Integrating Vision and Language Models for Interpretable Resectability Prediction of Pancreatic Ductal Adenocarcinoma

Submitted on: **08-03-2024**

Muhammad Amir Bin Mohd Azman and Leonard Bereska
amir.muhammad.amir.bin.mohd.azman@student.uva.nl
l.f.bereska@uva.nl
University of Amsterdam and Amsterdam University Medical Center
Amsterdam, The Netherlands

1 INTRODUCTION

Pancreatic Ductal Adenocarcinoma (PDAC), the predominant malignant form of pancreatic cancer, is the seventh leading cause of cancer death globally despite only being the 14th most common cancer [23]. The preoperative assessment of resectability for PDAC presents an ongoing challenge for healthcare professionals, mainly due to the risk of overestimating the tumor's resectability. Currently, surgical resection of the tumor combined with systemic therapy is the only potentially curative treatment option available for PDAC patients [28]. Despite advances in medical research and treatment methodologies, due to its aggressive nature, interobserver variability, and limited responsiveness to current treatment options, healthcare professionals are cautious about proceeding with high-risk surgeries [17, 20, 26].

As a highly heterogeneous disease, genetic mutations and diverse contributing factors make it difficult to pinpoint the potential origins of the tumor [11, 21]. However, the National Comprehensive Cancer Network (NCCN) guidelines recommend conducting a preoperative diagnostic from imaging through a multi-phase computed tomography (CT) to classify resectability [12]. Specifically, medical professionals are tasked with relying on their observations, as the determination of resectability depends on their capacity to identify factors such as the tumor's relationship with blood vessels, and the existence of metastatic disease [2]. Despite these potential limitations and risk of human error, the assessment of inter-observer agreement between healthcare professionals for both overall resectability status and specific vascular involvement has been positive as experience and personal interest in pancreatic cancer are more influential than the number of patients at a hospital or the hospital's specialty in ensuring assessment consistency [12].

Resectability is summarized into three classification categories: Resectable (RES), Borderline Resectable (BR), and Unresectable Locally Advanced (LA). Valuable noninvasive and informative information regarding the entire tumor is extracted using radiological imaging. RES PDAC is described as a tumor without contact with the superior mesenteric or common hepatic artery. In contrast, BR PDAC is a tumor that contacts less than 180 degrees of either the superior mesenteric artery or common hepatic artery. Meanwhile, LA PDAC is defined as a tumor that makes contact with more than 180 degrees of the superior mesenteric artery or common hepatic artery [12].

Hence, adopting a non-invasive, interpretable AI-powered model could serve as a transformative tool for the preoperative evaluation of PDAC while distancing itself from radiomic feature extraction. A novel standard in the accuracy and reliability of PDAC resectability predictions will be developed through the use of integrating visual data through Vision Transformers (ViT) and textual data through Large Language Models (LLM) transformer models may provide leverage in improving patient management and move away from radiomic feature extraction.

2 RELATED WORK

This section reviews existing literature on automated PDAC detection, deep learning models in PDAC research, and the emergence of transformer technology in cancer research. It highlights the novelty and necessity of the proposed approach compared to traditional methods and the potential of multimodal transformers in overcoming existing predictive limitations.

2.1 PDAC Research

2.1.1 Segmentation approach for PDAC Detection. In identifying PDAC, radiologists must evaluate various features that indicate the presence of the tumor, including a mass within the pancreas, dilated pancreatic ducts, and vascular involvement [4]. Thus, semi-supervised learning for accurate segmentation of tumors and vasculature from CT scans demonstrates the potential to enhance the observability precision of vascular involvement by leveraging radiomic features to provide a comprehensive analysis of tumor characteristics [5]. Segmentation is a process that involves dividing an image into multiple regions to simplify, providing a more meaningful representation for analysis. Therefore, the reliability of radiomic features is heavily influenced by the segmentation quality since they are quantitative measures extracted from medical images that describe tissue and lesion characteristics such as heterogeneity and shape [13]. For the model to learn effectively, it is essential that experienced radiologist refine and manually pre-segment the volume of interest in the CT scans [14]. Moreover, deep learning emerged as a promising approach for developing segmentation models in pancreatic research [8]. Following this advancement, implementing deep learning models and radiomics on automated detection of pancreatic tumors through contrast-enhanced

CT scans is possible. Subsequently, the success in segmentation created opportunities to develop additional features for PDAC detection [32].

