to enhance the network's robustness.

To evaluate the model's performance, a board-certified pathologist (D.F.) reviewed the Grad-CAM activation maps and characterized gland formation patterns that were picked up by the deep learning model for differentiation between the basal-like and classical molecular subtypes. This process identified gland formation as a contributing factor to the classification.

Table 6 The List of the Misclassified Cases From the External Test Data Set

Case ID	Number of slides	Number of basal-like patches	Number of classical patches	PurIST label	PurIST score
PDAC_VS13_35184	5	1008	619	Classical	0.001
PDAC_VS15_28948	1	55	42	Classical	0.0207
PDAC_VS13_26747	8	1662	112	Classical	0.0505

ID, identifier; PDAC, pancreatic ductal adenocarcinoma; PurIST, Purity Independent Subtyping of Tumors.

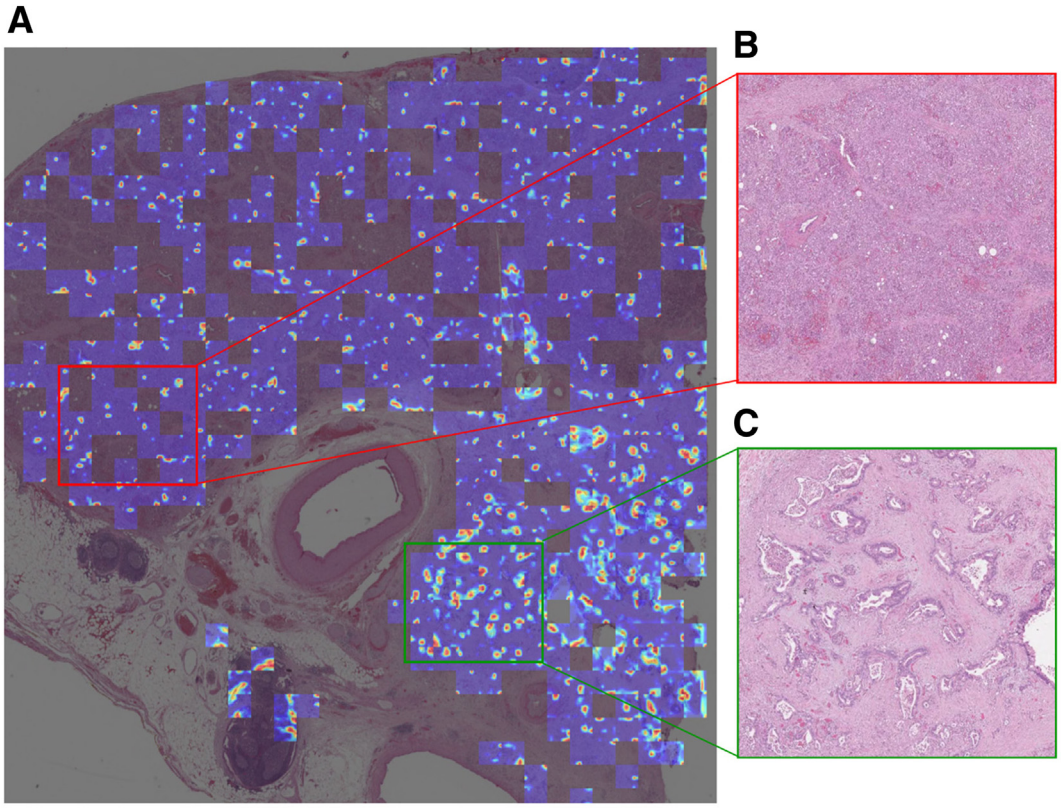

Figure 4 Screenshot from one of the misclassified cases from the external test data set. **A:** The overlaid Grad-CAM activation map on the slides reveals numerous patches erroneously labeled as tumors, which are sections of pancreatic parenchyma. In other words, the tumor/benign classifier struggled to accurately distinguish between tumor patches and pancreatic backgrounds. **B:** A view in a larger magnification of the pancreas background areas that are classified as tumor patches. **C:** A magnified view of an area that highlights the classifier's ability to detect metastatic pancreatic cancer within a lymph node, where the predominant feature is a glandular structure with minimal epithelial volume.

Conclusion

This study highlights the promise of deep learning models in the identification of the molecular subtypes of PDAC from routine histopathology slides and in light of recent advances,[18] and provides further evidence that such subtypes can be identified from frozen needle biopsy samples. A crucial aspect in future research involves assessing the clinical applicability and significance of the proposed deep learning approach. Although the model demonstrated promising performance on challenging frozen samples from needle biopsies of metastatic PDACs, it is crucial to ascertain its ability to accurately classify whole slide pathology images from diverse sources, including primary tumor tissue samples.

