Original Research Article

Improving feature extraction from histopathological images through a fine-tuning ImageNet model

Xingyu Li [a], Min Cen [a], Jinfeng Xu [b], Hong Zhang [a,*], Xu Steven Xu [c,*]

[a] Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China
[b] Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong
[c] Clinical Pharmacology and Quantitative Science, Genmab Inc., Princeton, New Jersey, USA

ABSTRACT

Background: Due to lack of annotated pathological images, transfer learning has been the predominant approach in the field of digital pathology. Pre-trained neural networks based on ImageNet database are often used to extract "off-the-shelf" features, achieving great success in predicting tissue types, molecular features, and clinical outcomes, etc. We hypothesize that fine-tuning the pre-trained models using histopathological images could further improve feature extraction, and downstream prediction performance.

Methods: We used 100 000 annotated H&E image patches for colorectal cancer (CRC) to fine-tune a pre-trained Xception model via a 2-step approach. The features extracted from fine-tuned Xception (FTX-2048) model and Image-pretrained (IMGNET-2048) model were compared through: (1) tissue classification for H&E images from CRC, same image type that was used for fine-tuning; (2) prediction of immune-related gene expression, and (3) gene mutations for lung adenocarcinoma (LUAD). Five-fold cross validation was used for model performance evaluation. Each experiment was repeated 50 times.

Findings: The extracted features from the fine-tuned FTX-2048 exhibited significantly higher accuracy (98.4%) for predicting tissue types of CRC compared to the "off-the-shelf" features directly from Xception based on ImageNet database (96.4%) (P value = 2.2×10^{-6}). Particularly, FTX-2048 markedly improved the accuracy for stroma from 87% to 94%. Similarly, features from FTX-2048 boosted the prediction of transcriptomic expression of immune-related genes in LUAD. For the genes that had significant relationships with image features (P < 0.05, n = 171), the features from the fine-tuned model improved the prediction for the majority of the genes (139; 81%). In addition, features from FTX-2048 improved prediction of mutation for 5 out of 9 most frequently mutated genes (STK11, TP53, LRP1B, NF1, and FAT1) in LUAD.

Conclusions: We proved the concept that fine-tuning the pretrained ImageNet neural networks with histopathology images can produce higher quality features and better prediction performance for not only the same-cancer tissue classification where similar images from the same cancer are used for fine-tuning, but also cross-cancer prediction for gene expression and mutation at patient level.

Introduction

Recent advancements of artificial intelligence and deep learning algorithms have greatly improved the current pathological workflows, such as tissue classification, and disease grading, etc.[1,2] These computer vision algorithms are usually based on convolutional neural networks (CNNs), which has also been successfully applied for tumor detection, prediction of histological subtypes, genomic mutations, transcriptomic expression, molecular subtyping, and patient prognosis.[1,3–15]

Deep learning computer vision models often rely on a large number of annotated images.[16] Unfortunately, until now, databases with annotated pathological images are still very limited. Therefore, transfer learning has been the predominant approach in the field of digital pathology. Particularly, use of "off-the-shelf" features directly extracted from the layer just prior to the final layer of pre-trained networks based on large-scale ImageNet database has been a popular approach to pre-processing the images for downstream prediction exercises.[17–19] It is worth to point out that even though the ImageNet models are not trained on pathological images, the current workflow (i.e., prediction using features extracted with pre-trained ImageNet models) generally provides a satisfactory performance and has been shown to outperform training the same neural network architecture from scratch using limited histopathological images.[12,20,21]

Another transfer learning approach is to fine-tune the middle layers of these pre-trained ImageNet neural networks with histopathological images.
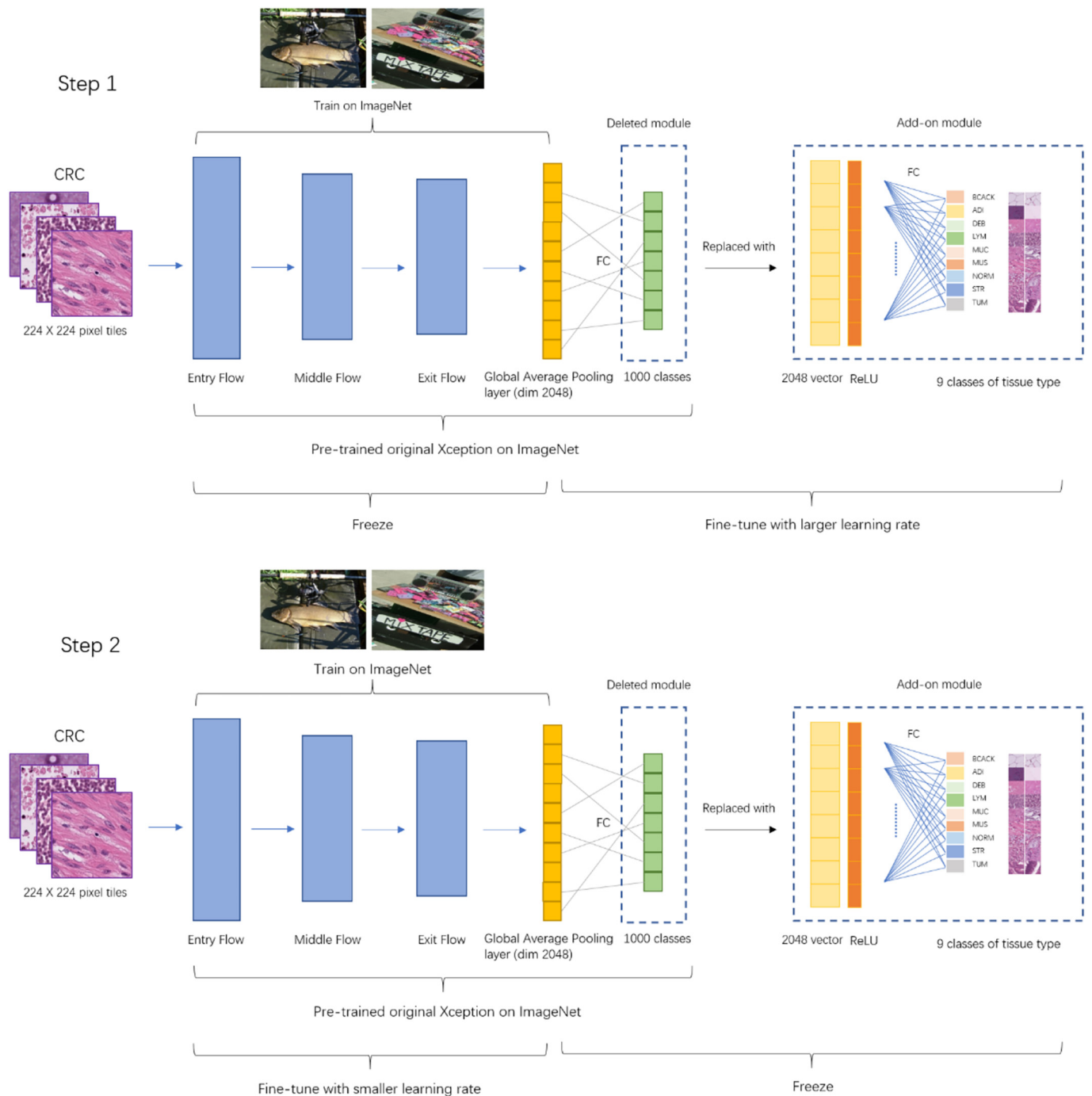
(a)