2.1.2 *Machine learning approaches in PDAC Resectability.*

Throughout this section, the machine learning approaches demonstrate remarkable accuracy, offering the potential to deliver expert confidence in their decision. However, throughout the studies, model performances are heavily criticized due to validation availability. The study by Baecker et al. discusses potential limits in data such as the age group boundaries, reduced predictive capability of the model when excluding patient data from the last three months prior to pancreatic ductal adenocarcinoma (PDAC) diagnosis, and omission of predictive factors (eg. back pain). Nonetheless, the insights by Jia et al. show that various models, such as both a custom neural network (PrismNN) and a logistic regression (PrismLR), can sufficiently classify resectability through the use of only data from the Electronic Health Record (EHR). EHR data encompasses diverse textual and numerical information such as demographics, diagnoses, laboratory test results, medication orders, and procedural codes. On the other hand, attempts are being made to identify PDAC through only visual data. Alternatively, efficient visual deep learning networks such as ResNet and DenseNet combined with supervised learning algorithms have also shown promising results in validation studies [19]. Long Short-Term Memory network, CE-ConvLSTM proposed by Yao et al., has been utilized to address the limitations of current predictive models by predicting survival outcomes and resection margin status. AI-based models such as VasQnet and VasQNet with rule-based decision deferral (RBDD) could use AI-generated segmentations to classify tumor resectability. Additionally, the VasQNet extension can compare classification with ground-truth observations made by radiologists [6].

Moreover, the emergence of Transformers has provided insights into PDAC research. A dual approach by Dong et al. developing a texture-aware transformer that combines CNN and transformer modules demonstrates improved accuracy. The model integrates features from the CNN and transformer blocks to facilitate the fusion of cross-modality features. Thus, classifying the degree of vascular invasion and predicting overall survival times creates a possible prognosis for PDAC patients.

2.2 Transformer Research

Transformers are in contention to replace Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) due to their scalability and parallelizability [24]. Moreover, in cancer research, more comprehensive information about the tumor and its relationship with surrounding structures can be achieved due to transformers' inherent interpretability. Xia et al. discusses the increase with healthcare professionals seeking a second confirmation through an interpretable approach. Nonetheless, the evaluation of multimodal transformers remains an emerging area of medical research.

2.2.1 *Vision Transformers.*

The Vision Transformer (ViT), introduced by Dosovitskiy et al., is a model that is overtaking CNN in image recognition tasks after its success in natural language processing (NLP); the advantages include its key component of self-attention. Unlike CNNs, transformers can weigh the importance of different patches of images that are linearly embedded while attaching positional embeddings; this enables the model to capture contextual meaning. In PDAC research, a study utilizes an anatomy-guided transformer on non-contrast CT scans that accurately detects resectable pancreatic masses by distinguishing PDAC from other pancreatic abnormalities or normal pancreas [29]. The models surpass the diagnostic performance of human radiologists in the study's evaluations as enhancement of sensitivity and specificity is evident [29]. However, non-contrast CT scans may not provide as detailed information about the vascular structures, tissue differentiation, or organ perfusion as the standard contrast-enhanced scans.

2.2.2 *Large Language Model Transformers.*

LLMs can generate rich contextual embeddings for text that enable them to capture semantic information effectively. Prominent LLMs such as Generative Pre-trained Transformer (GPT) and Bidirectional Encoder Representations from Transformers (BERT) are emerging due to their adaptability and ability to handle multi-modal inputs and generalize across various domains. However, they have different applications. In medical applications, LLMs obtain the potential to be alternative expert systems when fine-tuned on domain-specific information. Therefore, LLMs have the potential to be trained with the newest medical research to facilitate improving access to medical knowledge and supporting clinical decision-making [18]. Given that the heterogeneity of the disease makes it difficult for experts to summarize the extent of the severity since LLMs can consider various variables such as EHR characteristics and medical history. Thus, LLMs are able to extract embeddings from these variables and those from EHRs [7].

