Published in final edited form as:

Annu Int Conf IEEE Eng Med Biol Soc. 2022 July ; 2022: 475–479. doi:10.1109/EMBC48229.2022.9871547.

Neural Transformers for Intraductal Papillary Mucosal Neoplasms (IPMN) Classification in MRI images

F. Proietto Salanitri[1], **G. Bellitto**[1], **S. Palazzo**[1], **I. Irmakci**[2], **M. Wallace**[4], **C. Bolan**[4], **M. Engels**[4], **S. Hoogenboom**[4], **M. Aldinucci**[3], **U. Bagci**[2], **D. Giordano**[1], **C. Spampinato**[1]

[1]PeRCeiVe Lab, Department of Electrical, Electronic and Computer Engineering, University of Catania, Italy.

[2]Department of Radiology and BME, Northwestern University, Chicago, IL, USA

[3]Computer Science Department, University of Torino, Torino, Italy

[4]Mayo Clinic, Jacksonville, FL, USA

Abstract

Early detection of precancerous cysts or neoplasms, i.e., Intraductal Papillary Mucosal Neoplasms (IPMN), in pancreas is a challenging and complex task, and it may lead to a more favourable outcome. Once detected, grading IPMNs accurately is also necessary, since low-risk IPMNs can be under surveillance program, while high-risk IPMNs have to be surgically resected before they turn into cancer. Current standards (Fukuoka and others) for IPMN classification show significant intra- and inter-operator variability, beside being error-prone, making a proper diagnosis unreliable. The established progress in artificial intelligence, through the deep learning paradigm, may provide a key tool for an effective support to medical decision for pancreatic cancer. In this work, we follow this trend, by proposing a novel AI-based IPMN classifier that leverages the recent success of transformer networks in generalizing across a wide variety of tasks, including vision ones. We specifically show that our transformer-based model exploits pre-training better than standard convolutional neural networks, thus supporting the sought architectural universalism of transformers in vision, including the medical image domain and it allows for a better interpretation of the obtained results.

I. INTRODUCTION

Pancreatic cancer, also known as pancreatic ductal adeno-carcinoma (PDAC), is a growing public health issue around the world. In the United States in 2021, an estimated 60,430 new cases of pancreatic cancer will be diagnosed, with 48,220 people dying from the disease [1]. Pre-cancerous cysts or neoplasms in the pancreatic ducts are known as Intraductal Papillary Mucosal Neoplasms (IPMN) and can develop anywhere in the pancreas' ductal zone. Grading the severity of IPMNs is an important diagnosis step: most IPMNs are low-grade, and should be monitored over time; high-grade IPMNs, however, may turn into invasive cancer if left untreated. In these cases, surgery is the first choice to prevent them from

perceive@dieei.unict.it .

expanding into malignant pancreatic tumors. Therefore, there is an unmet for early detection techniques of IPMNs, in order to identify which IPMNs may lead to cancer. Automated image analysis in radiology imaging plays a key role in diagnosis, treatment and intervention of pancreas diseases; thus there is a strong potential for machine learning tools to support IPMN grade prediction that can serve better than the current radiographic standards. The most popular imaging modalities for the pancreas are computed tomography (CT) and magnetic resonance imaging (MRI). In the last few years, transformer architectures [2], [3] have raised as a valid alternative to standard convolutional networks on a variety of different tasks. More specifically, transformers enable learning arbitrary functions and consists of two main operation blocks: first, an attention-based block for modeling inter-element relations; second, a multilayer perceptron (MLP) modeling relations intra-element. A sequence of attention and MLP blocks intertwined with residual connections and normalization has showed to allow for generalization over multiple tasks. Following this trend, herein we propose an automated IPMN classifier based on transformer architecture. We, in particular, show how transformers generalize better than standard and state-of-the-art CNNs (namely, DenseNet, AlexNet, etc.) also for extremely complex tasks, as IPMN classification, while providing similar accuracy to the state of the art IPMN classification study with deep learning [4].