Fig. 1. (a) Flow chart of fine-tuning the Xception model. (b) The first experiment: comparison of IMGNET-2048 and FTX-2048 for tissue classification in colorectal cancer. (c) The second and third experiments: comparison of IMGNET-2048 and FTX-2048 for prediction of mRNA expression and gene mutations in lung cancer. IMGNET-2048: features extracted from the original Xception model based on ImageNet; FTX-2048: features extracted from the fine-tuned Xception model.

So far, very few fine-tuned models have been developed for digital pathology tasks; all focused on strongly supervised image classifications. Khan et al. fine-tuned an ImageNet-based model using breast histopathology images.[22] It was demonstrated that the fine-tuned model outperformed the transfer learning from ImageNet database and improved the detection of prostate cancer. In addition, a stepwise fine-tuning scheme was proposed for boosting the classification of gastric pathology image classification.[23] Most recently, Ahmed et al.[24] showed that fine-tuning improved the accuracy of Inception-V3 and VGG-16 networks on pathological images. Riasatian et al. proposed that training a deep network by a variety of
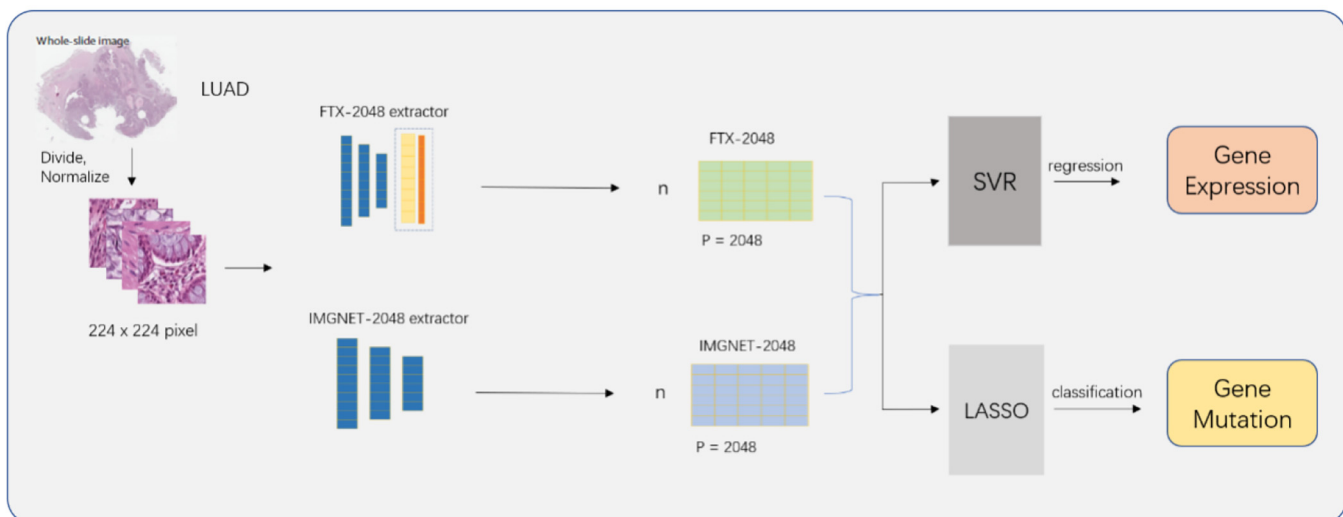
(b)
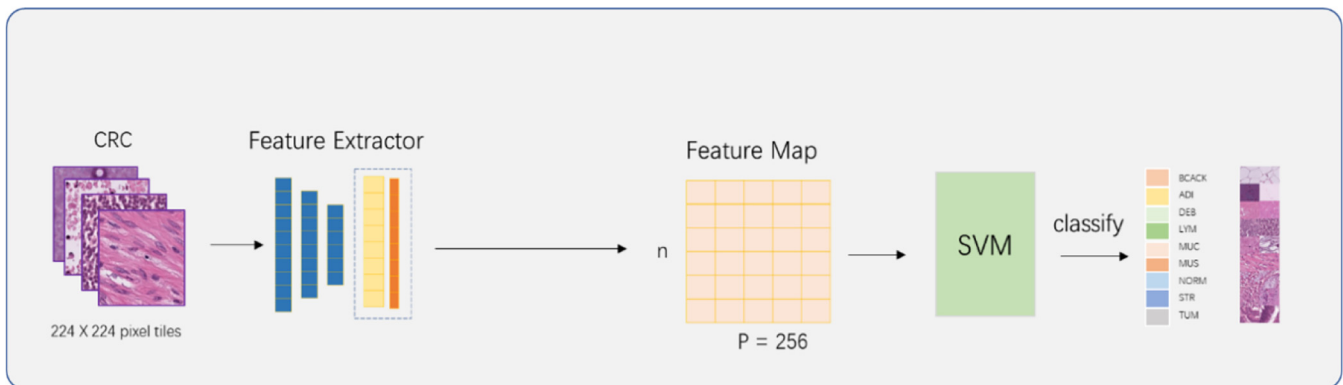
(c)