2.2.3 *Multi-modal Transformers.*

Multi-modal transformers aim to leverage and integrate information from multiple data modalities, such as text and images. The findings from Xu et al. emphasize the self-attention mechanism, allowing the model to focus on relevant parts of the input from different modalities.

3 RESEARCH QUESTIONS AND OBJECTIVES

The primary research question is: How can integrating Vision and Language data enhance the interpretability and accuracy of resectability for Pancreatic Ductal Adenocarcinoma?

Sub-questions include:

- What are the optimal features that can be extracted from vision and language data?
- How does the accuracy of resectability prediction change when addressing predictive limitations in the context of scarce data?
- How can the choice of fusion strategies influence the importance of either modality?

- What are the challenges and opportunities in integrating Vision Transformer (ViT) with Language Model (LLM) Transformer for multimodal prediction?

4 HYPOTHESES AND CONTRIBUTIONS

The thesis hypothesizes that integrating ViT and LLM Transformers will significantly enhance the accuracy of PDAC resectability predictions, offering a scalable and explainable AI solution.

Explainable Artificial Intelligence (XAI) in medical research is essential as models have to be more transparent, interpretable, and understandable for experts. Thus, ViT and LLM's could excel in its explanatory abilities and provide insights into the model's decision-making process. By incorporating a wide range of information across different modalities, leveraging ViTs and LLMs in healthcare settings will allow medical professionals to benefit from improved diagnostic accuracy, personalized treatment plans, and efficient patient care. Alternative methods like deep learning can enable the acquisition of new patterns and offer alternate ways of utilizing data.

5 MATERIALS AND METHODS

The methodology involves training a ViT on imaging data, utilizing an LLM Transformer to process structured text data from health records, and combining both for multimodal prediction. This includes processing raw imaging and structured EHR data, fine-tuning pre-trained models, and employing a fusion approach for final classification.

5.1 Dataset

The study will utilize a dataset comprising 126 scans indicating locally advanced PDAC and 344 scans for resectable and borderline cases from the Amsterdam University Medical Center. Additionally, EHR Data for each patient is provided.

5.2 Software

The open-source machine learning framework used is PyTorch due to its ease of use. The libraries to test pre-trained models will be extracted from the Hugging Face library.

5.3 Vision Transformer

5.3.1 Data Augmentation. Since the data has been segmented, augmentation through rotation in the axial direction and selecting cropped regions to shift are viable options discussed in previous research. However, other augmentation techniques will be applied to maximize accuracy.

5.3.2 Data Balancing. A vital consideration throughout the vision transformer literature is the limitation on data validation. However, data handling methods such as the Difference Enhancement-Random Sampling (DERS) and nested N-fold Cross-Validation described by Jiang et al. and [9] are a viable option to overcome these limitations.

5.3.3 Model Selection. The model selection will be determined based on fine-tuning pre-trained ViT models on the medical images. The model will aim to extract relevant features that could indicate the resectability of PDAC, such as tumor size, location, and involvement of surrounding tissues or vessels. To do this, the model's final layer must be modified to output feature vectors instead of the classification labels. Popular transformer architectures are DEiT and BEiT with research shown that on medical classification, vision models pre-trained on ImageNet have seen to be effective, thus, other options such as COCO, JFT-300M, and OpenImages can be potentially used to pre-train the transformer.

Model complexity will be regulated by implementing intrinsic methods into the transformer architecture.

5.3.4 Norm regularization. L1 regularization in the transformer layers can be used to encourage sparsity in the model parameters by adding a penalty to the cost of the weight. The benefits include preventing overfitting and enhancing the model's data generalization.

5.3.5 Activation Functions. Replace traditional activation functions with laterally inhibiting activation functions such as the Softmax Linear Units (SoLUs).

5.3.6 Weight Regularization. Dropout is a regularization technique that affects the network's structure by randomly omitting the connections of a subset of neurons during training. Dropout can be applied to the transformer layers in various forms, this includes applying it in the embedding, attention, Feed-Forward Networks (FFN) or intermediate layers.