The major contributions of this study are the following:

1. Our work on IPMN classification is an important application contribution, which is not widely done due to the difficulty nature of the problem, and hence there is a very limited published research on this task using MRI data with deep learning. Our method can provide a significant state-of-the-art baseline to be compared with for further MRI pancreas research just before critical surgery decision or surveillance.
2. Our study contributes to the recent AI research in the strive to demonstrate architectural universalism of Transformers that can be used in a wide variety of tasks using little inductive bias, beside validating their better interpretability than CNN counterparts. To our best of knowledge, transformers have never been tried on high-risk medical diagnoses tasks before, particularly for pancreas imaging research.

II. RELATED WORK

While significant progress has been made for automated approaches to segment the pancreas and its cysts [5], the use of advanced machine learning algorithms to perform fully automatic risk-stratification of IPMNs is still limited. Some recent works, employing machine learning techniques for predicting the risk of malignancy in IPMN, have used endoscopic ultrasound (EUS) images [6], [7] yielding high accuracy of 94.0%, outperforming both human diagnosis (56%) and conventional guidelines (40–68%). CT imaging has been also adopted for investigating IPMN as in [8], [9] where low-level imaging features, such as texture, strength, and shape, have extracted from segmented cysts or pancreas for IPMN classification. Recently, deep learning methods based on standard convolutional neural networks have been proposed to diagnose IPMN from MRI scans [10], [11], [4]. Sarfaraz et. al. [10] proposed an architecture for automated IPMN classification based on feature extraction with canonical correlation using a pre-trained 3D CNN, while [11] propose a novel CNN for recognizing high grade dysplasia or cancer on MR-images, yielding promising results. Finally, Rodney et al. [4] constructed two novel ”inflated” CNN architectures, InceptINN and DenseINN, for the task of diagnosing IPMN from multisequence (T1 and T2) MRI obtaining an accuracy of about 73% in grading IPNM into three classes (no risk, low and high-risk). In this work, we employ transformers that are specific neural architectures originally proposed for machine translation tasks [2]. Transformer-based models in NLP are generally pre-trained on large corpora and then fine-tuned for the task at hand [12], [13]. Their increasing interest to vision tasks starts with Vision Transformers [3] and Detection Transformer [14]. Recently, several methods have explored transformer-based architectures for medical image analysis mainly for segmentation tasks [19], [21], [20]. However, these method employ an hybrid architecture combining both convolutions and transformers. Our approach builds upon pure vision transformers and employs a strategy similar to that one employed in NLP (as in [12], [13]), i.e., pre-training transformers on natural images and then fine-tuning them to MRI IPNM images. Experimental results show that our pre-trained transformers perform significantly better than state-of-the-art CNN classifiers.

III. METHOD

In our study, we follow the recently emerging approach of *Transformers* [2] for vision tasks. In particular, we use the ViT [3] setting, in which the encoder of the original transformer model is used on a sequence of image “patches”. However, since [3] is trained on natural images, it is necessary to adapt the input representation to be able to process MRI scans, which are instead composed by an aggregation of multiple slices, providing anatomical volumetric information.

Fig. 1 describes the proposed procedure in detail. We use T1- and T2-weighted MRI scans of the same patients in an early fusion fashion to enrich diagnostic and anatomical (localization) information. For each modality, we first sample k=9 consecutive slices and use them to create a single image, rearranging the selected slices in a $\sqrt{k} \times \sqrt{k}$ grid. k can be set differently depending on the memory availability and z-direction resolution of the MRI scan. In our experiments, we optimize this number to appreciate full anatomical information of the pancreas. The two images (one for each modality) are then concatenated along the channel dimension: the resulting tensor, of size $\sqrt{k}H \times \sqrt{k}W \times 2$, with $H \times W$ being the original size of each slice, is provided as input to the transformer. Without loss of generality, let us assume that $H = W$. As in [3], the input image is then divided into N patches of size $P \times P$, where $N = \frac{kHW}{P^2}$. As a result of this procedure, an input image $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ becomes a sequence of 2D patches $\mathbf{X}_p \in \mathbb{R}^{N \times (P \times P \times C)}$ with C being the channel dimension. The 2D patches are then flattened into vectors of size P^2C and projected to an *embedding space* of size D, obtaining a sequence of *token embeddings*. As a last pre-processing step, learnable positional encodings are summed to token embeddings, producing the actual input data sequence to the transformer. We extend the token sequence with a special *class* token, whose state at the output of the transformer describes the overall input image representation for classification purposes [3], [15], [12].