Author Contributions

A.B. and D.F.S. conceived the study; P.A., H.F., and A.B. designed experiments; P.A. and A.D. developed computational and visualization tools; D.F. annotated data sets; D.F. and D.F.S. provided feedback for the detected morphologic feature; P.A. performed experiments; P.A., H.F., D.F., D.F.S., and A.B. interpreted and analyzed the data; P.A. and H.F. wrote the first draft of the manuscript; S.J.M.J. advised on machine learning analysis and provided computational resources; D.F., J.T., J.K., D.R., D.F.S., and J.Na. contributed pathology expertise; and J.Ne. contributed to sample acquisition and management. All the authors critically reviewed the manuscript for important intellectual content and approved the final manuscript.

Disclosure Statement

None declared.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.ajpath.2024.08.006*.

References

1. Oettle H, Neuhaus P, Hochhaus A, Hartmann JT, Gellert K, Ridwelski K, Niedergethmann M, Zülke C, Fahlke J, Arning MB: Adjuvant chemotherapy with gemcitabine and long-term outcomes among patients with resected pancreatic cancer: the CONKO-001 randomized trial. JAMA 2013, 310:1473–1481
2. Ferrone CR, Pieretti-Vanmarcke R, Bloom JP, Zheng H, Szymonifka J, Wargo JA, Thayer SP, Lauwers GY, Deshpande V,