5.3.7 Loss function. Similarity in tumor cells contribute to the existence of noise, thus, the symmetric cross-entropy loss function can prevent overfitting [27]. Other potential insights and considerations during the research that could significantly benefit the model include adversarial robustness, continual learning, modularity, and compression [22].

5.4 Large Language Model

The choice of architecture will depend on the approach and the type of data that will be chosen. An encoder architecture will focus on extracting features from the EHR text. By extracting these features only, the integration with features extracted from CT scans will be feasible. Therefore, pre-trained models such as BERT are optimal for this scenario. An extension to the methodology would be to include decoders, as they efficiently generate descriptive captions for images and can transfer patient information into a comprehensive narrative. For example, GPT could generate explanations based on the trained images, which could also prove immensely beneficial for transparency. Therefore, after analysis of the EHR data, the approach can be judged based on the most optimal feature extraction for the input of the multimodal [1].

5.4.1 Data Pre-processing. Data Augmentation will be applied to encourage robustness and generalizability of the model due to the limited amount of data available.

5.5 Multimodal approach

The fusion mechanism is a core feature distinguishing multimodal transformers from unimodal models. Moreover, two multi-modal architectures differ based on the required results: the encoder and the encoder-decoder. The encoder architecture will be more significant in this research as the expected feature extraction from the ViT and LLM will be combined at the input level of the transformer. Pre-trained models such as CLIP and OWL-VIT successfully make a pair prediction of image and text [1].

5.5.1 Fusion methods. The fusion method will depend on which modality will be more influential in representation, the computational and complexity demands of each modality, and the importance of interpretability.

Sequential fusion would be a viable option as it is useful when one modality provides contextual information for another. Thus, processing one type of data first, most likely images, and then using the output to generate a caption could be significant for the final classification. On the other hand, early fusion can convert images into textual descriptions through image captioning while combining additional text inputs from the raw EHR data before processing them for classification.

6 EXPECTED OUTCOMES AND POTENTIAL CHALLENGES

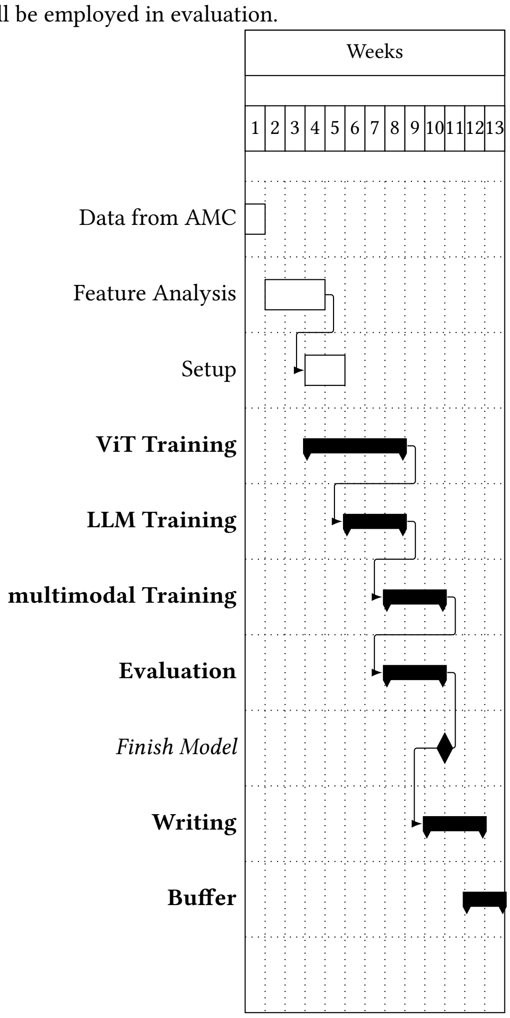
The project expects to develop a cutting-edge predictive model that leverages the integration of Vision Transformer (ViT) and Large Language Models (LLMs) for the early detection and accurate assessment of pancreatic ductal adenocarcinoma (PDAC) resectability. Additionally, the project will hope to advance in intrinsic transformer model interpretability, offering insights into decision-making processes, potentially through novel visualization techniques that map model focus areas against clinical significance. Hopefully, a substantial increase in the accuracy of PDAC resectability assessments can lead to more personalized and effective treatment plans, potentially reducing unnecessary surgeries. The main challenge is to effectively integrate visual and textual data and balancing the dataset to mitigate class sample bias.