Formally, the input $\mathbf{z}_0$ to the transformer is defined as:

$$\mathbf{z}_0 = \left[\mathbf{x}_{\text{class}},\ \mathbf{x}_p^1\mathbf{E}, \ldots,\ \mathbf{x}_p^N\mathbf{E}\right] + \mathbf{E}_{\text{pos}}, \quad (1)$$

where each $\mathbf{x}_p^i \in \mathbb{R}^{P^2C}$ is a flattened patch vector, $\mathbf{E} \in \mathbb{R}^{\left(P^2C\right) \times D}$ is the embedding matrix and $\mathbf{E}_{\text{pos}} \in \mathbb{R}^{(N+1) \times D}$ is the positional encoding matrix.

The transformer encoder [2] alternates multi-head self-attention and MLP (multilayer perceptron) blocks. These blocks are then intertwined with layer normalization and residual connections (see Fig. 1), as follows:

$$\mathbf{z}'_l = \text{MSA}(\text{LN}(\mathbf{z}_{l-1})) + \mathbf{z}_{l-1}, \quad (2)$$

$$\mathbf{z}_l = \text{MLP}(\text{LN}(\mathbf{z}'_l)) + \mathbf{z}'_l, \quad (3)$$

where $l = 1 \cdots L$ identifies the transformer layer, $\text{LN}(\cdot)$ performs layer normalization, MLP represents a multilayer perceptor, and $\text{MSA}(\cdot)$ computes the standard *query-key-value* multi-head self-attention [2].

At the last transformer layer, the output embedding corresponding to *class* token is finally used for classification into 3 classes, since the MRI dataset includes normal scans, low-grade and high-grade IPMN lesions:

$$\mathbf{y} = \text{LN}\left(\mathbf{z}_L^0\right), \quad (4)$$

with $\mathbf{y}$ being the vector of output class scores.

IV. EXPERIMENTAL RESULTS

A. Dataset

We evaluate the accuracy of our proposed IPMN risk assessment method in MRI (with both T1 and T2 modalities). We use a total of 139 scans from distinct patients, retrospectively collected at Mayo Clinic [4]. Patients have either IPMN cysts detected in their pancreases or they are normal control cases selected to match the IPMN patients. Out of 139 cases, 58 (42%) were male; mean (standard deviation) age was 65.3 (11.9) years. 22% had normal pancreas; 34%, low-grade dysplasia; 14%, high-grade dysplasia; and 29%, adenocarcinoma [11]. Two expert radiologists graded the cases in a pathology report after surgery: 0) normal, 1) low-grade IPMN, and 2) high-grade IPMN. We did not consider invasive carcinoma in our analysis as they are outside the scope of IPMN risk stratification.

MRI images were resized (in the transverse plane) to 256×256 pixels. Voxel spacing of MRI scans were varying from 0.468 mm to 1.406 mm. We applied a set of preprocessing steps: N4 bias field correction followed by an edge-preserving Gaussian smoothing, and intensity standardization procedure to normalize MRI scans across patients, scanners, and time. All MRIs were performed using Siemens scanners 1.5 or 3 T (Siemens, Berlin, Germany).

