



# Prognostic Value of Transfer Learning Based Features in Resectable Pancreatic Ductal Adenocarcinoma

Yucheng Zhang<sup>1</sup>, Edrize M. Lobo-Mueller<sup>2</sup>, Paul Karanicolas<sup>3</sup>, Steven Gallinger<sup>4</sup>, Masoom A. Haider<sup>4,5†</sup> and Farzad Khalvati<sup>1,6,7\*†</sup>

<sup>1</sup> Department of Medical Imaging, University of Toronto, Toronto, ON, Canada, <sup>2</sup> Department of Diagnostic Imaging and Department of Oncology, Faculty of Medicine and Dentistry, Cross Cancer Institute, University of Alberta, Edmonton, AB, Canada, <sup>3</sup> Department of Surgery, Sunnybrook Health Sciences Centre, Toronto, ON, Canada, <sup>4</sup> Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada, <sup>5</sup> Joint Department of Medical Imaging, Sinai Health System, University Health Network, University of Toronto, Toronto, ON, Canada, <sup>6</sup> Research Institute, The Hospital for Sick Children, Toronto, ON, Canada, <sup>7</sup> Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada

## OPEN ACCESS

### Edited by:

Frank Emmert-Streib,  
Tampere University, Finland

### Reviewed by:

Junjie Zhang,  
Royal Bank of Canada, Canada  
Shailesh Tripathi,  
Tampere University of  
Technology, Finland

### \*Correspondence:

Farzad Khalvati  
Farzad.Khalvati@utoronto.ca

<sup>†</sup> These authors share  
senior authorship

### Specialty section:

This article was submitted to  
Medicine and Public Health,  
a section of the journal  
Frontiers in Artificial Intelligence

**Received:** 10 April 2020

**Accepted:** 24 August 2020

**Published:** 05 October 2020

### Citation:

Zhang Y, Lobo-Mueller EM,  
Karanicolas P, Gallinger S, Haider MA  
and Khalvati F (2020) Prognostic  
Value of Transfer Learning Based  
Features in Resectable Pancreatic  
Ductal Adenocarcinoma.  
Front. Artif. Intell. 3:550890.  
doi: 10.3389/frai.2020.550890

**Background:** Pancreatic Ductal Adenocarcinoma (PDAC) is one of the most aggressive cancers with an extremely poor prognosis. Radiomics has shown prognostic ability in multiple types of cancer including PDAC. However, the prognostic value of traditional radiomics pipelines, which are based on hand-crafted radiomic features alone is limited.

**Methods:** Convolutional neural networks (CNNs) have been shown to outperform radiomics models in computer vision tasks. However, training a CNN from scratch requires a large sample size which is not feasible in most medical imaging studies. As an alternative solution, CNN-based transfer learning models have shown the potential for achieving reasonable performance using small datasets. In this work, we developed and validated a CNN-based transfer learning model for prognostication of overall survival in PDAC patients using two independent resectable PDAC cohorts.

**Results:** The proposed transfer learning-based prognostication model for overall survival achieved the area under the receiver operating characteristic curve of 0.81 on the test cohort, which was significantly higher than that of the traditional radiomics model (0.54). To further assess the prognostic value of the models, the predicted probabilities of death generated from the two models were used as risk scores in a univariate Cox Proportional Hazard model and while the risk score from the traditional radiomics model was not associated with overall survival, the proposed transfer learning-based risk score had significant prognostic value with hazard ratio of 1.86 (95% Confidence Interval: 1.15–3.53,  $p$ -value: 0.04).

**Conclusions:** This result suggests that transfer learning-based models may significantly improve prognostic performance in typical small sample size medical imaging studies.

**Keywords:** transfer learning, radiomics, prognosis, pancreatic cancer, survival analysis

## INTRODUCTION

Pancreatic Ductal Adenocarcinoma (PDAC) is one of the most aggressive malignancies with poor prognosis (Stark and Eibl, 2015; Stark et al., 2016; Adamska et al., 2017). Evidence suggested that surgery can improve overall survival in resectable PDAC cohorts (Stark et al., 2016; Adamska et al., 2017). However, the 5-year survival rate of patients who went through surgery is still low (Fatima et al., 2010). Thus, it is important to identify high-risk and low-risk surgical candidates so that healthcare providers can make personalized treatment decisions (Khalvati et al., 2019a). In resectable patients, clinicopathologic factors such as tumor size, margin status at surgery, and histological tumor grade have been studied as biomarkers for prognosis (Ahmad et al., 2001; Ferrone et al., 2012; Khalvati et al., 2019a). However, many of these biomarkers can only be assessed after the surgery and thus, the opportunity for patient-tailored neoadjuvant therapy is lost. Recently, quantitative medical imaging biomarkers have shown promising results in prognostication of the overall survival for cancer patients, providing an alternative solution (Kumar et al., 2012; Parmar et al., 2015; Lambin et al., 2017).

As a rapidly developing field in medical imaging, radiomics is defined as the extraction and analysis of a large number of quantitative imaging features from medical images including CT and MRI (Kumar et al., 2012; Lambin et al., 2012; Khalvati et al., 2019b). The conventional radiomic analysis pipeline consists of four steps as shown in **Figure 1**. Following this pipeline, several radiomic features have been shown to be significantly associated with clinical outcomes including overall survival or recurrence in different cancer sites such as lung, head and neck, and pancreas (Aerts et al., 2014; Coroller et al., 2015; Carneiro et al., 2016; Cassinotto et al., 2017; Chakraborty et al., 2017; Eilaghi et al., 2017; Lao et al., 2017; Zhang et al., 2017; Attiyeh et al., 2018; Yun et al., 2018; Sandrasegaran et al., 2019). Using these radiomic features, patients can be categorized into low-risk or high-risk groups guiding clinicians to design personalized treatment plans (Chakraborty et al., 2018; Varghese et al., 2019). Although limited work has been done in the context of PDAC, recent studies have confirmed the potential of new quantitative imaging biomarkers for resectable PDAC prognosis (Eilaghi et al., 2017; Khalvati et al., 2019a).

