MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231062

Transfer Learning for Mortality Prediction in Non-Small Cell Lung Cancer with Low-Resolution Histopathology Slide Snapshots

Matthew CLARK[a], Christopher MEYER[a], Jaime RAMOS-CEJUDO[b,c], Danne C. ELBERS[b,d], Karen PIERCE-MURRAY[b], Rafael FRICKS[e], Gil ALTEROVITZ[d,e], Luigi RAO[f,g], Mary T. BROPHY[b,h], Nhan V. DO[b,h], Robert L. GROSSMAN[a] and Nathanael R. FILLMORE[b,d,1]

[a] *Center for Translational Data Science, University of Chicago, Chicago, IL*
[b] *VA Boston Healthcare System, Boston, MA*
[c] *New York University Grossman School of Medicine, New York, NY*
[d] *Harvard Medical School, Boston, MA*
[e] *National Artificial Intelligence Institute, Dept. of Veterans Affairs, Washington, DC*
[f] *Dept. of Pathology, Walter Reed National Military Medical Center, Bethesda, MD*
[g] *Office of the Surgeon General, US Army Medical Command, Falls Church, VA*
[h] *Boston University School of Medicine, Boston, MA*

Abstract. High-resolution whole slide image scans of histopathology slides have been widely used in recent years for prediction in cancer. However, in some cases, clinical informatics practitioners may only have access to low-resolution snapshots of histopathology slides, not high-resolution scans. We evaluated strategies for training neural network prognostic models in non-small cell lung cancer (NSCLC) based on low-resolution snapshots, using data from the Veterans Affairs Precision Oncology Data Repository. We compared strategies without transfer learning, with transfer learning from general domain images, and with transfer learning from publicly available high-resolution histopathology scans. We found transfer learning from high-resolution scans achieved significantly better performance than other strategies. Our contribution provides a foundation for future development of prognostic models in NSCLC that incorporate data from low-resolution pathology slide snapshots alongside known clinical predictors.

Keywords. deep learning, transfer learning, prognosis, pathology, medical images

1. Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide, and non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer cases in the United States [1]. Prognostic models that integrate multiple sources of information to predict outcomes are of clinical interest in NSCLC due to the heterogeneity of this disease [2]. Although information sources have largely come in the form of structured clinical data, recent advances in the field of deep learning have allowed researchers to effectively leverage unstructured data, including digitized histopathology slides and other medical

[1] Corresponding Author: Nathanael Fillmore, email: Nathanael.Fillmore@va.gov

image data, to achieve strong performance on a number of prediction and classification tasks [3]. Specifically, success in use of digital pathology slides for predicting prognosis in NSCLC has been achieved through the use of architectures trained from scratch on large databases of high-resolution scans of whole slide images (WSI), such as those from The Cancer Genome Atlas (TCGA) [4].

However, clinical informatics practitioners may only have access to low-resolution snapshots of digital pathology slides, not high-resolution digital scans, and therefore, models developed using these snapshots would potentially offer a more practical path to widespread uptake. Whereas clinical data warehousing efforts have produced large databases of clinical data that can be used to define both predictors and outcomes, medical images are often omitted from these warehouses due to information technology constraints or institutional policy driven by an intended business intelligence or clinical operations use case [5]. In contrast, low-resolution snapshots may be more widely available in order to enable on-screen display or through inclusion in reports.

For example, the Veterans Affairs Precision Oncology Data Repository (VA-PODR) is a large, deidentified data repository of clinical, genomic, and imaging data on VA patients with cancer [6]. VA-PODR includes comprehensive clinical information from the VA's nationwide electronic health record system and cancer registry and molecular alteration data from targeted tumor sequencing. VA-PODR does not contain any high-resolution WSIs, because these are not widely available within the VA, but does contain low-resolution histopathology slide snapshots which were included in molecular alteration reports to physicians. Thus, while it would be infeasible develop prognostic models based on high-resolution WSIs within VA-PODR, it is potentially feasible to do so using low-resolution snapshots.

In this study, our objective was to evaluate the feasibility of training prognostic models in this scenario where only low-resolution digital snapshots of histopathology images are available, using data on NSCLC patients from VA-PODR. We compared strategies without transfer learning, with transfer learning from general domain images, and with transfer learning from publicly available high-resolution histopathology scans. We hypothesized that transfer-learning from high-resolution histopathology scans would achieve promising performance, providing a foundation for future development of prognostic models in NSCLC that incorporate data from low-resolution pathology slide snapshots alongside known clinical predictors.

2. Methods

This work's primary data source is VA-PODR [6]. In addition, two additional data sources are used for pre-training: TCGA [4], which contains publicly available data aimed at molecularly characterizing cancer tissue samples, and ImageNet [7], a database of annotated general domain images from the world-wide web. VA-PODR was accessed through the Veterans Precision Oncology Data Commons (VPODC) [8], while TCGA data was acquired through the Genomic Data Commons [9]. Analysis was conducted using deidentified data in the VPODC computational environment.