7 RISK ASSESSMENT

The risk assessment will cover potential data and methodological challenges, with a contingency plan outlined for each identified risk. Firstly, computation and complexity of creating a vision transformer from scratch without a pre-trained model may require an abundance amount of time. Therefore to prevent time constraints, Fine-tuning a pre-trained model will be developed first. Secondly, the structure of the EHR data may result in insufficient features for the LLM to extract, therefore, open-source datasets will be considered to augment the data [25]. Alternatively, if the LLM proves to be inaccurate, the option of a Decision Tree or Reinforcement Learning to analyze the EHR data is viable. Ultimately, if the models do not fuse to make an accurate model, a pre-trained multi-modal model will be used.

8 PROJECT PLAN

A detailed project plan will include milestones for building prediction models, integrating multimodal learning, and intrinsic interpretability strategies. The overarching plan is to create a successful multimodal prediction model that can classify the resectability of PDAC through ViT and LLM. The first milestone will focus on building a vision-only pre-trained ViT, in the same time the second half will be dedicated to the non-pre-trained vision model. Once completed, the focus will shift to using a pre-trained LLM on EHR data. Alternatively, if the LLM proves to be inaccurate, the option of a Decision Tree or Reinforcement Learning to analyze the EHR data is viable. Ultimately, a fusion technique will be applied between the two models. A pre-trained multimodal will be dedicated during the weeks of "evaluation" and "finish model". The intrinsic interpretability strategies will be employed in evaluation.



REFERENCES

[1] [n. d.]. The Transformer model family. https://huggingface.co/docs/transformers/model_summary

[2] Henry E. Badgery, Tjuntu Muhlen-Schulte, John R. Zalcberg, Bianka D'souza, Jan F. Gerstenmaier, Craig Pickett, Jaswinder Samra, Daniel Croagh, Farhan Ahmed, Emel Allan, Inna Amelina, Somaiah Aroori, Nalayini Balendran, Ee J. Ban, David Bartlett, Roger Berry, Abraham Bezuidenhout, Ricky Bhogal, Candice Bolan, Louisa Bolm, Cristiana Bonifacio, Adina Borsaru, David Burnett, Nick Butterfield, Roberto Cannella, Giovanni Capretti, Carlos Carnelli, Nikolaos Chatzizacharias, Felix Chingoli, Linda Chu, Wai pong Chu, Toshimasa Clark, Saxon Connor, Stephen Currin, Jurstine Daruwalla, Domenico De Santis, Ankita Dhawan, Bruno Di Muzio, Marco Dioguardi Burgio, Joel Dunn, Georgios Gemenetzis, Mark Goodwin, Andrew Gray, Christopher Halloran, George Harisis, Heather Harris, Jin He, Elizabeth Hecht, Asmus Heumann, Russell Hodgson, Cathryn Hui, Petr Kalinin, Jorg Kleeff,

Hyun S. Ko, Ksenia Kozyreva, Numan Kutaiba, Joel Lewin, Christopher Lim, Fabian Lohöfer, Benjamin Loveday, Giovanni Marchegiani, Katerina Mastrocostas, Vasileios K. Mavroeidis, Neil Merrett, Martina Nebbia, John Neoptolemos, Denis Nesterov, Charles Pilgrim, Bradley N. Reames, Alistair Rowcroft, Melanie Seale, Sabbir Siddique, Marty Smith, Gavin Sugrue, Jatin Thakur, Rose Thomas, Walid Tibermacine, Parag Tolat, Charikleia Triantopoulou, Stanislav Trofimov, Faik G. Uzunoglu, Federica Vernuccio, Diederick W. De Boo, Laurence Webber, Alice Wei, Verity Wood, and Marc Zins. 2023. Determination of "borderline resectable" pancreatic cancer – A global assessment of 30 shades of grey. *HPB* 25, 11 (Nov. 2023), 1393–1401. https://doi.org/10.1016/j.hpb.2023.07.883