Despite recent progress, radiomics analytics solutions have a major limitation in terms of performance. The performance of radiomics models relies on the amount of information that radiomics features can capture from medical images (Kumar et al., 2012). Most radiomics features represent morphology, first order, or texture information from the regions of interest (Van Griethuysen et al., 2017). The equations of these radiomic features are often manually designed. This is a sophisticated and time-consuming process, requiring prior knowledge of image processing and tumor biology. Consequently, a poor design of

the feature bank may fail to extract important information from medical images, having a significant negative impact on the performance of prognostication. In contrast, the ability of deep learning for automatic feature extraction has been proven and shown to achieve promising performances in different medical imaging tasks (Shen et al., 2017; Yamashita et al., 2018; Yasaka et al., 2018).

A convolutional neural network (CNN) (Schmidhuber, 2014; LeCun et al., 2015) performs a series of convolution and pooling operations to get comprehensive quantitative information from input images (LeCun et al., 2015). Compared to hand-crafted radiomic features that are pre-designed and fixed, the coefficients of CNNs are modified in the training process. Hence, the final features generated from a successfully trained CNN are tuned to be associated with the target outcomes (e.g., overall survival, recurrence). It has been shown that CNN architectures are effective in different medical imaging tasks such as segmentation for head and neck anatomy and diagnosis for the retinal disease (Dalmış et al., 2017; De Fauw et al., 2018; Nikolov et al., 2018; Irvin et al., 2019).

However, to train a CNN from scratch, millions of parameters need to be tuned. This requires a large sample size which is not feasible to collect in most medical imaging studies (Du et al., 2018). As an alternative solution, CNN-based transfer learning is more suitable for medical imaging tasks since it can achieve a comparable performance using a limited amount of data (Pan and Yang, 2010; Chuen-Kai et al., 2015).

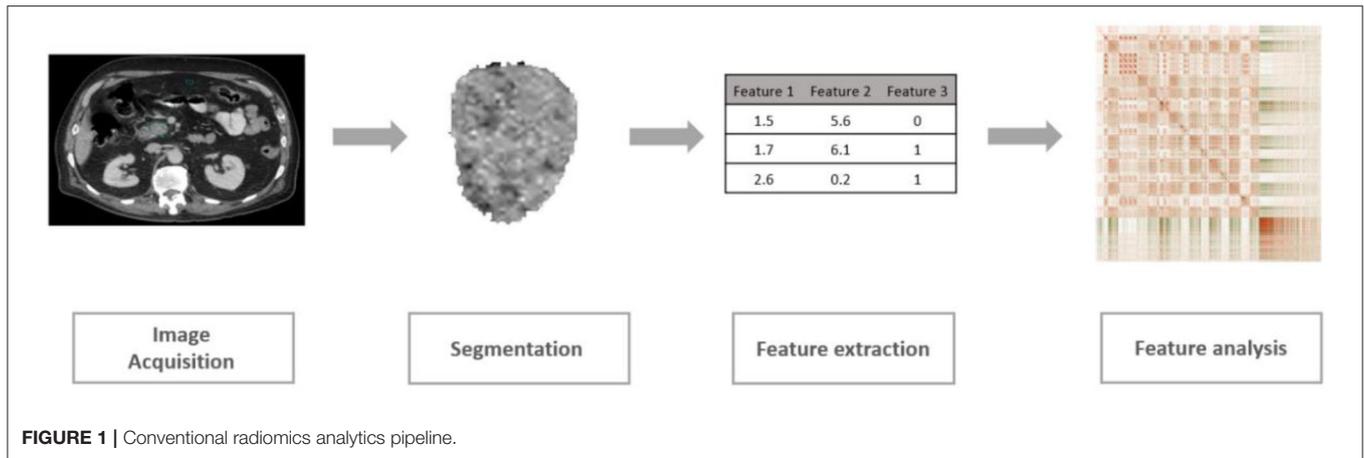
CNN-based transfer learning is defined as taking images from a different domain such as natural images (e.g., ImageNet) to build a pretrained model and then apply the pretrained model to target images (e.g., CT images of lung cancer) (Ravishankar et al., 2017). The idea of transfer learning is based on the assumption that the structure of a CNN is similar to the human visual cortex as both are composed of layers of neurons (Pan and Yang, 2010; Tan et al., 2018). Top layers of CNNs can extract general features from images while deeper layers are able to extract information that is more specific to the outcomes (Yosinski et al., 2014).

Transfer learning utilizes this property, training top layers using another large dataset while finetuning deeper layers using data from the target domain. For example, the ImageNet dataset contains more than 14 million images (Russakovsky et al., 2015). Hence, pretraining a model using this dataset would help the model learn how to extract general features using initial layers. Given that many image recognition tasks are similar, top (shallower) layers of the pretrained network can be transferred to another CNN model. In the next step, deeper layers of the CNN model can be trained using the target domain images (Torrey and Shavlik, 2009). Since the deeper layers are more target-specific, finetuning them using the images from the target domain may help the model quickly adapt to the target outcome, and hence, improve the overall performance.

In medical imaging, the target dataset is often so small that it is impractical to properly finetune the deeper layers. Consequently, in practice, a pretrained CNN can be used as a feature extractor (Hertel et al., 2015; Lao et al., 2017). Given that convolution layers can capture high-level and informative details from images, passing the target domain images through these layers allows

---

**Abbreviations:** ROC, Receiver operating characteristic; AUC, Area under the ROC curve; CT, Computed tomography; CI, Confidence interval; CNN, Convolutional neural network; GLCM, Gray-Level Co-occurrence matrix; NSCLC, Non-small-cell lung cancer; PDAC, Pancreatic ductal adenocarcinoma; ROI, Region of interest; SMOTE, Synthetic minority over-sampling technique.



extractions of features. These features can be further used to train a classifier for the target domain, enabling building a high-performance transfer learning model using a small dataset.