Fig. 1 (*continued*).

tumor types will provide better image features.[25] Ren et al. proposed that unsupervised domain adaptation can improve the performance of the classification task.[26] Fagerblom et al. proposed a meta-learning that could be applied to the digital pathology.[27] Finally, Dahaene et al. proposed self-supervision to improve the prediction on weakly supervised tasks.[28]

We hypothesize that fine-tuning the middle layers of pre-trained ImageNet models using histopathological images could improve feature extraction, and consequently the performance of the downstream prediction tasks, e.g., not only simple tasks like strongly supervised image classifications where pathologists' annotations are available, but also for more complex tasks such as prediction of other important clinical outcomes or molecular features (e.g., patient survival, transcriptomic expression, genomic alterations, etc.) at patient level where there is no prior knowledge regarding which regions of tissue and its appearance are predictive. So far, no fine-tuned neural networks are made available in the public domain as feature extractors.

In this work, we aimed to prove the concept that fine-tuning the ImageNet-based neural networks with histopathological images can improve the quality of extracted features and downstream predictions. We used 100 000 annotated H&E image patches for colorectal cancer (CRC) to fine-tune a pre-trained Xception model. The image features produced by the fine-tuned Xception model have at least 2 advantages. First, they boosted the performance of

strongly supervised tissue classification of CRC histological images. Second, they improved the performances of weakly supervised prediction tasks on across-cancer types (i.e., predictions of transcriptomic mRNA expression levels and genomic mutations in lung cancer patients).[1,29] The developed fine-tuned feature extractors for pathological images are available at https://github.com/1996lixingyu1996/Transfer_learning_Xception_pathology.

Material and methods

Datasets

The NCT-CRC dataset consists of 2 sub-datasets from colorectal cancer (CRC) patients, NCT-CRC-100k and NCT-CRC-7k. In total, 107 000 images (sized 224 × 224) are available and labelled with 1 of 9 classes (e.g., adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM)). The NCT-CRC-100k dataset (100 000 images) were used to fine-tune the pre-trained Xception model based on ImageNet, while NCT-CRC-7k (7180 images) was used to independently test the performance of the features extracted from the fine-tuned model and the pretrained ImageNet model for prediction of the tissue classification.[30–37]
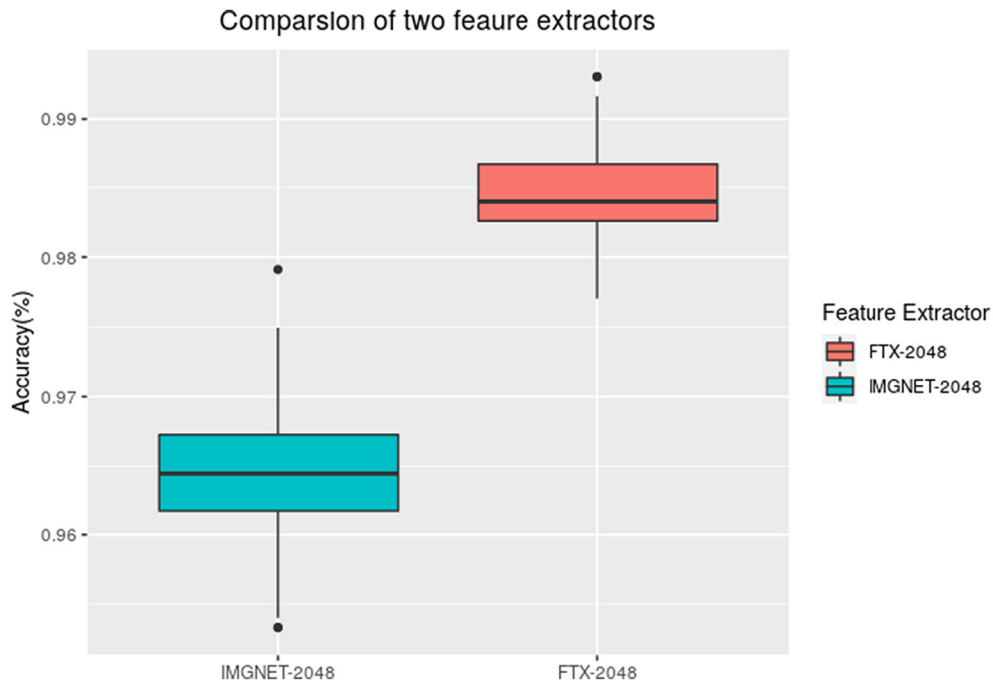
Fig. 2. Comparison of overall accuracy for tissue classification in colorectal cancer between IMGNET-2048 and FTX-2048. IMGNET-2048: features extracted from the original Xception model based on ImageNet; FTX-2048: features extracted from the fine-tuned Xception model.