[3] Aileen Baecker, Sungjin Kim, Harvey A. Risch, Teryl K. Nuckols, Bechien U. Wu, Andrew E. Hendifar, Stephen J. Pandol, Joseph R. Pisegna, and Christie Y. Jeon. 2019. Do changes in health reveal the possibility of undiagnosed pancreatic cancer? Development of a risk-prediction model based on healthcare claims data. *PLOS ONE* 14, 6 (June 2019), e0218580. https://doi.org/10.1371/journal.pone.0218580

[4] Andrew E Becker, Yasmin G Hernandez, Harold Frucht, and Aimee L Lucas. 2014. Pancreatic ductal adenocarcinoma: Risk factors, screening, and early detection. *World Journal of Gastroenterology : WJG* 20, 32 (Aug. 2014), 11182–11198. https://doi.org/10.3748/wjg.v20.i32.11182

[5] Jacqueline I. Bereska, Boris V. Janssen, C. Yung Nio, Marnix P. M. Kop, Geert Kazemier, Olivier R. Busch, Femke Struik, Henk A. Marquering, Jaap Stoker, Marc G. Besselink, Inez M. Verpalen, and for the Pancreatobiliary and Hepatic Artificial Intelligence Research (PHAIR) consortium. 2024. Artificial intelligence for assessment of vascular involvement and tumor resectability on CT in patients with pancreatic cancer. *European Radiology Experimental* 8, 1 (Feb. 2024), 18. https://doi.org/10.1186/s41747-023-00419-9

[6] Jacqueline Isabel Bereska, Henk Marquering, Marc Besselink, Jaap Stoker, and Inez Verpalen. 2023. Improving the Reliability of Medical Diagnostic Models through Rule-Based Decision Deferral. In *Proceedings of the Inaugural 2023 Summer Symposium Series 2023*. AAAI Press, 122–126. https://researchinformation.amsterdamumc.org/en/publications/improving-the-reliability-of-medical-diagnostic-models-through-ru

[7] Sandi Besen. 2023. LLM Embeddings — Explained Simply. https://pub.aimind.so/llm-embeddings-explained-simply-f7536d3d0e4b

[8] Jinzheng Cai, Le Lu, Zizhao Zhang, Fuyong Xing, Lin Yang, and Qian Yin. 2016. Pancreas Segmentation in MRI Using Graph-Based Decision Fusion on Convolutional Neural Networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells (Eds.). Vol. 9901. Springer International Publishing, Cham, 442–450. https://doi.org/10.1007/978-3-319-46723-8_51

[9] Hexin Dong, Jiawen Yao, Yuxing Tang, Mingze Yuan, Yingda Xia, Jian Zhou, Hong Lu, Jingren Zhou, Bin Dong, Le Lu, Zaiyi Liu, Li Zhang, Yu Shi, and Ling Zhang. 2023. Improved Prognostic Prediction of Pancreatic Cancer Using Multi-phase CT by Integrating Neural Distance and Texture-Aware Transformer. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023 (Lecture Notes in Computer Science)*, Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor (Eds.). Springer Nature Switzerland, Cham, 241–251. https://doi.org/10.1007/978-3-031-43904-9_24

[10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. (2020). https://doi.org/10.48550/ARXIV.2010.11929

[11] Pierre-Olivier Frappart and Thomas G. Hofmann. 2020. Pancreatic Ductal Adenocarcinoma (PDAC) Organoids: The Shining Light at the End of the Tunnel for Drug Response Prediction and Personalized Medicine. *Cancers* 12, 10 (Oct. 2020), 2750. https://doi.org/10.3390/cancers12102750