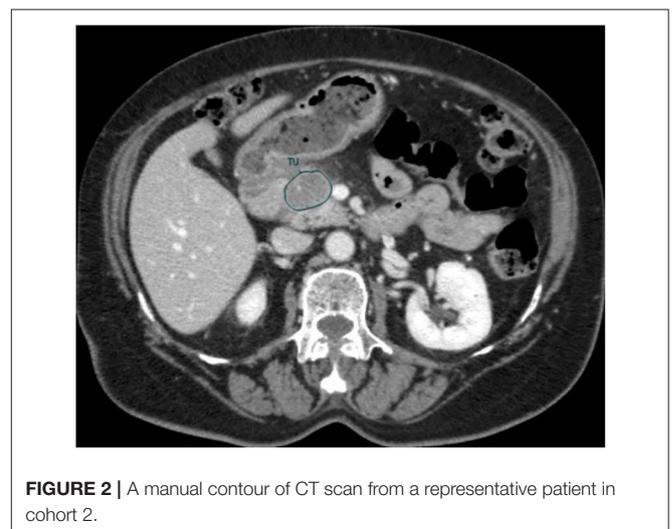
In this study, using two independent small sample size resectable PDAC cohorts, we evaluated the prognosis performance of a transfer learning model and compared its performance to that of a traditional radiomics model. The goal of the prognostication was to dichotomize PDAC patients who were candidates for curative-intent surgery to high-risk and low-risk groups. We found that the transfer learning model provides better prognostication performance compared to the conventional radiomics model, suggesting the potential of transfer learning in a typical small sample size medical imaging study.

## METHODS

### Dataset

Two cohorts from two independent hospitals consisting of 68 (Cohort 1) and 30 (Cohort 2) patients were enrolled in this retrospective study. All patients underwent curative intent surgical resection for PDAC from 2007–2012 to 2008–2013 in Cohort 1 and Cohort 2, respectively, and they did not receive other neoadjuvant treatment. Preoperative portal venous phase contrast-enhanced CT images were used. Overall survival (including survival as duration and death as the event) was collected as the primary outcome and it was calculated as the duration from the date of preoperative CT scan until death. To exclude the confounding effect of postoperative complications, patients who died within 90 days after the surgery were excluded. Institutional review board approval was obtained for this study from both institutions (Khalvati et al., 2019a).

An in-house developed Region of Interest (ROI) contouring tool (ProCanVAS Zhang et al., 2016) was used by a radiologist with 18 years of experience who completed the contours blind to the outcome (overall survival). Following the protocol, the slices were contoured with the largest visible 2D cross-section of the tumor on the portal venous phase. When the boundary of the tumor was not clear, it was defined by the presence of pancreatic or common bile duct cut-off and the review of pancreatic phase



images (Khalvati et al., 2019a). An example of the contour is shown in **Figure 2**.

### Radiomics Feature Extraction

Radiomics features were extracted using the PyRadiomics library (Van Griethuysen et al., 2017) (version 2.0.0) in Python. Voxels with Hounsfield unit under  $-10$  and above  $500$  were excluded so that the presence of fat and stents will not affect the values of the features. The bin width (number of gray levels per bin) was set to 25. In total, 1,428 radiomic features were extracted from CT images within the ROI for both cohorts. **Table 1** lists different classes of features used in this study (Khalvati et al., 2019a).

### Transfer Learning

We developed a transfer learning model (LungTrans) pretrained by CT images from non-small-cell lung cancer (NSCLC) patients. The Lung CT dataset was published on Kaggle for Lung Nodule Analysis (LUNA16), containing CT images from 888 lung cancer patients and the outcome (malignancy or not) (Armato et al., 2011). All input ROIs were resized to  $32 \times 32$  greyscale. An

8-layer CNN was trained from scratch using LUNA16 CT images with batch size 16 and learning rate 0.001 (Figure 3). This configuration was shown to have high performance in differentiating malignancy vs. normal tissue in the LUNA16 competition (De Wit, 2017). In addition, given small ROI sizes of data in this study (32×32) and the fact that images are grayscale instead of RGB color, off-the-shelf deep CNNs such as ResNet (He et al., 2015) do not provide adequate performance. Each convolutional layer except for Conv\_5 has Kernel size as 3×3 with stride of 1 with zero padding. Conv\_5 has 2×2 kernel size and stride of 1 without padding. All the Max Pooling layers have 2×2 kernel size. Previous research has shown that top layers in

the CNN extract generic features from the image, while bottom layers can extract features specific to the tasks (Yosinski et al., 2014; Paul et al., 2019). Since our pretrained domain (lung CT) and target domain (PDAC CT) are rather similar, we extracted features from the bottom layer. In addition, the number of features (coefficients) in the CNN significantly decreases as the layers become deeper, due to Max pooling. If we picked a layer above the final layer, the number of extracted features would increase significantly. Considering the sample size of our training (68) and test (30) datasets, all the convolution layers were frozen and features were extracted from the end of the CNN (Conv\_5). As a result, for each ROI from PDAC CT images, 64 features were extracted. This was the ideal number of intermediate features tested in LUNA16 dataset (De Wit, 2017).