The TCGA LUAD dataset included frozen whole slide images (WSIs) for patients with lung adenocarcinoma. After deleting the pathological images with very small contents and erroneous contents, the data contained a total of 344 different patients, with a total of 900 pathological images. The mRNA-expression data for TCGA LUAD were downloaded using the R package TCGAbiolink, while the mutation data were downloaded using the GDC Data Transfer Tool.[38,39]

Fine-tuning the pre-trained ImageNet model

The fine-tuning training process included 2 steps (Fig. 1a). The first step was to train an add-on module at a relatively big learning rate. We replaced the final classification layer of the original Xception model (1000 classes) with a 9-class classification layer for the 9 different tissue types for CRC, and added a fully connected layer after the original Xception GlobalAveragePooling layer
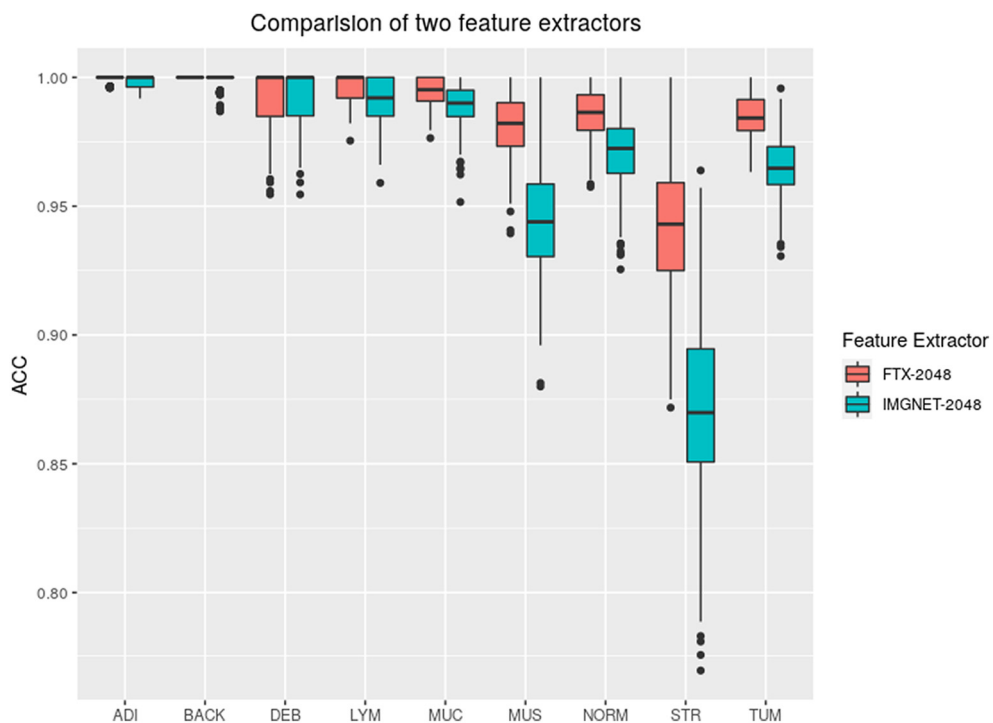

Fig. 3. Comparison of accuracy for classification of different tissue types in colorectal cancer between IMGNET-2048 and FTX-2048. IMGNET-2048: features extracted from Xception model based on ImageNet. FTX-2048: features extracted from fine-tuned Xception model.
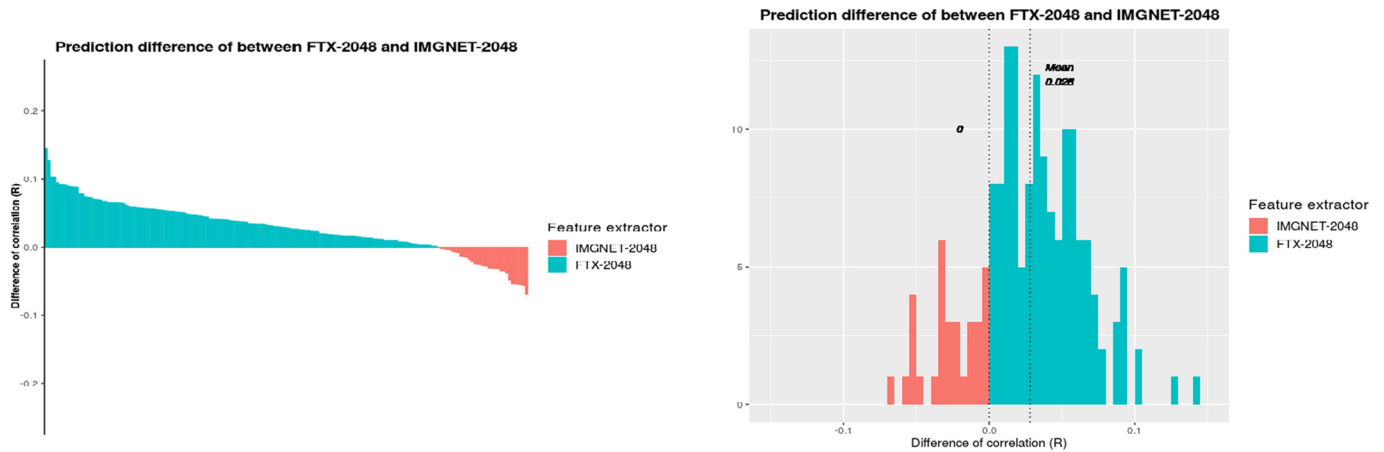
Fig. 4. Difference in correlation (observed vs. predicted mRNA-expressions for immune-related genes) between FTX-2048 and IMGNET-2048. IMGNET-2048: features extracted from Xception model based on ImageNet. FTX-2048: features extracted from fine-tuned Xception model.