[12] Fabio Giannone, Giovanni Capretti, Mohammed Abu Hilal, Ugo Boggi, Donata Campra, Carla Cappelli, Riccardo Casadei, Raffaele De Luca, Massimo Falconi, Gabriele Giannotti, Luca Gianotti, Roberto Girelli, Paola Gollini, Davide Ippolito, Giorgio Limerutti, Lorenzo Maganuco, Valeria Malagnino, Giuseppe Malleo, Mario Morone, Cristina Mosconi, Federica Mrakic, Diego Palumbo, Roberto Salvia, Salvatore Sgroi, Alessandro Zerbi, and Gianpaolo Balzano. 2021. Resectability of Pancreatic Cancer Is in the Eye of the Observer: A Multicenter, Blinded, Prospective Assessment of Interobserver Agreement on NCCN Resectability Status Criteria. *Annals of Surgery Open: Perspectives of Surgical History, Education, and Clinical Approaches* 2, 3 (Sept. 2021), e087. https://doi.org/10.1097/AS9.0000000000000087

[13] Yan Han, Gregory Holste, Ying Ding, Ahmed Tewfik, Yifan Peng, and Zhangyang Wang. 2022. Radiomics-Guided Global-Local Transformer for Weakly Supervised Pathology Localization in Chest X-Rays. https://doi.org/10.48550/arXiv.2207.04394 arXiv:2207.04394 [cs].

[14] Erik W Ingwersen, Jacqueline I Bereska, Alberto Balduzzi, Boris V Janssen, Marc G Besselink, Geert Kazemier, Giovanni Marchegiani, Giuseppe Malleo, Henk A Marquering, C Yung Nio, Riccardo De Robertis, Roberto Salvia, Ewout W Steyerberg, Jaap Stoker, Femke Struik, Inez M Verpalen, Freek Daams, and for the Pancreatobiliary and Hepatic Artificial Intelligence Research (PHAIR) consortium. 2023. Radiomics preoperative-Fistula Risk Score (RAD-FRS) for pancreatoduodenectomy: development and external validation. *BJS Open* 7, 5 (Sept. 2023), zrad100. https://doi.org/10.1093/bjsopen/zrad100

[15] Kai Jia, Steven Kundrot, Matvey B. Palchuk, Jeff Warnick, Kathryn Haapala, Irving D. Kaplan, Martin Rinard, and Limor Appelbaum. 2023. A pancreatic cancer risk prediction model (Prism) developed and validated on large-scale US clinical data. *eBioMedicine* 98 (Dec. 2023), 104888. https://doi.org/10.1016/j.ebiom.2023.104888

[16] Zhencun Jiang, Zhengxin Dong, Lingyang Wang, and Wenping Jiang. 2021. Method for Diagnosis of Acute Lymphoblastic Leukemia Based on ViT-CNN Ensemble Model. *Computational Intelligence and Neuroscience* 2021 (Aug. 2021), e7529893. https://doi.org/10.1155/2021/7529893

[17] Mee Joo Kang and Sun-Whe Kim. 2021. Paradigm shift for defining the resectability of pancreatic cancer. *Annals of Hepato-Biliary-Pancreatic Surgery* 25, 4 (Nov. 2021), 451–455. https://doi.org/10.14701/ahbps.2021.25.4.451

[18] Mert Karabacak and Konstantinos Margetis. [n. d.]. Embracing Large Language Models for Medical Applications: Opportunities and Challenges. *Cureus* 15, 5 ([n. d.]), e39305. https://doi.org/10.7759/cureus.39305

[19] Woohyung Lee, Hyo Jung Park, Hack-Jin Lee, Eunsung Jun, Ki Byung Song, Dae Wook Hwang, Jae Hoon Lee, Kyongmook Lim, Namkug Kim, Seung Soo Lee, Jae Ho Byun, Hyoung Jung Kim, and Song Cheol Kim. 2022. Preoperative data-based deep learning model for predicting postoperative survival in pancreatic cancer patients. *International Journal of Surgery* 105 (Sept. 2022), 106851. https://doi.org/10.1016/j.ijsu.2022.106851

[20] Luis Marti-Bonmati, Leonor Cerdá-Alberich, Alexandre Pérez-Girbés, Roberto Díaz Beveridge, Eva Montalvá Orón, Judith Pérez Rojas, and Angel Alberich-Bayarri. 2022. Pancreatic cancer, radiomics and artificial intelligence. *The British Journal of Radiology* 95, 1137 (Sept. 2022), 20220072. https://doi.org/10.1259/bjr.20220072