The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

B. Training Procedure

We use the Vision-Transformer pre-trained on 300 million images [16] and released in [3]. During training, we fine-tune all transformers layers with the training data from the MRI dataset. MRI slices are cropped around the pancreas areas by expert physicians for all scans, and each set of 9 consecutive slices, extracted in a sliding window fashion, is arranged in a 3 × 3 grid (from top-left to bottom-right), where each cell of the grid is filled by a 64×64 MRI slice (see Fig. 1). Input MRI scans are re-oriented using the RAS axes convention and normalized, individually, between 0 and 1. Data augmentation is performed through random horizontal flipping and random 90-degrees rotation (identically applied to all slices within a grid). We minimize the cross-entropy loss with gradient descent using the Adam optimizer (learning rate: 0.003) and batch size of 8, for a total of 3000 epochs. At inference time, we classify each input MRI by feeding the sequence of 9 central slices to the model.

We employ the same training and evaluation procedure for CNN models used as baselines, i.e., DenseNet-121 [22], AlexNet [23] ResNet18 [24], EfficientNet b5 [25] and MobileNet v2 [26]. Experiments are performed on a NVIDIA RTX 3090 GPU.The proposed approach was implemented in PyTorch and MONAI; all code will be publicly released upon publication.

C. Performance

We perform 10-fold nested cross-validation in order to estimate the accuracy of the proposed approach and the methods under comparison. Results are reported in Table I, where the proposed model largely outperforms the CNN models, confirming the better generalization capabilities of transformer-based architectures compared to standard convolutional models.

We also evaluate the role of early fusion and of combining the T1 and T2 modalities, by assessing classification performance when the model receives only one modality at a time (either T1 or T2) and when performing late fusion. In this case, we train two transformer models, one for each modality, and we then concatenate the two class tokens before classification. Performance is reported in Table II that demonstrates how using T1 and T2 in an early fusion setting yields the highest performance.

It has to be noted that, comparing on the same dataset, the performance achieved by our transformer-based approach is slightly lower than those obtained in [4], i.e., about 73%. However, the architecture in [4] was specifically designed and tuned for solving the IPMN classification problem, while our transformer architecture is general, designed for natural image classification and applied directly without significant architectural changes to IPMN classification problem. This is remarkable, as we demonstrate that a general architecture performs similarly to an ad-hoc one for a complex task with limited training data.

D. Interpretability of results

Transformers allow for a more direct interpretation of its internal representations through visualizing the attention weights [3], thus supporting the sought interpretability necessary in safety-critical contexts as the medical domain one. We apply Attention Rollout [17] to track down the information propagated from the input layer to the embeddings in the higher layers. Thus, we average attention weights of all heads of each transformer layer and then multiply these averages across all layers. Fig. 3 shows some examples of interpretabilty maps in cases of correct cyst classification.

It can be observed how our transformer-based model focuses its attention mainly on cysts, thus it provides robust predictions. Conversely, CNN-based models lead to weak decisions, as their attention maps (estimated using Grad-Cam [18]) reveal that features not strictly related to cysts are used for classification (see Fig. 2, top row). Finally, although our model fails in some cases, as demonstrated by the classification accuracy in Tab. I, its attention maps often point to the correct cyst regions (see Fig. 2, bottom row); thus, the wrong prediction is due to either using directly raw data, rather than a more powerful representation, or lack of enough training data.

V. Conclusion

In this work, our overall goal was to classify pancreas (IPMN) cysts automatically. We utilized *transformers* for the first time for pancreas risk predictions and obtained promising results that can be used for MRI-based IPMN risk stratification routinely. Compared to the (few) existing methods, transformers showed higher performance overall. We found that training transformer for IPMN risk stratification is easier than conventional CNN based systems and generalizes better. Furthermore, the proposed transformer-based classifier allows for better interpretation of results than standard CNNs, revealing how it employs cues exclusively related to cysts, providing more robustness to the automated diagnosis than the comparing methods. These findings highlight the contribution that transformers can give to the future research in medical image understanding, in general, and IPMN classification, in particular, beside contributing the recent AI research efforts towards universal architectures.