(before the 9-classification layer; referred to as Add-on Module Part B thereafter). The layer before the GlobalAveragePooling is referred as Part A (pretrained on ImageNet). These 2 layers were used to incorporate histopathological images (colorectal cancer images) into the ImageNet model which was not trained using pathological images. With the new, pathological images, the fine-tuned model may be better equipped to handle similar medical images, facilitating extracting features with higher quality. We first fixed Part A with the parameters pre-trained on the ImageNet dataset and trained the add-on module Part B (2 fully connected layers with

2048 neurons, a ReLU activation layer, and 9-neuron classifier layer) with an Adam optimizer at a learning rate of 4×10^{-4} (20 iterations; batch size = 128).

The second step was to adjust the pre-trained module at a smaller learning rate. We fixed part B, which was trained during Step 1, and fine-tuned Part A using the Adam optimizer at a learning rate of 5×10^{-5} (10 iterations; batch size = 128). We extracted the 2048 neurons from the final layer of Part B of the fine-tuned Xception model as the features (FTX-2048).
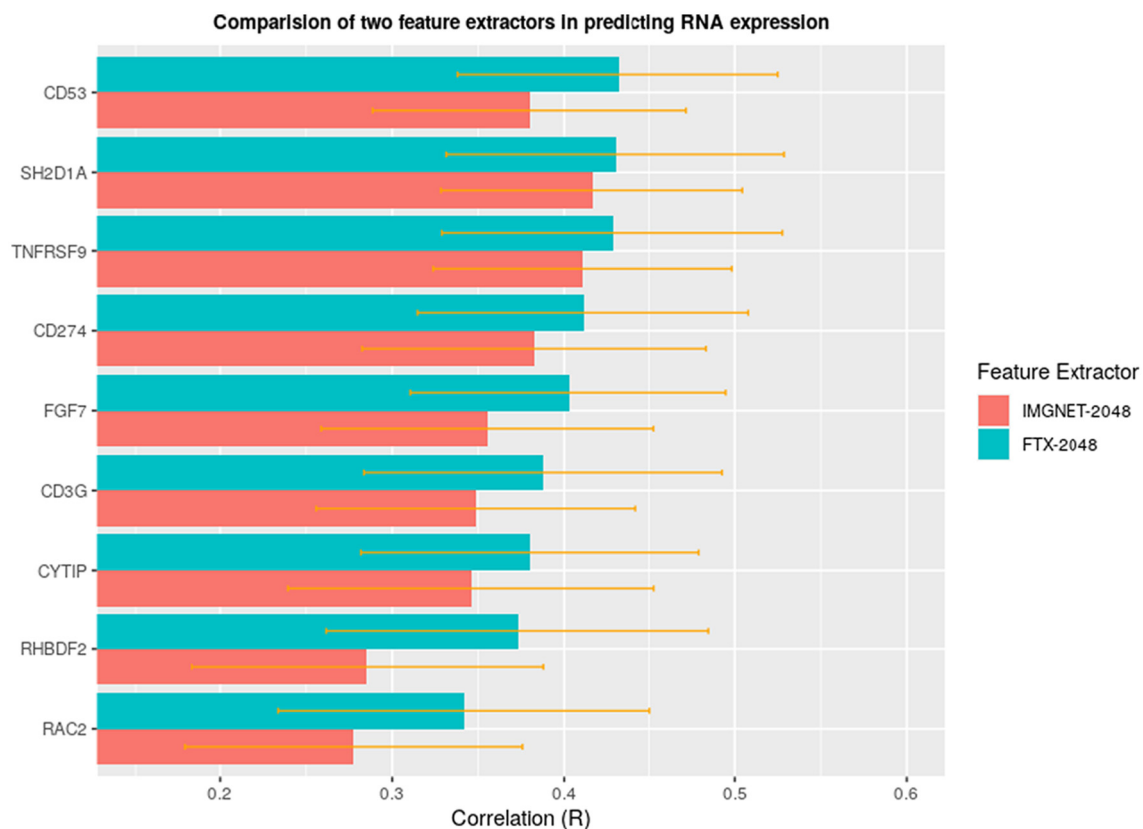

Fig. 5. Correlation between observed and predicted mRNA-expressions for selected immune-related genes (CD274, CD3G, TNFRSF9, FGF7, CYTIP, RAC2, RHBDF2, CD53, and SH2D1A) for FTX-2048 and IMGNET-2048. IMGNET-2048: features extracted from Xception model based on ImageNet. FTX-2048: features extracted from fine-tuned Xception model.

For feature extraction using the pretrained Xception model (IMGNET-2048), we replaced the last fully connected layer in the original model with a fully connected layer with 2048 neurons, a linear rectification activation function (ReLU), and a fully connected classification layer at the top. Categorical crossentropy was chosen as the loss function.

Comparison of features from pre-trained and fine-tuned models

We used the features extracted from fine-tuned Xception (FTX-2048) model, and Image-pretrained (IMGNET-2048) model and self-supervision model MoCo v2 to compare the 3 feature extractors for 2 different experiments with 5-fold cross validation (Fig. 1b–c). Each experiment was repeated 50 times.