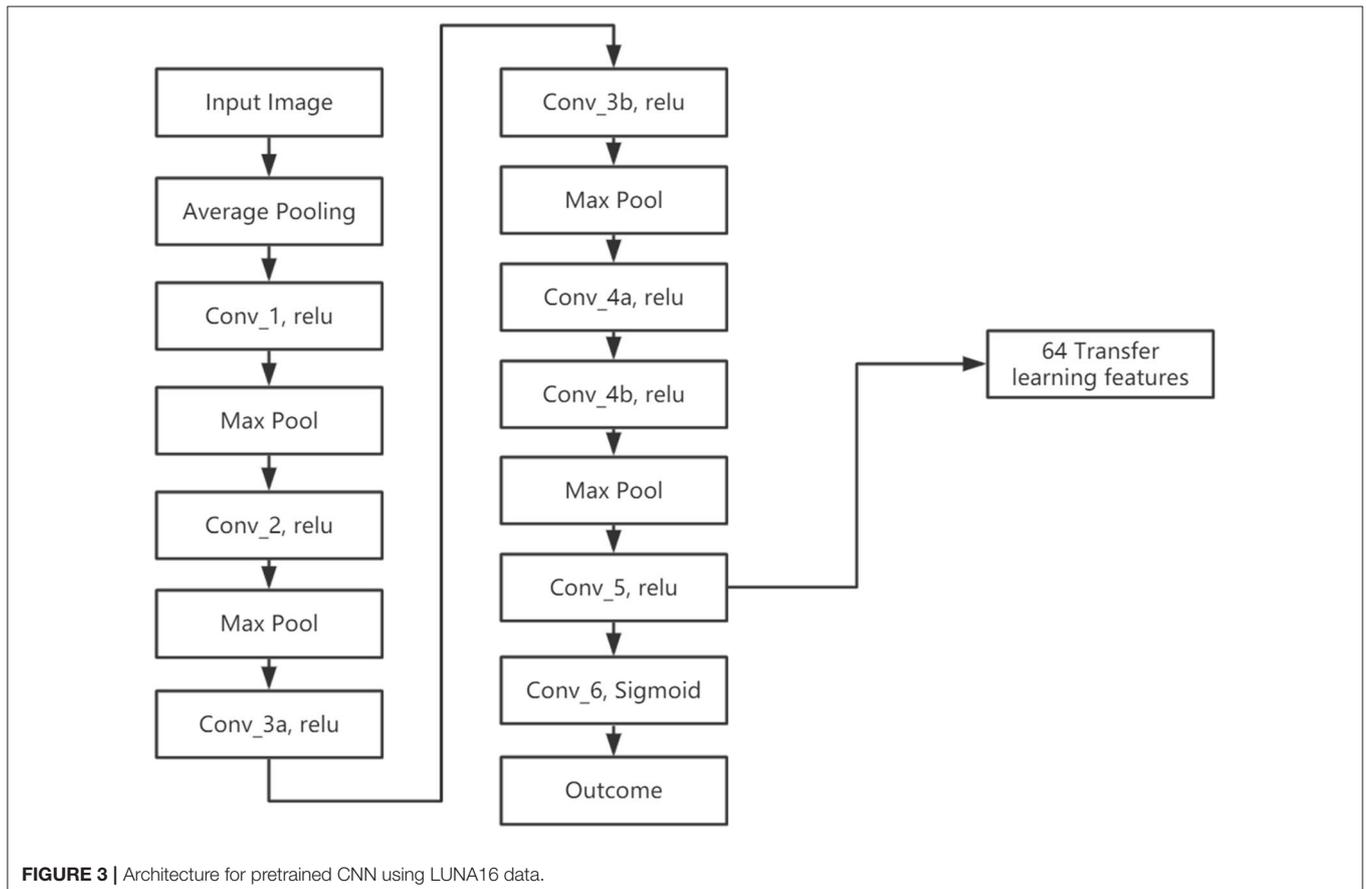
**TABLE 1** | List of radiomic feature classes and filters.

First-order features	Histogram-based features
Second-order texture features	Features extracted from Gray-Level Co-Occurrence matrix (GLCM)
Morphology features	Features based on the shape of the region of interest
Filters	No filter, exponential, gradient, logarithm, square, square-root, local binary pattern, wavelet

**TABLE 2A** | Summary of models' performances in AUC.

	Training cohort (n = 68) (5-Fold cross validation)	Test cohort (n = 30)
PyRadiomics model	0.57 (95% CI: 0.42–0.73)	0.54 (95% CI: 0.32–0.76)
Transfer learning model	0.72 (95% CI: 0.58–0.86)	0.81 (95% CI: 0.64–0.98)

Tables 2B,C show Confusion Matrix for Random Forest models using PyRadiomics and LungTrans features, respectively, in the test cohort.



## Prognostic Models

To have a proper and robust validation, training and test datasets were collected from two different institutions. In Cohort 1 (training cohort,  $n = 68$ ), two prognostic models for overall survival were trained using features extracted from conventional radiomics feature bank (PyRadiomics) and transfer learning model (LungTrans). The prognosis models were built using the Random Forest classifier, which is a common classifier in radiomics analytic pipeline, with 500 decision trees (Chen and Ishwaran, 2012; Zhang et al., 2017). Random Forest classifier is highly data-adaptive, which have shown the potential to handle large P small N problem by choosing the best subset of features for classification (Chen and Ishwaran, 2012). The “data-adaptive” characteristic makes the random forest a good candidate for our study where transfer learning and PyRadiomics offered different numbers of features. The number of variables available for splitting at each tree node (mtry) was determined by the best performing mtry option in the training cohort. Due to the imbalanced outcome in the training data, (Cohort 1: 52 Deaths vs. 16 Survivals), a data balancing algorithm, SMOTE (Ryu et al.,

2002), was applied in the training process to artificially balance the training data.

The prognostic values of these two models were evaluated in Cohort 2 ( $n = 30$ , 15 Deaths vs. 15 Survivals) using the area under the receiver operating characteristic (ROC) curve (AUC). DeLong test, as one of the common comparison tests, was used to test the difference between the two ROC curves (DeLong et al., 1988). To further assess the prognosis values, the predicted probabilities of death generated from the two classifiers were used as risk scores in survival analyses. These risk scores were tested in Cohort 2 using univariate Cox Proportional Hazards Model for their Hazard Ratio and Wald test  $p$ -value (Cox, 1972). These analyses were done in R (version 3.5.1) using “caret,” “pROC,” and “survival” packages (Kuhn, 2008; Therneau, 2020).

## RESULTS

### Prognostic Models Performance

Using features from the PyRadiomics feature bank, the Random Forest model yielded AUC of 0.54 [95% Confidence Interval (CI): 0.32–0.76] in the test cohort (Cohort 2) (mtry: 2). In contrast, using LungTrans features, the AUC of the Random Forest model reached 0.81 (95% CI: 0.64–0.98) in the test cohort (mtry: 17). The performances of these two models for both training and test cohorts are listed in **Table 2A**. We performed a 5-fold cross-validation to produce AUCs for the training cohort. The AUCs for the test cohort were generated using the models trained by the training cohort.

To investigate the prognostic value of each PyRadiomics features, variable importance indices were calculated using the Caret Package in R. The top ten features were first order entropy, first order uniformity, first order interquartile range, GLSZM gray level non-uniformity normalized, GLRLM run length non-uniformity normalized, GLCM cluster tendency, NGTDM busyness, GLSZM small area high gray level emphasis, GLSZM low gray level zone emphasis, and GLSZM large area high gray level emphasis. This confirming previous studies in