The patient population in VA-PODR is defined as patients with NSCLC, identified based on structured cancer registry data on site and histology, for whom deidentified snapshots of hematoxylin and eosin (H&E)-stained histopathology slide images are available in VA-PODR. Similarly, in TCGA, we included patients with a diagnosis of either lung adenocarcinoma (LUAD) or lung squamous cell carcinoma (LUSC) for whom H&E-stained histopathology whole-slide images were available. We excluded patients whose follow-up was censored at less than 1 year after diagnosis.

The outcome of interest was 1-year survival, starting from the date of NSCLC diagnosis. In order to derive features for prediction, we preprocessed each pathology slide image using methods defined in prior work [10]. Briefly, each image was segmented into 299x299 pixel non-overlapping tiles, and all tiles which contained more than 50% whitespace (defined by having a value <220 for all values in the RGB color space) were omitted from the dataset. Each tile from each image is associated with the same label as to 1-year survival status of the patient from which they originated. We did not use any features other than those derived from pathology images (e.g., age or stage), since the focus of this work was to determine how best to make use of pathology image snapshot-derived features, leaving integration with clinical features for future work.

Model development and evaluation was conducted using Monte Carlo validation, with 10 random splits of VA-PODR patients into train (30%), tune (30%), and test (40%) sets. In each split, we trained a neural network model of 1-year survival based on Google's Inception v3 architecture using the train set [11]. Inception v3 is a convolutional neural network which has been used successfully in many medical image tasks including in NSCLC [10,12]. Models were pre-trained using one of four strategies described below. Learning rate and batch size hyperparameters were optimized based on performance in the tune set. Mean performance evaluation measures in the test set were reported, averaged across splits, as well as 95% confidence intervals (CIs) based on 1000 bootstrap replicates per split with the percentile method. The primary evaluation measure was mean area under the receiver operating characteristic curve (AUROC), and the receiver operating characteristic (ROC) curve was plotted. Statistical significance of difference in mean AUROC between models was evaluated through bootstrapping.

We evaluated four different models based on different pre-training strategies. ImageNet Pretrained Model: We started with Google's ImageNet-pretrained model [11]. This model was then further trained with VA-PODR data as described above. TCGA Pretrained Model: We pretrained a model using TCGA data as follows. Patients were randomly allocated into train (80%) and tune (20%) sets. We did not create a separate test set in TCGA, since our purpose was only to pretrain a model for use in VA-PODR and we wanted to maximize our training data. This model was then further trained with VA-PODR data. We also evaluated two models that do not use transfer learning. VA-PODR Only Model: This model was based only on VA-PODR data, starting from randomly initialized weights, with no pretraining. TCGA Only Model: This model was the TCGA-pretrained model with no further training on VA-PODR.

3. Results

We identified 461 patients with 484 histopathology slides meeting the inclusion criteria in VA-PODR, generating 22,153 tiles. Similarly, 892 patients with 2296 histopathology slides were identified in TCGA, generating 8,250,175 tiles. Patient characteristics of both VA-PODR and TCGA are detailed elsewhere [4,6]. The pathology slide images available in VA-PODR and TCGA differ substantially. Specifically, the TCGA dataset contains conventional gigapixel scale WSIs taken at 20x magnification. In contrast, pathology slide images in VA-PODR are low-resolution snapshots of tissue slides taken at 200x magnification which generated images, on average, only 1000 pixels wide.

We initially sought to train a deep learning model to predict 1-year survival using pathology slide snapshot images from VA-PODR alone, as described in the methods (*VA-PODR Only Model*). However, we observed results no better than random guessing (mean AUROC 0.51, 95% CI 0.40–0.60; Figure 1A). We also attempted to train using VA-PODR data alone with a larger training set (80% train, 10% tune, 10% test), but did not observe any improvement.

After pretraining on TCGA data, we evaluated model performance on VA-PODR with no additional training on VA-PODR data (*TCGA Only Model*), but this also performed no better than the model trained only on VA-PODR (mean AUROC 0.53, 95% CI 0.41–0.62; Figure 1B; P=0.40). However, by pretraining on TCGA and further training on VA-PODR (*TCGA Pretrained Model*), we observed substantially better results and were able to predict 1-year survival significantly better than our VA-PODR only model (mean AUROC 0.67, 95% CI 0.57-0.78; Figure 1C; P=0.02). In contrast, pretraining with general-purpose images from ImageNet and further training on VA-PODR (*ImageNet Pretrained Model*) did not perform significantly better than our VA-PODR only model (mean AUROC 0.61, 95% CI 0.45-0.73; Figure 1D; P=0.19).

We compared our results to a random forests baseline model that includes standard clinical features commonly derived from pathology slides, specifically histology and grade, and observed mean AUROC 0.62 (95% CI 0.52-0.73), which is not significantly better than our neural network trained only on VA-PODR data (P=0.07).