The first experiment was to predict tissue types using the fine-tuning process with similar data and images, i.e., annotated images from CRC-HE-7K for CRC. We extracted features from the CRC-HE-7K with 2 feature extractors, FTX-2048 and IMGNET-2048. We used a multi-class classifier constructed using linear support vector classifiers to compare 2 feature extractors.

The second experiment was to predict patient-level transcriptomic expression for immune-related genes of LUAD, a tumor type different from CRC. The TCGA-LUAD whole slide images were divided into small tiles. The tiles were then standardized to 0.25 mpp and color normalized with Macenko's method.[40] FTX-2048 and IMGNET-2048 were then used to extract 2048 features from each image tile. The features of all tiles from a patient were averaged to produce a final representation of 2048 features for that patient. Gene expressions values were log transformed before modeling. Support vector regression (SVR) implemented in the R package e1071 was used to model and predict the log RNA-expression.

The third experiment was to predict cross-cancer patient-level gene mutations in LUAD. FTX-2048 and IMGNET-2048 were used to extract 2048 features from image tiles as in the second experiment. The features were averaged over different tiles of a patient. Least absolute shrinkage and selection operator (LASSO) implemented in the R package glmnet was used for model fitting and prediction for gene mutations at patient level.

Results

Classification of tissue types in CRC

We compared FTX-2048 with IMGNET-2048 in terms of prediction of tissue type for CRC where similar images were used for fine-tuning. Fig. 2 clearly demonstrates the superiority of the features extracted from the fine-tuned Xception model (FTX-2048). Overall, the extracted features with fine-tuned Xception (FTX-2048) exhibited a significantly higher accuracy (98.4%) compared to the "off-the-shelf" features extracted using the original Xception based on ImageNet database (IMGNET-2048) (96.4%, P value $= 2.2 \times 10^{-6}$). Prediction of stroma tissue has been a challenge for the pretrained ImageNet models.[1] In our present experiment, IMGNET-2048 produced a suboptimal performance for stroma tissue, and the accuracy was only approximately 87% (Fig. 3), whereas FTX-2048 remarkably improved the accuracy to 94%. Substantial improvement was also observed for the prediction of muscle (98% by FTX-2048 vs. 94% by IMGNET-2048), tumor (99% by FTX-2048 vs. 97% by IMGNET-2048), and normal tissue (99% by FTX-2048 vs. 97% by IMGNET-2048). Evident improvement of accuracy was seen for mucin and lymphocytes as well. Both FTX-2048 and IMGNET-2048 had accuracies close to 100% for adipose, debris, and background.
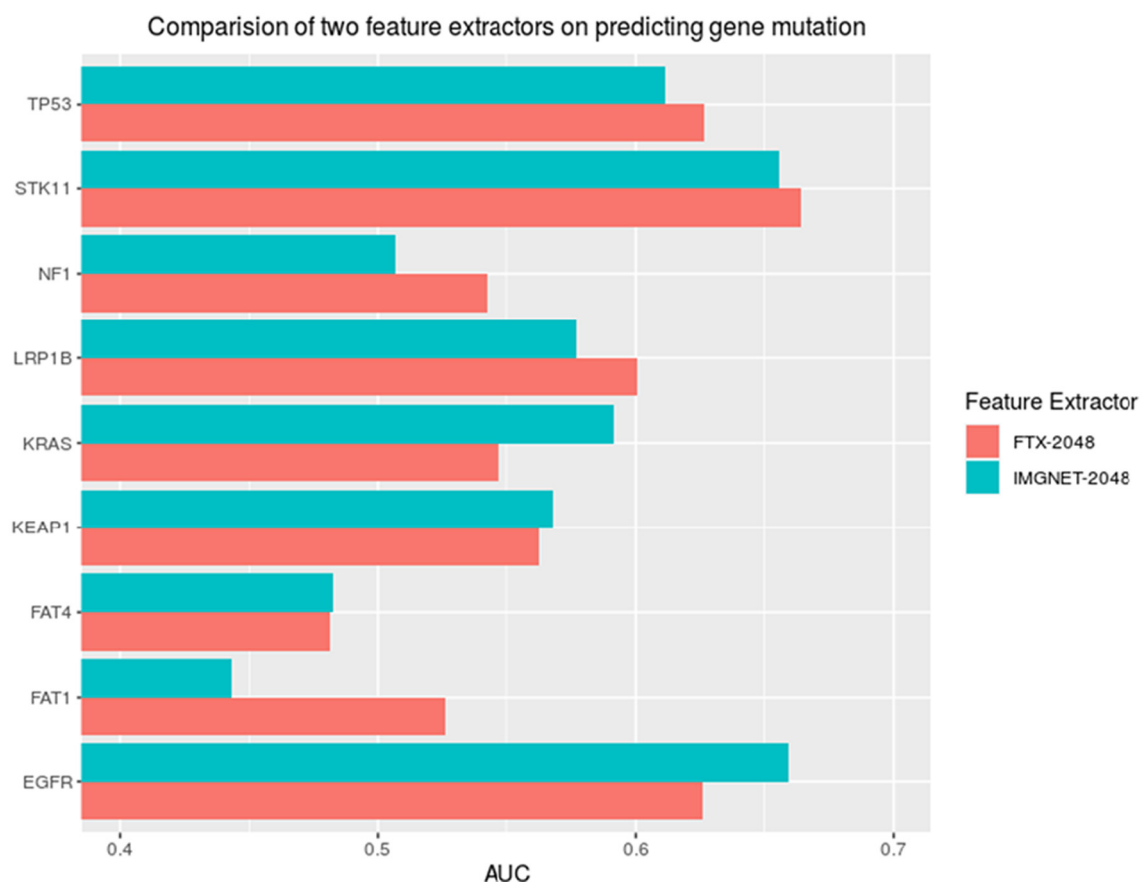
Fig. 6. Comparison of area under the curve (AUC) for prediction of mutation of nine frequently mutated genes in LUAD between IMGNET-2048 and FTX-2048. IMGNET-2048: features extracted from Xception model based on ImageNet. FTX-2048: features extracted from fine-tuned Xception model.