**TABLE 2B** | Confusion Matrix of PyRadiomics model in the test cohort.

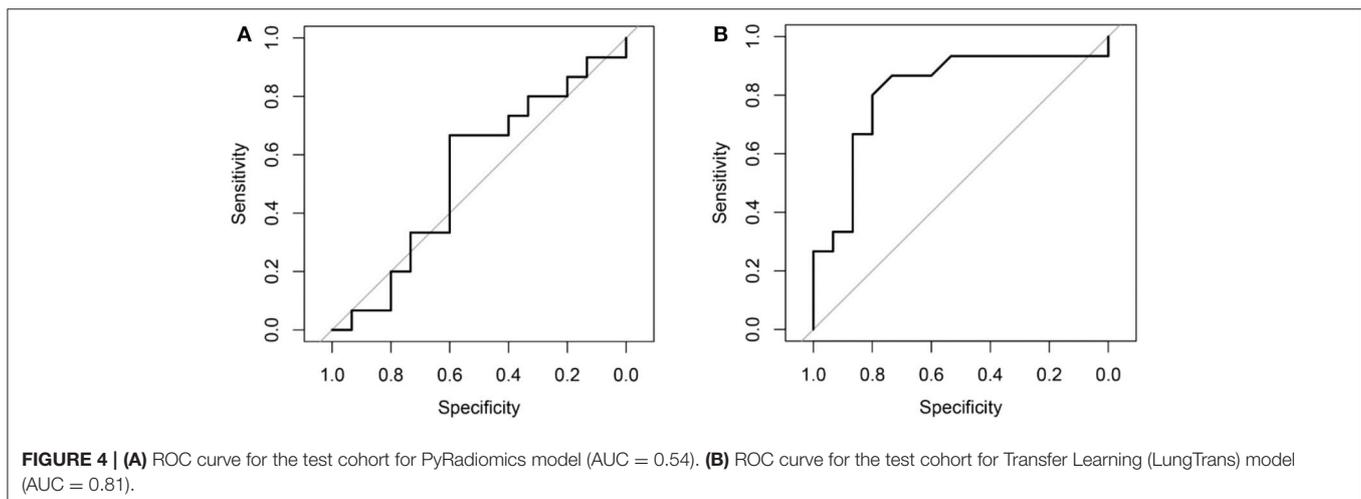
Test cohort	Deceased patients	Survived patients
Predicted death	12	10
Predicted survival	3	5

Accuracy: 0.57, Sensitivity: 0.8, Specificity: 0.33, Precision: 0.55.

**TABLE 2C** | Confusion matrix of transfer learning model in the test cohort.

Test cohort	Deceased patients	Survived patients
Predicted Death	13	4
Predicted Survival	2	11

Accuracy: 0.80, Sensitivity: 0.87, Specificity: 0.73, Precision: 0.76.



this field where similar radiomic features have been reported to be prognostic of PDAC (Eilaghi et al., 2017; Chu et al., 2019; Khalvati et al., 2019a; Li et al., 2020). It is worth noting that morphologic features were not ranked as top features in the list. This may be attributed to the challenges associated with contouring the PDAC regions of interest, leading to the low robustness of morphology features.

Comparing the ROC curves using Delong ROC test (DeLong et al., 1988), the LungTrans (Transfer Learning) prognosis model had significantly higher performance than that of PyRadiomics feature bank with a  $p$ -value of 0.0056 (AUC of 0.81 vs. 0.54). This result indicated that the transfer learning model based on lung CT images (LungTrans) significantly improved the prognostic performance compared to that of the traditional radiomics methods (PyRadiomics). **Figure 4** shows the ROC curves for the two models for the test cohort.

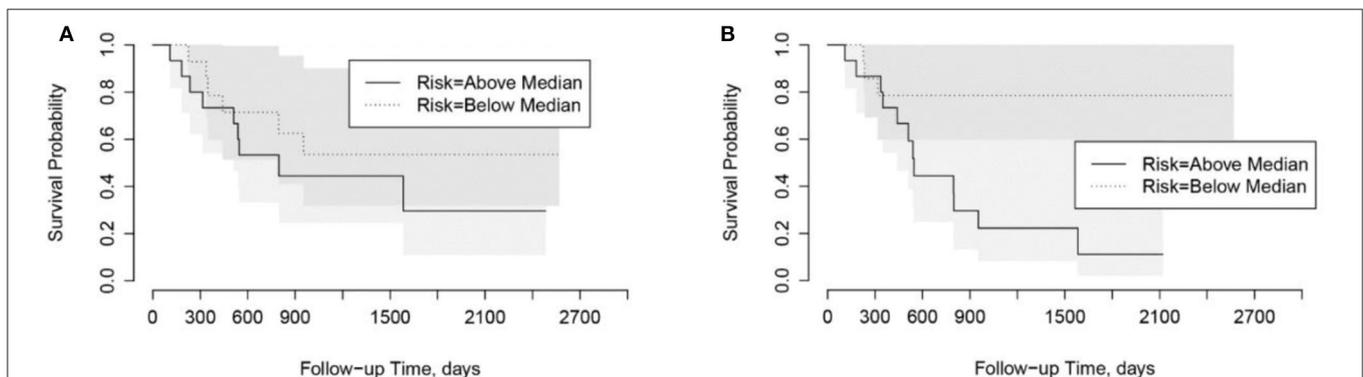
## Risk Score

In univariate Cox Proportional Hazard analysis, the risk score from the PyRadiomics model was not associated with overall survival. In contrast, the risk score from the LungTrans model had significant prognostic value with a Hazard Ratio of 1.86 [95% Confidence Interval (CI): 1.15–3.53],  $p$ -value: 0.04 as shown in **Table 3**.

Using the risk scores, patients can be categorized into low-risk or high-risk groups based on the median values. As shown in Kaplan-Meier plots in **Figure 5**, the LungTrans model was able to differentiate patients with high risk from those with low risk. This result further confirms that the transfer learning feature extractor pretrained by NSCLC CT images is capable of providing prognostic information for PDAC patients.

**TABLE 3** | Performance of risk score models in Cox Proportional Hazard analysis.

	Hazard ratio and CI	$p$
PyRadiomics based risk score	1.03 (95% CI: 0.60–1.76)	0.91
Transfer learning based risk score	1.86 (95% CI: 1.15–3.53)	0.04



**FIGURE 5** | Kaplan-Meier plots for overall survival in Cohort 2. **(A)** PyRadiomics based risk score ( $p = 0.91$ ). **(B)** Transfer Learning (LungTrans) based risk score ( $p = 0.04$ ).

## DISCUSSION

In this study, we developed and compared two prognostic models for overall survival for resectable PDAC patients using the PyRadiomics and transfer learning features banks pretrained by lung CT images (LungTrans). The LungTrans model achieved significantly better prognosis performance compared to that of the traditional radiomics approach (AUC of 0.81 vs. 0.54). This result suggested that the transfer learning approach has the potential of significantly improving prognosis performance in the resectable PDAC cohort using CT images.