ACKNOWLEDGMENT

This research was partially supported by the following grants: 1) Rehastart funded by PO FESR 2014-2020 Sicilia, Azione 1.1.5; project number 08ME6201000222, CUP: G79J18000610007); 2) the “Adaptive Brain-Derived Artificial Intelligence Methods for Event Detection” – BrAIn funded by the “Programma per la ricerca di ateneo UNICT 2020–22 linea 3”; 3) NIH grants R01-CA246704-01 and R01-CA240639-01.

References

[1]. American Cancer Society, “Cancer facts & figures,” American Cancer Society, 2021.

[2]. Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Aidan N Gomez Lukasz Kaiser, and Polosukhin Illia, “Attention is all you need,” arXiv preprint arXiv:1706.03762, 2017.

[3]. Dosovitskiy Alexey, Beyer Lucas, Kolesnikov Alexander, Weissenborn Dirk, Zhai Xiaohua, Unterthiner Thomas, Dehghani Mostafa, Minderer Matthias, Heigold Georg, Gelly Sylvain, et al. , “An image is worth 16×16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.

[4]. Rodney LaLonde Irene Tanner, Nikiforaki Katerina, Georgios Z Papadakis Pujan Kandel, Bolan Candice W, Wallace Michael B, and Bagci Ulas, “Inn: inflated neural networks for ipmn diagnosis,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019, pp. 101–109.

[5]. Zhou Yuyin, Xie Lingxi, Fishman Elliot K, and Yuille Alan L, “Deep supervision for pancreatic cyst segmentation in abdominal ct scans,” in International conference on medical image computing and computer-assisted intervention. Springer, 2017, pp. 222–230.

[6]. Kuwahara Takamichi, Hara Kazuo, Mizuno Nobumasa, Okuno Nozomi, Matsumoto Shimpei, Obata Masahiro, Kurita Yusuke, Koda Hiroki, Toriyama Kazuhiro, Onishi Sachiyo, et al. , “Usefulness of deep learning analysis for the diagnosis of malignancy in intraductal papillary mucinous neoplasms of the pancreas,” Clinical and translational gastroenterology, vol. 10, no. 5, 2019.

[7]. Gorris Myrte, Hoogenboom Sanne A, Wallace Michael B, and van Hooft Jeanin E, “Artificial intelligence for the management of pancreatic diseases,” Digestive Endoscopy, vol. 33, no. 2, pp. 231–241, 2021. [PubMed: 33065754]

[8]. Hanania Alexander N, Bantis Leonidas E, Feng Ziding, Wang Huamin, Tamm Eric P, Katz Matthew H, Maitra Anirban, and Koay Eugene J, “Quantitative imaging to evaluate malignant potential of ipmns,” Oncotarget, vol. 7, no. 52, pp. 85776, 2016. [PubMed: 27588410]

[9]. Gazit Lior, Chakraborty Jayasree, Attiyeh Marc, Langdon-Embry Liana, Allen Peter J, Do Richard KG, and Simpson Amber L, “Quantification of ct images for the classification of high-and low-risk pancreatic cysts,” in Medical Imaging 2017: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2017, vol. 10134, p. 101340X.

[10]. Hussein Sarfaraz, Kandel Pujan, Corral Juan E, Bolan Candice W, Wallace Michael B, and Bagci Ulas, “Deep multi-modal classification of intraductal papillary mucinous neoplasms (ipmn) with canonical correlation analysis,” in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 800–804.

[11]. Juan E Corral Sarfaraz Hussein, Kandel Pujan, Candice W Bolan Ulas Bagci, and Michael B Wallace, “Deep learning to classify intraductal papillary mucinous neoplasms using magnetic resonance imaging,” Pancreas, vol. 48, no. 6, pp. 805–810, 2019. [PubMed: 31210661]

[12]. Devlin Jacob, Chang Ming-Wei, Lee Kenton, and Toutanova Kristina, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.