Prediction of mRNA expression in LUAD

The fine-tuned Xception model trained with CRC H&E images was used to extract image features for another cancer type, i.e., LUAD from TCGA, to predict mRNA expression at patient level. We applied FTX-2048 and IMGNET-2048 to randomly selected 907 immune-related genes from InnateDB's gene list. As expected, the features from FTX-2048 also outperformed the "off-the-shelf" features from IMGNET-2048 (Fig. 4). A _t_ test suggested that the correlations between mRNA expression and FTX-2048 features were significantly higher than those obtained using IMGNET-2048 features ($P < 1 \times 10^{-5}$). Fig. 5 shows examples where FTX-2048 provided a better prediction for some well-known immune genes such as CD274 (PDL1), CD3G (CD3 T cell), and TNFRSF9 (41BB), etc.

Prediction of gene mutation in LUAD

We also attempted to compare the prediction performance of FTX-2048 with IMGNET-2048 in terms of gene mutation for nine most frequently mutated genes in LUAD. Again, the features from the fine-tuned model provided a higher or similar AUC for the majority of the genes (7 out of 9 genes) (Fig. 6). The prediction performance was improved in 5 out of 9 genes (STK11, TP53, LRP1B, NF1, and FAT1). For FAT4 and KEAP1, FTX-2048 and IMGNET-2048 provided similar AUCs. On the other hand, IMGNET-2048 produced higher AUCs for EGFR and KRAS. Although no statistically significant difference in model predictions were detected ($P = 0.50$), the lack of statistical significance was mainly due to small number of genes were included in this experiment.

Model interpretability

The models in this manuscript were based on mean-pooling of feature values from different patches of individual WSIs. We examined the morphological appearance of the image patch that was the closest to the mean features of each individual patient's WSI. This revealed the difference in the morphological appearance of the "mean" image patches between FTX-2048 and IMGNET-2048, which may explain the difference in prediction performance of the 2 feature extractors. Fig. 7 illustrates the image patches closest to the mean features extracted based on FTX-2048 and IMGNET-2048 for 5 randomly selected individual WSIs.

Discussion

Cancer is a leading cause of death worldwide, accounting for 19.3 million new cancer cases and nearly 10 million deaths by the end of 2020.[41] Digital pathology has become a powerful tool in cancer research.[42,43] However, due to the lack of annotated histopathological images, the current digital pathology workflows are primarily built based on "off-the-shelf" features extracted from ImageNet models, where natural images are the predominant sources of the database.

In this work, we proved the concept that fine-tuning middle layers of the current ImageNet neural networks (e.g., Xception) with histopathology images can produce features with a higher quality and better prediction performance for not only the classification of tissue types in CRC where similar images from CRC patients were used for fine-tuning, but also gene expression and mutations in another type of tumor (LUAD), even though no images from LUAD were used to fine-tune the feature extractors. The experiments suggested that the features from the fine-tuned models possess richer pathological information than "off-the-shelf" features directly from pre-trained models based on ImageNet dataset. Furthermore, the fine-tuned features can improve the downstream prediction performance for different tasks (e.g., tissue segmentation/classification, molecular gene expression and mutations, etc) and different cancer types. Our work highlights the importance of continuing to build on the existing fine-tuned models with more annotated pathological images in the future.[44]

Code availability

https://github.com/1996lixingyu1996/Transfer_learning_Xception_pathology

Data availability

The TCGA dataset is publicly available at the TCGA portal (https://portal.gdc.cancer.gov). The public TCGA clinical data is available at https://xenabrowser.net/datapages/.

Xception model weights are available at https://github.com/fchollet/deep-learning-models/releases/download/v0.4/xception_weights_tf_dim_ordering_tf_kernels_notop.h5.

We randomly selected 907 immune-related genes from InnateDB's gene from https://www.innatedb.com/annotatedGenes.do?type=innatedb

Competing interests

The authors declare no potential conflicts of interest

Acknowledgements

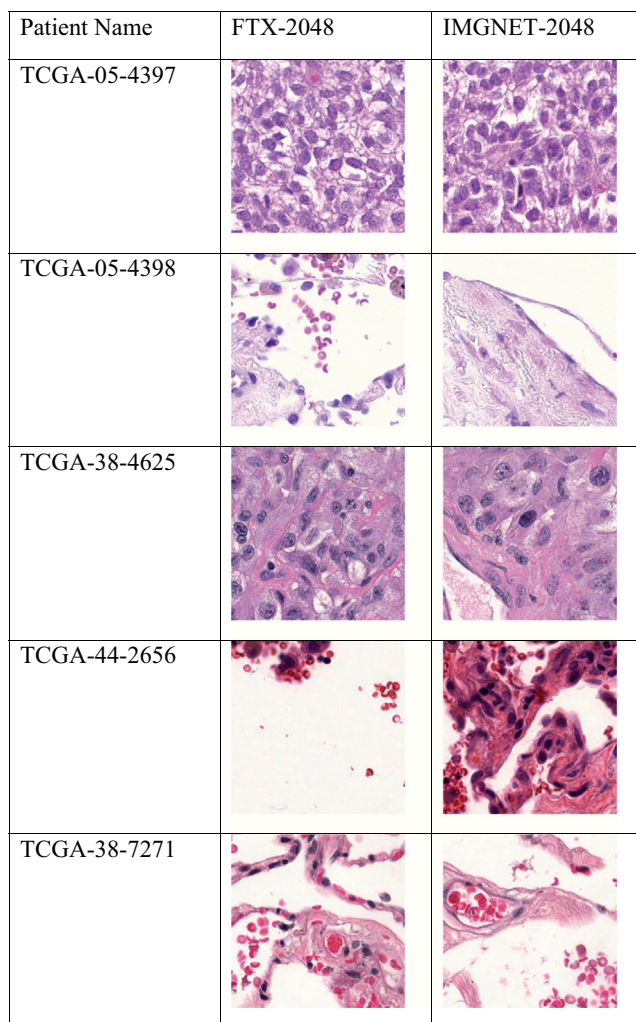| Patient Name	FTX-2048	IMGNET-2048
TCGA-05-4397		
TCGA-05-4398		
TCGA-38-4625		
TCGA-44-2656		
TCGA-38-7271		