Previous transfer learning studies in medical imaging research often utilized ImageNet pretrained models (Chuen-Kai et al., 2015; Lao et al., 2017). In our study, we used a lung CT pretrained CNN (LungTrans) as feature extractor and showed the potential of transfer learning in a typical small sample size setting. Although CNNs are capable of achieving high performance in image recognition tasks, training these networks needs a large sample size. If a CNN with the same architecture as LungTrans was trained from scratch in the training cohort (Cohort 1), it could not provide any prognostic value in the test cohort (Cohort 2) (AUC of  $\sim 0.50$ ). Transfer learning, unlike conventional deep learning methods which need large datasets, can achieve reasonable performance using a limited number of samples, making it suitable for most medical imaging studies. Although the training cohort in our study was small ( $n = 68$ ), in the PDAC test cohort, our transfer learning model had positive predictive value (Precision) of 76%, demonstrating its prognostic value in finding high-risk patients. This may significantly benefit patients by providing personalized neoadjuvant or adjuvant therapy for better prognosis.

Although the proposed transfer learning model outperformed the conventional radiomics model, this was not an indication to discard radiomic features altogether. These hand-crafted features have been shown to be prognostic for survival and recurrence in different cancer sites (Kumar et al., 2012; Balagurunathan et al., 2014; Haider et al., 2017). In the PDAC radiomics field, more than forty features have been found to be significantly associated with tissue classification or overall survival for PDAC patients (e.g., sum entropy, cluster tendency, dissimilarity, uniformity,

and busyness) (Cassinotto et al., 2017; Chakraborty et al., 2017; Attiyeh et al., 2018; Yun et al., 2018; Chu et al., 2019; Sandrasegaran et al., 2019; Li et al., 2020; Park et al., 2020). Furthermore, a few radiomics features have been found to be associated with tumor heterogeneity and genomics profile (Lambin et al., 2012; Itakura et al., 2015; Rizzo et al., 2016; Li et al., 2018). Hence, radiomics features can provide unique information about the lesions. Thus, studying the associations between radiomics and transfer learning features, together with feature fusion analysis, may further improve the prognostication performance in future research.

Despite achieving promising results, we should also note that the differences between NSCLC and PDAC are substantial, in terms of their biological profiles and prognoses, and thus, they may not have similar appearances in CT images. This is a limitation of the present study. A larger PDAC dataset would allow us to address these differences and test different transfer learning approaches in the context of PDAC prognosis. For example, finetuning a few layers of the CNN pretrained by NSCLC CT images using PDAC CT images would allow the network extract features that may further adapt to the PDAC images and lead to better performance.

In this study, we aimed to improve the accuracy of the survival model using the transfer learning approach. For diseases with poor prognosis, including PDAC, providing binary survival classifications offers limited information for clinicians for decision making since the survival rates are usually low. It would be more beneficial to provide time vs. risk information, e.g., identify the high-risk time intervals for a resectable PDAC patient using CT images. Future studies may choose to combine the transfer learning-based features extraction methods with the recent work on deep learning-based survival models (e.g., DeepSurv Katzman et al., 2018) to provide more practical prognosis information for personalized care.

## CONCLUSION

Deep transfer learning has the potential to improve the performance of prognostication for cancers with limited sample sizes such as PDAC. In this work, the proposed transfer learning model outperformed a predefined radiomics model for prognostications in resectable PDAC cohorts.

## REFERENCES

- Adamska, A., Domenichini, A., and Falasca, M. (2017). Pancreatic ductal adenocarcinoma: current and evolving therapies. *Int. J. Mol. Sci.* 18:1338. doi: 10.3390/ijms18071338
- Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 5:4006. doi: 10.1038/ncomms5006
- Ahmad, N. A., Lewis, J. D., Ginsberg, G. G., Haller, D. G., Morris, J. B., Williams, N. N., et al. (2001). Long term survival after pancreatic resection for pancreatic adenocarcinoma. *Am. J. Gastroenterol.* 96, 2609–2615. doi: 10.1111/j.1572-0241.2001.04123.x

## DATA AVAILABILITY STATEMENT

The datasets of Cohort 1 and Cohort 2 analyzed during the current study are available from the corresponding author on reasonable request pending the approval of the institution(s) and trial/study investigators who contributed to the dataset.

## ETHICS STATEMENT

This study was reviewed and approved by the research ethics boards of University Health Network, Sinai Health System, and Sunnybrook Health Sciences Centre. For this retrospective study the informed consent was obtained for Cohort 1 and the need for informed consent was waived for Cohort 2.

## AUTHOR CONTRIBUTIONS

YZ, MAH, and FK contributed to the design of the concept. EML, SG, PK, MAH, and FK contributed in collecting and reviewing the data. YZ and FK contributed to the design and implementation of quantitative imaging feature extraction and machine learning modules. All authors contributed to the writing and reviewing of the paper and read and approved the final manuscript.

## FUNDING

This study was conducted with support of the Ontario Institute for Cancer Research (PanCuRx Translational Research Initiative) through funding provided by the Government of Ontario, the Wallace McCain Centre for Pancreatic Cancer supported by the Princess Margaret Cancer Foundation, the Terry Fox Research Institute, the Canadian Cancer Society Research Institute, and the Pancreatic Cancer Canada Foundation. This study was also supported by charitable donations from the Canadian Friends of the Hebrew University (Alex U. Soyka). The funding bodies had no role in the design of the study, collection, analysis, and interpretation of data, or in writing the manuscript.

## ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at arXiv (Zhang et al., 2019).

- Armato, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., et al. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* 38, 915–931. doi: 10.1118/1.3528204
- Attiyeh, M. A., Chakraborty, J., Doussot, A., Langdon-Embry, L., Mainarich, S., Gönen, M., et al. (2018). Survival prediction in pancreatic ductal adenocarcinoma by quantitative computed tomography image analysis. *Ann. Surg. Oncol.* 25, 1034–1042. doi: 10.1245/s10434-017-6323-3
- Balagurunathan, Y., Kumar, V., Gu, Y., Kim, J., Wang, H., Liu, Y., et al. (2014). Test–retest reproducibility analysis of lung CT image features. *J. Digit. Imaging* 27, 805–823. doi: 10.1007/s10278-014-9716-x
- Carneiro, G., Oakden-Rayner, L., Bradley, A. P., Nascimento, J., and Palmer, L. (2016). “Automated 5-year mortality prediction using deep learning and radiomics features from chest computed tomography,” in *Autom. 5-year*

- Mortal. Predict. Using Deep Learn. Radiomics Featur. from Chest Comput. Tomogr.* doi: 10.1109/ISBI.2017.7950485
- Cassinotto, C., Chong, J., Zogopoulos, G., Reinhold, C., Chiche, L., Lafourcade, J. P., et al. (2017). Resectable pancreatic adenocarcinoma: Role of CT quantitative imaging biomarkers for predicting pathology and patient outcomes. *Eur. J. Radiol.* 90, 152–158. doi: 10.1016/j.ejrad.2017.02.033
- Chakraborty, J., Langdon-Embry, L., Cunanan, K. M., Escalon, J. G., Allen, P. J., Lowery, M. A., et al. (2017). Preliminary study of tumor heterogeneity in imaging predicts two year survival in pancreatic cancer patients. *PLoS ONE*. 12:e0188022. doi: 10.1371/journal.pone.0188022
- Chakraborty, J., Midya, A., Gazit, L., Attiyeh, M., Langdon-Embry, L., Allen, P. J., et al. (2018). CT radiomics to predict high-risk intraductal papillary mucinous neoplasms of the pancreas. *Med. Phys.* 45, 5019–5029. doi: 10.1002/mp.13159
- Chen, X., and Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics* 99, 323–329. doi: 10.1016/j.ygeno.2012.04.003
- Chu, L. C., Park, S., Kawamoto, S., Fouladi, D. F., Shayesteh, S., Zinreich, E. S., et al. (2019). Utility of CT radiomics features in differentiation of pancreatic ductal adenocarcinoma from normal pancreatic tissue. *Am. J. Roentgenol.* 213, 349–357. doi: 10.2214/AJR.18.20901
- Chuen-Kai, S., Chung-Hisang, C., Chun-Nan, C., Meng-Hsi, W., and Edward, Y. C. (2015). “Transfer representation learning for medical image analysis,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.
- Coroller, T. P., Grossmann, P., Hou, Y., Rios Velazquez, E., Leijenaar, R. T., Hermann, G., et al. (2015). CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother. Oncol.* 114, 345–350. doi: 10.1016/j.radonc.2015.02.015
- Cox, D. R. (1972). Regression models and life-tables. *J. R. Statist. Soc.* 34, 187–220. doi: 10.1111/j.2517-6161.1972.tb00899.x
- Dalmış, M. U., Litjens, G., Holland, K., Setio, A., Mann, R., Karssemeijer, N., et al. (2017). Using deep learning to segment breast and fibroglandular tissue in MRI volumes. *Med. Phys.* 44, 533–546. doi: 10.1002/mp.12079
- De Fauw, J., LedSAM, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* 24, 1342–1350. doi: 10.1038/s41591-018-0107-6
- De Wit, J. (2017). Kaggle datascience bowl 2017. *GitHub/kaggle\_ndsb* (2017).
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837–845. doi: 10.2307/2531595
- Du, S. S., Wang, Y., Zhai, X., Balakrishnan, S., Salakhutdinov, R., and Singh, A. (2018). “How many samples are needed to estimate a convolutional neural network?” in *Conference on Neural Information Processing Systems*.
- Eilaghi, A., Baig, S., Zhang, Y., Zhang, J., Karanicolas, P., Gallinger, S., et al. (2017). CT texture features are associated with overall survival in pancreatic ductal adenocarcinoma – a quantitative analysis. *BMC Med. Imaging* 17:38. doi: 10.1186/s12880-017-0209-5
- Fatima, J., Schnellrdorfer, T., Barton, J., Wood, C. M., Wiste, H. J., Smyrk, T. C., et al. (2010). Pancreatoduodenectomy for ductal adenocarcinoma: Implications of positive margin on survival. *Arch. Surg.* 145, 167–172. doi: 10.1001/archsurg.2009.282
- Ferrone, C. R., Pieretti-Vanmarcke, R., Bloom, J. P., Zheng, H., Szymoniaka, J., Wargo, J. A., et al. (2012). Pancreatic ductal adenocarcinoma: long-term survival does not equal cure. *Surgery* 152, S43–S49. doi: 10.1016/j.surg.2012.05.020
- Haider, M. A., Vosough, A., Khalvati, F., Kiss, A., Ganeshan, B., Bjarnason, G. A., et al. (2017). CT texture analysis: a potential tool for prediction of survival in patients with metastatic clear cell carcinoma treated with sunitinib. *Cancer Imaging* 17:4. doi: 10.1186/s40644-017-0106-8
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). “Deep residual learning for image,” in *Recognition* 770–77. doi: 10.1109/CVPR.2016.90
- Hertel, L., Barth, E., Käster, T., and Martinetz, T. (2015). “Deep convolutional neural networks as generic feature extractors,” in *2015 International Joint Conference on Neural Networks (IJCNN)*.
- Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., et al. (2019). “CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison,” in *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*.
- Itakura, H., Achrol, A. S., Mitchell, L. A., Loya, J. J., Liu, T., Westbroek, E. M., et al. (2015). Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. *Sci. Transl. Med.* 7:303ra138. doi: 10.1126/scitranslmed.aaa7582
- Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y., et al. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* 18:24. doi: 10.1186/s12874-018-0482-1
- Khalvati, F., Zhang, Y., Baig, S., Lobo-Mueller, E. M., Karanicolas, P., Gallinger, S., et al. (2019a). Prognostic value of CT radiomic features in resectable pancreatic ductal adenocarcinoma. *Nat. Sci. Reports.* 9:5449. doi: 10.1038/s41598-019-41728-7
- Khalvati, F., Zhang, Y., Wong, A., and Haider, M. A. (2019b). Radiomics. *Encyclopedia Biomed Eng.* 2, 597–603. doi: 10.1016/B978-0-12-801238-3.99964-1
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i05
- Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., et al. (2012). Radiomics: the process and the challenges. *Magn. Reson. Imaging* 30, 1234–1248. doi: 10.1016/j.mri.2012.06.010
- Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., et al. (2017). Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* 14, 749–762. doi: 10.1038/nrclinonc.2017.141
- Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., et al. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* 48, 441–446. doi: 10.1016/j.ejca.2011.11.036
- Lao, J., Chen, Y., Li, Z. C., Li, Q., Zhang, J., Liu, J., et al. (2017). A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* 7:10353. doi: 10.1038/s41598-017-10649-8
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Li, K., Yao, Q., Xiao, J., Li, M., Yang, J., Hou, W., et al. (2020). Contrast-enhanced CT radiomics for predicting lymph node metastasis in pancreatic ductal adenocarcinoma: a pilot study. *Cancer Imaging* 20:12. doi: 10.1186/s40644-020-0288-3
- Li, Y., Qian, Z., Xu, K., Wang, K., Fan, X., Li, S., et al. (2018). MRI features predict p53 status in lower-grade gliomas via a machine-learning approach. *NeuroImage Clin.* 17, 306–311. doi: 10.1016/j.nicl.2017.10.030
- Nikolov, S., Blackwell, S., Mendes, R., De Fauw, J., Meyer, C., Hughes, C., et al. (2018). Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *arXiv:1809.04430v1*.
- Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Park, S., Chu, L., Hruban, R. H., Vogelstein, B., Kinzler, K. W., Yuille, A. L., et al. (2020). Differentiating autoimmune pancreatitis from pancreatic ductal adenocarcinoma with CT radiomics features. *Diagn. Interv. Imaging.* 1, 770–778. doi: 10.1016/j.diii.2020.03.002
- Parmar, C., Grossmann, P., Bussink, J., Lambin, P., and Aerts, H. J. W. L. (2015). Machine learning methods for quantitative radiomic biomarkers. *Sci. Rep.* 5:13087. doi: 10.1038/srep13087
- Paul, R., Schabath, M., Balagurunathan, Y., Liu, Y., Li, Q., Gillies, R., et al. (2019). Explaining deep features using radiologist-defined semantic features and traditional quantitative features. *Tomogr.* 5, 192–200. doi: 10.18383/j.tom.2018.00034
- Ravishankar, H., Sudhakar, P., Venkataramani, R., Thiruvankadam, S., Annangi, P., Babu, N., et al. (2017). “Understanding the mechanisms of deep transfer learning for medical images,” in *Deep Learning and Data Labeling for Medical Applications*. 188–196. doi: 10.1007/978-3-319-46976-8\_20
- Rizzo, S., Petrella, F., Buscarino, V., De Maria, F., Raimondi, S., Barberis, M., et al. (2016). CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. *Eur. Radiol.* 26, 32–42. doi: 10.1007/s00330-015-3814-0
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