[13]. Radford Alec, Wu Jeffrey, Child Rewon, Luan David, Amodei Dario, and Sutskever Ilya, “Language models are unsupervised multitask learners,” 2018.

[14]. Carion Nicolas, Massa Francisco, Synnaeve Gabriel, Usunier Nicolas, Kirillov Alexander, and Zagoruyko Sergey, “End-to-end object detection with transformers,” in European Conference on Computer Vision. Springer, 2020, pp. 213–229.

[15]. Touvron Hugo, Cord Matthieu, Douze Matthijs, Massa Francisco, Sablayrolles Alexandre, and Jégou Hervé, “Training data-efficient image transformers & distillation through attention,” arXiv preprint arXiv:2012.12877, 2020.

[16]. Sun Chen, Shrivastava Abhinav, Singh Saurabh, and Gupta Abhinav, “Revisiting unreasonable effectiveness of data in deep learning era,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 843–852.

[17]. Abnar Samira and Zuidema Willem, “Quantifying attention flow in transformers,” arXiv preprint arXiv:2005.00928, 2020.

[18]. Ramprasaath R Selvaraju Michael Cogswell, Das Abhishek, Vedantam Ramakrishna, Parikh Devi, and Batra Dhruv, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

[19]. Chen Jieneng and Lu Yongyi and Yu Qihang and Luo Xiangde and Adeli Ehsan and Wang Yan and Lu Le and Yuille Alan L and Zhou Yuyin “Transunet: Transformers make strong encoders for medical image segmentation,” arXiv preprint arXiv:2102.04306, 2021.

[20]. Hatamizadeh Ali and Tang Yucheng and Nath Vishwesh and Yang Dong and Myronenko Andriy and Landman Bennett and Roth Holger R and Xu Daguang “Unetr: Transformers for 3d medical

image segmentation," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,2022, pp. 574–584.

[21]. Xie Yutong and Zhang Jianpeng and Shen Chunhua and Xia Yong "CoTr: Efficiently Bridging CNN and Transformer for 3D Medical Image Segmentation," arXiv preprint arXiv:2103.03024, 2021

[22]. Huang Gao and Liu Zhuang and Van Der Maaten Laurens and Weinberger Kilian Q "Densely connected convolutional networks" in Proceedings of the IEEE conference on computer vision and pattern recognition,2017, pp. 4700–4708

[23]. Krizhevsky Alex and Sutskever Ilya and Hinton, Geoffrey E "Imagenet classification with deep convolutional neural networks" in Advances in neural information processing systems,2012, 25, pp. 1097–1105

[24]. He Kaiming and Zhang Xiangyu and Ren Shaoqing and Sun Jian "Deep residual learning for image recognition" in Proceedings of the IEEE conference on computer vision and pattern recognition,2016, pp. 770–778

[25]. Tan Mingxing and Le Quoc "Efficientnet: Rethinking model scaling for convolutional neural networks" in International Conference on Machine Learning, 2019, pp. 6105–6114

[26]. Sandler Mark and Howard Andrew and Zhu Menglong and Zhmoginov Andrey and Chen Liang-Chieh "Mobilenetv2: Inverted residuals and linear bottlenecks" in Proceedings of the IEEE conference on computer vision and pattern recognition,2018, pp. 4510–4520