Fig. 7. Image patches closest to the mean features extracted based on FTX-2048 and IMGNET-2048 for 5 randomly selected individual WSIs.

References

1. Kather J, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med 2019;16, e1002730.
2. Campanella G, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25(8):1301–1309.
3. Ehteshami Bejnordi B, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama 2017;318(22):2199–2210.
4. Wang D, et al. Deep Learning for Identifying Metastatic Breast Cancer. Preprint at: https://arxiv.org/abs/1606.05718 2016.
5. Hong R, et al. Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. Cell Rep Med 2021;2(9): 100400.
6. Tsou P, Wu C-J. Mapping driver mutations to histopathological subtypes in papillary thyroid carcinoma: applying a deep convolutional neural network. J Clin Med 2019:8. https://doi.org/10.3390/jcm8101675.
7. Kather JN, et al. Pan-cancerimage-based detection of clinically actionable genetic alterations. Nat Cancer 2020;1(8):789–799.
8. Chang P, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. Am J Neuroradiol 2018;39(7):1201–1207.
9. Schaumberg AJ, Rubin MA, Fuchs TJ. *H&E-stained Whole Slide Image Deep Learning Predicts SPOP Mutation State in Prostate Cancer*. 2018. https://doi.org/10.1101/064279% JbioRxiv.Preprint at: https://www.biorxiv.org/content/10.1101/064279v9.
10. Schmauch B, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun 2020;11(1):3877.
11. Sirinukunwattana K, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut 2021;70(3):544–554.
12. Courtiol P, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nat Med 2019;25(10):1519–1525.
13. Yao J, et al. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med Image Anal 2020;65:101789.
14. Mobadersany P, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci 2018;115(13), E2970.
15. Skrede O-J, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet (London, England) 2020;395(10221):350–360.
16. Deng J, et al. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009.
17. Durand T, et al. WILDCAT: weakly supervised learning of deep ConvNets for image classification, pointwise localization and segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017.
18. Durand T, Thome N, Cord M. WELDON: weakly supervised learning of deep convolutional neural networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.
19. Li X, Plataniotis KN. How much off-the-shelf knowledge is transferable from natural images to pathology images? PLoS One 2020;15(10), e0240530.
20. Echle A, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology 2020;159(4):1406–1416.e11.
21. Kather JN, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019;25(7):1054–1056.
22. Hasan Khan UA, et al. Improving Prostate Cancer Detection with Breast Histopathology Images. arXiv:1903.05769 2019.
23. Qu J, et al. Gastric pathology image classification using stepwise fine-tuning for deep neural networks. J Healthc Eng 2018;2018:8961781.
24. Ahmed S, et al. Transfer learning approach for classification of histopathology whole slide images. Sensors (Basel) 2021;21(16).
25. Riasatian A, et al. Fine-Tuning and training of densenet for histopathology image representation using TCGA diagnostic slides. Med Image Anal 2021;70:102032.
26. Ren J, et al. Unsupervised domain adaptation for classification of histopathology whole-slide images. Front Bioeng Biotechnol 2019;7:102.
27. Fagerblom F, Stacke K, Molin J. Combatting out-of-distribution errors using model-agnostic meta-learning for digital pathology. SPIE Med Imaging 2021;11603.SPIE.
28. Dehaene O, et al. Self-Supervision Closes the Gap Between Weak and Strong Supervision in Histology. arXiv:2012.03583 2020.
29. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv:1610. 02357 2016.
30. Huang G, et al. Densely Connected Convolutional Networks. arXiv:1608.06993 2016.
31. Szegedy C, et al. Rethinking the Inception Architecture for Computer Vision. arXiv:1512. 00567 2015.
32. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 2014.
33. He K, et al. Deep Residual Learning for Image Recognition. arXiv:1512.03385 2015.
34. Sandler M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801. 04381 2018.
35. Chollet F. Xception: deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017.
36. Zoph B, et al. Learning transferable architectures for scalable image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018.
37. Szegedy C, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco, California, USA: AAAI Press; 2017. p. 4278–4284.
38. Mounir M, et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS Comput Biol 2019;15(3), e1006701.
39. Zhang Z, et al. Uniform genomic data analysis in the NCI Genomic Data Commons. Nat Commun 2021;12(1):1226.
40. Macenko M, et al. A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2009.
41. Ferlay J, et al. Cancer statistics for the year 2020: an overview. Int J Cancer 2021;149(4): 778–789.
42. Bera K, et al. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019;16(11):703–715.
43. Cui M, Zhang DY. Artificial intelligence and computational pathology. Lab Invest 2021;101(4):412–422.
44. Wang D, et al. Deep Learning for Identifying Metastatic Breast Cancer. arXiv:1606.05718 2016.