- Ryu, S., Lee, H., Lee, D. K., Kim, S. W., Kim, C. E., Chawla, N. V., et al. (2002). SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Sandrasegaran, K., Lin, Y., Asare-Sawiri, M., Taiyini, T., and Tann, M. (2019). CT texture analysis of pancreatic cancer. *Eur. Radiol.* 29, 1067–1073. doi: 10.1007/s00330-018-5662-1
- Schmidhuber, J. (2014). Deep learning in neural networks: an overview. *Neural Networks* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003
- Shen, D., Wu, G., and Suk, H.-I. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248. doi: 10.1146/annurev-bioeng-071516-044442
- Stark, A., and Eibl, G. (2015). “Pancreatic Ductal Adenocarcinoma,” in *Pancreapedia: The Exocrine Pancreas Knowledge Base*. Version 1.0 (Ann Arbor, MI: Michigan Publishing; University of Michigan Library), 1–9. doi: 10.3998/panc.2015.14
- Stark, A. P., Ikoma, N., Chiang, Y. J., Estrella, J. S., Das, P., Minsky, B. D., et al. (2016). Long-term survival in patients with pancreatic ductal adenocarcinoma. *Surgery* 159, 1520–1527. doi: 10.1016/j.surg.2015.12.024
- Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., et al. (2018). “A survey on deep transfer learning,” in *Artificial Neural Networks and Machine Learning*. 270–279. doi: 10.1007/978-3-030-01424-7\_27
- Therneau, T. M. (2020). *A Package for Survival Analysis in R*. R package version 32–3.
- Torrey, L., and Shavlik, J. (2009). “Transfer learning,” in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, eds E.S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, A. J. S. López.
- Van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 77, e104–e107. doi: 10.1158/0008-5472.CAN-17-0339
- Varghese, B., Chen, F., Hwang, D., Palmer, S. L., De Castro Abreu, A. L., Ukimura, O., et al. (2019). Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images. *Sci. Rep.* 9:1570. doi: 10.1038/s41598-018-38381-x
- Yamashita, R., Nishio, M., Do, R. K. G., Togashi, K., Yamashita, R., Nishio, M., et al. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9, 611–629. doi: 10.1007/s13244-018-0639-9
- Yasaka, K., Akai, H., Abe, O., Kiryu, S., Yasaka, K., Akai, H., et al. (2018). Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology* 286, 887–896. doi: 10.1148/radiol.2017170706
- Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). “How transferable are features in deep neural networks?” in *Advances in Neural Information Processing Systems* 27 (NIPS 2014).
- Yun, G., Kim, Y. H., Lee, Y. J., Kim, B., Hwang, J. H., Choi, D. J., et al. (2018). Tumor heterogeneity of pancreas head cancer assessed by CT texture analysis: association with survival outcomes after curative resection. *Sci. Rep.* 8:7226. doi: 10.1038/s41598-018-25627-x
- Zhang, J., Baig, S., Wong, A., Haider, M. A., and Khalvati, F. (2016). A local ROI-specific Atlas-based segmentation of prostate gland and transitional zone in diffusion MRI. *J. Comput. Vis. Imaging Syst.* 2, 2–4. doi: 10.15353/vsnl.v2i1.113
- Zhang, Y., Lobo-Mueller, E. M., Karanicolas, P., Gallinger, S., Haider, M. A., Khalvati, F., et al. (2019). *Prognostic Value of Transfer Learning Based Features in Resectable Pancreatic Ductal Adenocarcinoma*. arXiv.
- Zhang, Y., Oikonomou, A., Wong, A., Haider, M. A., and Khalvati, F. (2017). Radiomics-based prognosis analysis for non-small cell lung cancer. *Nat. Sci. Rep.* 7:46349. doi: 10.1038/srep46349

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JZ declared a past co-authorship with one of the authors FK to the handling editor.

Copyright © 2020 Zhang, Lobo-Mueller, Karanicolas, Gallinger, Haider and Khalvati. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.