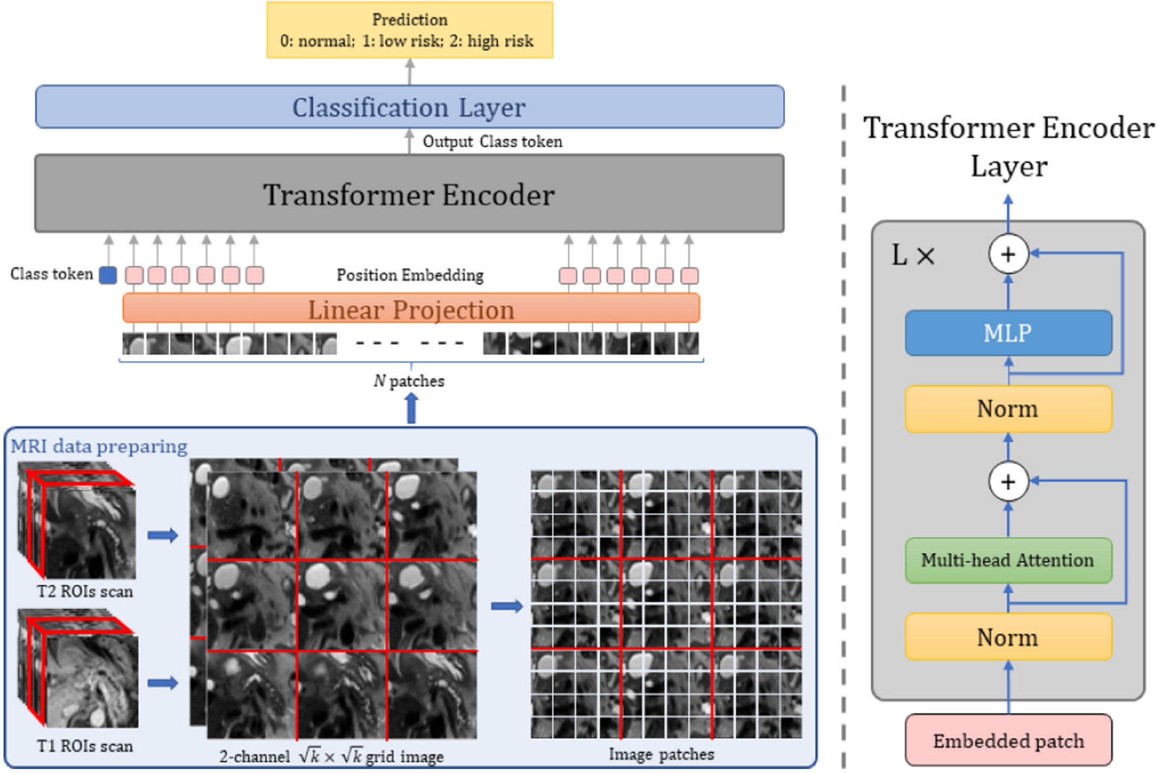

Fig. 1:

The proposed transformer-based architecture. T1 and T2 slices are concatenated along the channel dimension and sequences of 9 consecutive slices are arranged in a 3 × 3 grid. Patches are then extracted from the resulting image, and are used as input to the transformer architecture. After encoding the pach set through transformer layers (consisting of a cascade of multihead attention block and MLP layers), a special *classification token* encodes global image representation, and is used for final classification into three IPMN classes: *normal*, *low risk* and *high risk*.