4. Discussion

We compared strategies for training NSCLC prognostic models based on low-resolution digital snapshots of histopathology images from VA-PODR and found that transfer learning from publicly available high-resolution histopathology scans in TCGA achieved better performance than other approaches. Although the absolute AUC achieved remains modest (0.67), this is comparable to AUCs achieved by prior studies based on WSIs in TCGA in NSCLC [13], and modest AUC based on information derived from histopathology slides alone is not surprising, since numerous other factors are known to have prognostic value. Nevertheless, our contribution provides a necessary first step toward development of more complete models that integrate low-resolution histopathology snapshots with other known prognostic factors such as demographics, comorbidities, cancer stage, cancer treatment, and tumor genomics.

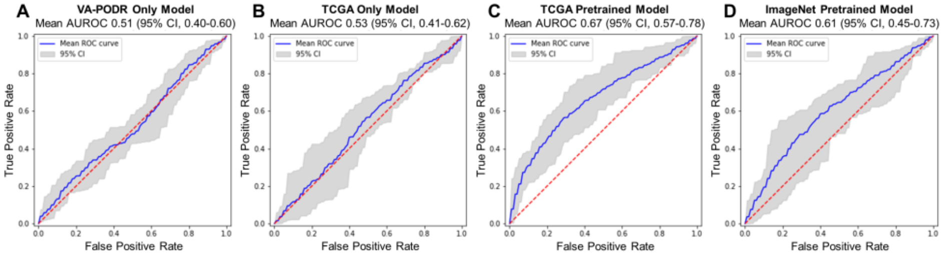

Figure 1. Receiver operator characteristic (ROC) curves in the VA-PODR test set for models without and with transfer learning. The solid blue curve shows the mean ROC curve across folds. The shaded region shows the bootstrapped 95% CI. The dashed straight red line corresponds to random guessing.

5. Conclusions

Transfer learning with a large like-domain, publicly available histopathology dataset of high-resolution WSIs from TCGA was successfully leveraged to improve performance of deep neural network models for NSCLC prognostication in a dataset of low-resolution snapshots from VA-PODR. This work provides a foundation for future work to integrate low-resolution snapshots with clinical data in prognostic models in institutions where such data is more readily available.

Acknowledgements

This work was supported by the VA Cooperative Studies Program, the Open Commons Consortium, and the Center for Translational Data Science at the University of Chicago. The views expressed are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018 Jan;68(1):7-30, doi: 10.3322/caac.21442.

[2] Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014 Aug;14(8):535-46, doi: 10.1038/nrc3775.

[3] Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, Yates LR, Jimenez-Linan M, Moore L, Gerstung M. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer. 2020 Aug;1(8):800-10, doi: 10.1038/s43018-020-0085-8.

[4] Cancer Genome Atlas Research Network; Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct;45(10):1113-20, doi: 10.1038/ng.2764.

[5] Gagalova KK, Leon Elizalde MA, Portales-Casamar E. What You Need to Know Before Implementing a Clinical Research Data Warehouse: Comparative Review of Integrated Data Repositories in Health Care Institutions. JMIR Form Res. 2020;4(8), doi:10.2196/17687.

[6] Elbers DC, Fillmore NR, Sung FC. The Veterans Affairs Precision Oncology Data Repository, a Clinical, Genomic, and Imaging Research Database. Patterns. 2020;1(6), doi:10.1016/j.patter.2020.100083.

[7] Russakovsky O, Deng J, Su H. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis Dec. 2015;115(3):211–52, doi:10.1007/s11263-015-0816-y.

[8] Do N, Grossman R, Feldman T. The Veterans Precision Oncology Data Commons: Transforming VA data into a national resource for research in precision oncology. Semin Oncol Aug-Oct. 2019;46(4-5):314–20, doi:10.1053/j.seminoncol.2019.09.002.

[9] Heath AP, Ferretti V, Agrawal S, An M, Angelakos JC, Arya R, Bajari R, Baqar B, Barnowski JH, Burt J, Catton A. The NCI Genomic Data Commons. Nat Genet. 2021;53(3):257-62, doi:10.1038/s41588-021-00791-5.

[10] Coudray N, Ocampo PS, Sakellaropoulos T. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med Oct. 2018;24(10), doi:10.1038/s41591-018-0177-5.

[11] Szegedy C, Vanhoucke V, Ioffe S. Rethinking the Inception Architecture for Computer Vision. 2016:2818–26.

[12] Kornblith S, Shlens J, Le QV. Do Better ImageNet Models Transfer Better? 2019:2656–66.

[13] Wulczyn E, Steiner DF, Xu ZY. Deep learning-based survival prediction for multiple cancer types using histopathology images. Plos One Jun. 2020;15(6), doi:10.1371/journal.pone.0233678.