Mino-Kenudson M, Fernández-del Castillo C, Lillemoe KD, Warshaw AL: Pancreatic ductal adenocarcinoma: long-term survival does not equal cure. Surgery 2012, 152:S43−S49
3. Vincent A, Herman J, Schulick R, Hruban RH, Goggins M: Pancreatic cancer. Lancet 2011, 378:607−620
4. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, Smyla JK, Anderson JM, Kim HJ, Bentrem DJ, Talamonti MS, Iacobuzio-Donahue CA, Hollingsworth MA, Yeh JJ: Virtual microdissection identifies distinct tumor-and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 2015, 47:1168−1178
5. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, Feiler HS, Ko AH, Olshen AB, Danenberg KL, Tempero MA, Spellman PT, Hanahan D, Gray JW: Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011, 17:500−503
6. Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C, et al: Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016, 531:47−52
7. Rashid NU, Peng XL, Jin C, Moffitt RA, Volmar KE, Belt BA, Panni RZ, Nywening TM, Herrera SG, Moore KJ, Hennessey SG, Morrison AB, Kawalerski R, Nayyar A, Chang AE, Schmidt B, Kim HJ, Linehan DC, Yeh JJ: Purity independent subtyping of tumors (PurIST), a clinically robust, single-sample classifier for tumor subtyping in pancreatic cancer. Clin Cancer Res 2020, 26:82−92
8. Naso JR, Topham JT, Karasinska JM, Lee MK, Kalloger SE, Wong H, Nelson J, Moore RA, Mungall AJ, Jones SJ, Laskin J, Marra MA, Renoof DJ, Schaeffer DF: Tumor infiltrating neutrophils and gland formation predict overall survival and molecular subgroups in pancreatic ductal adenocarcinoma. Cancer Med 2021, 10:1155−1165
9. Kalimuthu SN, Wilson GW, Grant RC, Seto M, O'Kane G, Vajpeyi R, Notta F, Gallinger S, Chetty R: Morphological classification of pancreatic ductal adenocarcinoma that predicts molecular subtypes and correlates with clinical outcome. Gut 2020, 69:317−328
10. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH: Patch-based convolutional neural network for whole slide tissue image classification. Edited by Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. pp. 2424−2433
11. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH: Deep learning for identifying metastatic breast cancer. arXiv 2016, [Preprint], https://doi.org/10.48550/arXiv.1606.05718
12. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NN, Tomaszewski J, González FA, Madabhushi A: Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. Sci Rep 2017, 7:46450
13. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017, 318:2199−2210
14. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, Moreira AL, Razavian N, Tsirigos A: Classification and mutation prediction from non−small cell lung cancer histopathology images using deep learning. Nat Med 2018, 24:1559−1567
15. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019, 25:1301−1309
16. Farnell DA, Huntsman D, Bashashati A: The coming 15 years in gynaecological pathology: digitisation, artificial intelligence, and new technologies. Histopathology 2020, 76:171−177
17. Nimgaonkar V, Krishna V, Krishna V, Tiu E, Joshi A, Vrabac D, Bhambhvani H, Smith K, Johansen JS, Makawita S, Musher B, Mehta A, Hendifar A, Wainberg Z, Sohal D, Fountzilas C, Singhi A, Rajpurkar P, Collisson EA: Development of an artificial intelligence-derived histologic signature associated with adjuvant gemcitabine treatment outcomes in pancreatic cancer. Cell Rep Med 2023, 4:101013
18. Saillard C, Delecourt F, Schmauch B, Moindrot O, Svrcek M, Bardier-Dupas A, Emile JF, Ayadi M, Rebours V, de Mestier L, Hammel P, Neuzillet C, Bachet JB, Iovanna J, Dusetti N, Blum Y, Richard M, Kermezli Y, Paradis V, Zaslavsky M, Courtiol P, Kamoun A, Nicolle R, Cros J: Pacpaint: a histology-based deep learning model uncovers the extensive intratumor molecular heterogeneity of pancreatic adenocarcinoma. Nat Commun 2023, 14:3459
19. Gutman DA, Cobb J, Somanna D, Park Y, Wang F, Kurc T, Saltz JH, Brat DJ, Cooper LA, Kong J: Cancer digital slide archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. J Am Med Inf Assoc 2013, 20:1091−1098
20. Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A: HistoQC: an open-source quality control tool for digital pathology slides. JCO Clin Cancer Inf 2019, 3:1−7
21. Boschman J, Farahani H, Darbandsari A, Ahmadvand P, Van Spankeren A, Farnell D, Levine AB, Naso JR, Churg A, Jones SJ, Yip S, Köbel M, Huntsman DG, Gilks CB, Bashashati A: The utility of color normalization for AI-based diagnosis of hematoxylin and eosin-stained pathology images. J Pathol 2022, 256:15−24
22. Reinhard E, Adhikhmin M, Gooch B, Shirley P: Color transfer between images. IEEE Comput Graph Appl 2001, 21:34−41
23. Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, Navab N: Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans Med Imag 2016, 35:1962−1971
24. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, Schmitt C, Thomas NE: A method for normalizing histology slides for quantitative analysis. Edited by 2009 IEEE International Symposium on Biomedical Imaging: from Nano to Macro. IEEE, 2009. pp. 1107−1110
25. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. Edited by Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. pp. 770−778
26. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L: Imagenet: a large-scale hierarchical image database. Edited by 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009. pp. 248−255
27. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN: Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer 2021, 124:686−696
28. Bilal M, Raza SEA, Azam A, Graham S, Ilyas M, Cree IA, Snead D, Minhas F, Rajpoot NM: Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. Lancet Digital Health 2021, 3:e763−e772
29. Dietterich TG, Lathrop RH, Lozano-Pérez T: Solving the multiple instance problem with axis-parallel rectangles. Artif Intell 1997, 89:31−71
30. Schirris Y, Gavves E, Nederlof I, Horlings HM, Teuwen J: DeepSMILE: self-supervised heterogeneity-aware multiple instance learning for DNA damage response defect classification directly from H&E whole-slide images. arXiv 2021, [Preprint], https://doi.org/10.48550/arXiv.2107.09405
31. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. arXiv 2014, [Preprint], https://doi.org/10.48550/arXiv.1409.1556
32. Simon M, Rodner E, Denzler J: ImageNet pre-trained models with batch normalization. arXiv 2016, [Preprint], https://doi.org/10.48550/arXiv.1612.01452
33. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-cam: visual explanations from deep networks via gradient-based localization. Edited by Proceedings of the IEEE International Conference on Computer Vision; 2017. pp. 618−626
34. Hayashi H, Higashi T, Miyata T, Yamashita Y, Baba H: Recent advances in precision medicine for pancreatic ductal adenocarcinoma. Ann Gastroenterol Surg 2021, 5:457−466

35. Pishvaian MJ, Blais EM, Brody JR, Lyons E, DeArbeloa P, Hendifar A, Mikhail S, Chung V, Sahai V, Sohal DP: Overall survival in patients with pancreatic cancer receiving matched therapies following molecular profiling: a retrospective analysis of the know your tumor registry trial. Lancet Oncol 2020, 21:508−518
36. Aung KL, Fischer SE, Denroche RE, Jang G-H, Dodd A, Creighton S, Southwood B, Liang S-B, Chadwick D, Zhang A, O'Kane GM, Albaba H, Moura S, Grant RC, Miller JK, Mbabaali F, Pasternack D, Lungu IM, Bartlett JMS, Ghai S, Lemire M, Holter S, Connor AA, Moffitt RA, Yeh JJ, Timms L, Kryzanowski PM, Dhani N, Hedley D, Notta F, Wilson JM, Moore MJ, Gallinger S, Knox JJ: Genomics-driven precision medicine for advanced pancreatic cancer: early results from the COMPASS trial. Clin Cancer Res 2018, 24:1344−1354