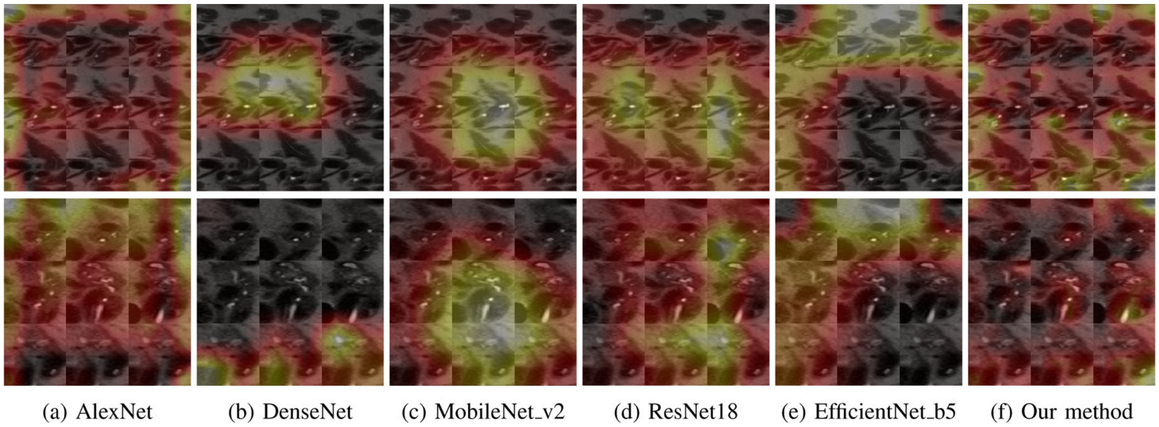

Fig. 2:

Comparison between the attention maps of AlexNet, DenseNet-121 and our model in case of correct (top row) and erroneous (bottom row) predictions on a 3×3 grid of MRI images.

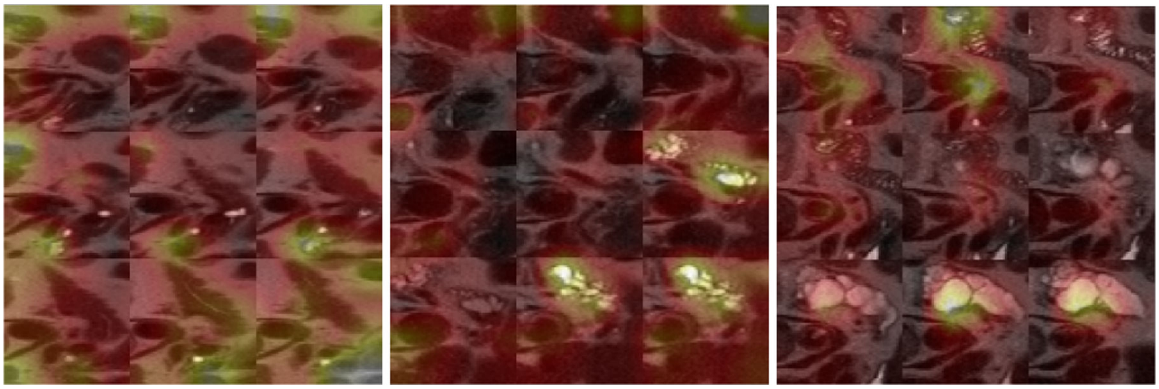

Fig. 3:

Attention maps of our transformer-based classifier on 3×3 grid of MRI images for correct IPMN classification.

TABLE I:

Performance of tested models with 10-fold nested cross-validation. We report results in term of mean ± standard deviation of metrics computed over all validation folds.

Method	Accuracy	Precision	Recall
AlexNet	0.42 ± 0.17	0.37 ± 0.15	0.39 ± 0.11
DenseNet	0.51 ± 0.12	0.54 ± 0.14	0.50 ± 0.14
ResNet18	0.53 ± 0.11	0.55 ± 0.23	0.32 ± 0.08
MobileNet_v2	0.43 ± 0.11	0.54 ± 0.26	0.35 ± 0.11
EfficientNet_b5	0.55 ± 0.10	0.60 ± 0.14	0.36 ± 0.08
Ours	**0.70** ± 0.11	**0.67** ± 0.19	**0.64** ± 0.12

TABLE II:

Performance of our model using different input data modality with 10-fold nested cross-validation. We report results in terms of mean ± standard deviation of metrics computed over all validation folds.

Method	Accuracy	Precision	Recall
T1	0.53 ± 0.08	0.60 ± 0.11	0.58 ± 0.14
T2	0.64 ± 0.12	0.64 ± 0.13	0.63 ± 0.11
T1+T2 modalities			
Late fusion	0.60 ± 0.16	0.61 ± 0.13	0.59 ± 0.11
Early fusion	**0.70** ± 0.11	**0.67** ± 0.19	**0.64** ± 0.12