[21] Natalia Michalak and Ewa Małecka-Wojciesko. 2023. Modifiable Pancreatic Ductal Adenocarcinoma (PDAC) Risk Factors. *Journal of Clinical Medicine* 12, 13 (Jan. 2023), 4318. https://doi.org/10.3390/jcm12134318

[22] Tilman Räuker, Anson Ho, Stephen Casper, and Dylan Hadfield-Menell. 2023. Toward Transparent AI: A Survey on Interpreting the Inner Structures of Deep Neural Networks. https://doi.org/10.48550/arXiv.2207.13243 arXiv:2207.13243 [cs].

[23] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* 71, 3 (May 2021), 209–249. https://doi.org/10.3322/caac.21660

[24] Matteo Tiezzi, Michele Casoni, Alessandro Betti, Tommaso Guidi, Marco Gori, and Stefano Melacci. 2024. On the Resurgence of Recurrent Models for Long Sequences – Survey and Research Opportunities in the Transformer Era. https://doi.org/10.48550/arXiv.2402.08132 arXiv:2402.08132 [cs].

[25] Luke A. Torre-Healy, Ryan R. Kawalerski, Ki Oh, Lucie Chrastecka, Xianlu L. Peng, Andrew J. Aguirre, Naim U. Rashid, Jen Jen Yeh, and Richard A. Moffitt. 2023. Open-source curation of a pancreatic ductal adenocarcinoma gene expression analysis platform (pdacR) supports a two-subtype model. *Communications Biology* 6 (Feb. 2023), 163. https://doi.org/10.1038/s42003-023-04461-6

[26] Eva Versteijne, Oliver J. Gurney-Champion, Astrid Van Der Horst, Eelco Lens, M. Willemijn Kolff, Jeroen Buijsen, Gati Ebrahimi, Karen J. Neelis, Coen R. N. Rasch, Jaap Stoker, Marcel Van Herk, Arjan Bel, and Geertjan Van Tienhoven. 2017. Considerable interobserver variation in delineation of pancreatic cancer on 3DCT and 4DCT: a multi-institutional study. *Radiation Oncology* 12, 1 (Dec. 2017), 58. https://doi.org/10.1186/s13014-017-0777-0

[27] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. 2019. Symmetric Cross Entropy for Robust Learning with Noisy Labels. (2019). https://doi.org/10.48550/ARXIV.1908.06112

[28] Kongyuan Wei and Thilo Hackert. 2021. Surgical Treatment of Pancreatic Ductal Adenocarcinoma. *Cancers* 13, 8 (April 2021), 1971. https://doi.org/10.3390/cancers13081971

[29] Yingda Xia, Jiawen Yao, Le Lu, Lingyun Huang, Guotong Xie, Jing Xiao, Alan Yuille, Kai Cao, and Ling Zhang. 2021. Effective Pancreatic Cancer Screening on Non-contrast CT Scans via Anatomy-Aware Transformers. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021 (Lecture Notes in Computer Science)*, Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert (Eds.). Springer International Publishing, Cham, 259–269. https://doi.org/10.1007/978-3-030-87240-3_25

[30] Peng Xu, Xiatian Zhu, and David A. Clifton. 2022. Multimodal Learning with Transformers: A Survey. (2022). https://doi.org/10.48550/ARXIV.2206.06488

[31] Jiawen Yao, Yu Shi, Le Lu, Jing Xiao, and Ling Zhang. 2020. DeepPrognosis: Preoperative Prediction of Pancreatic Cancer Survival and Surgical Margin via Contrast-Enhanced CT Imaging. (2020). https://doi.org/10.48550/ARXIV.2008.11853

[32] Ling Zhang, Yu Shi, Jiawen Yao, Yun Bian, Kai Cao, Dakai Jin, Jing Xiao, and Le Lu. 2020. Robust Pancreatic Ductal Adenocarcinoma Segmentation with Multi-Institutional Multi-Phase Partially-Annotated CT Scans. https://doi.org/10.48550/arXiv.2008.10652 arXiv:2008.10652 [cs, eess].