Background

- 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.

- Prognostic models that integrate multiple sources of information to predict outcomes are of clinical interest in NSCLC due to the heterogeneity of this disease.

- Recent advances in the field of deep learning have allowed researchers to effectively leverage unstructured data, including digitized histopathology slides and other medical image data, to achieve strong performance on a number of prediction and classification tasks.

- 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.

Objective

- 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 Veterans Affairs Precision Oncology Data Repository (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.

Data Sources

- This work's primary data source is VA-PODR, a large data repository of clinical, genomic, and imaging data on VA patients with cancer.

- In addition, two additional data sources are used for pre-training:
    - The Cancer Genome Atlas (TCGA), which contains publicly available data aimed at molecularly characterizing cancer tissue samples
    - ImageNet, a database of annotated general domain images from the world-wide web.

- VA-PODR was accessed through the Veterans Precision Oncology Data Commons (VPODC) hosted at the University of Chicago.

- TCGA data was acquired through the Genomic Data Commons.

- Analysis was conducted using deidentified data in the VPODC computational environment.

Population

- 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.

Outcome and Predictors

- 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.
  - 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 whom 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 & Evaluation

- 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.

  - Inception v3 is a convolutional neural network which has been used successfully in many medical image tasks including in NSCLC.

- 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.

Transfer Learning

- 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 data.

Summary of Data Available

- 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.

- 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.
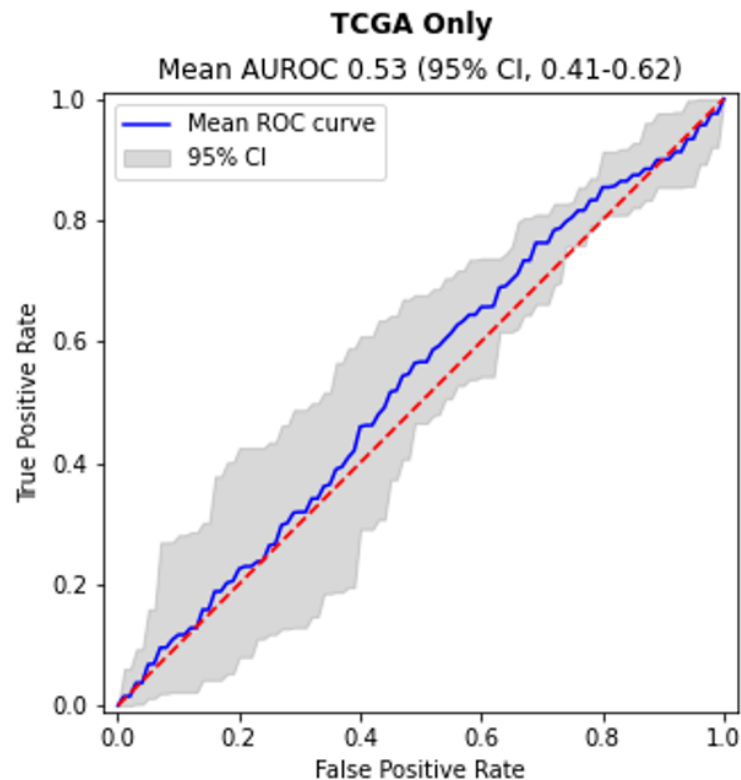
VA-PODR Only Model

- 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).

- 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.

VA-PODR Only
Mean AUROC 0.51 (95% CI, 0.40-0.60)
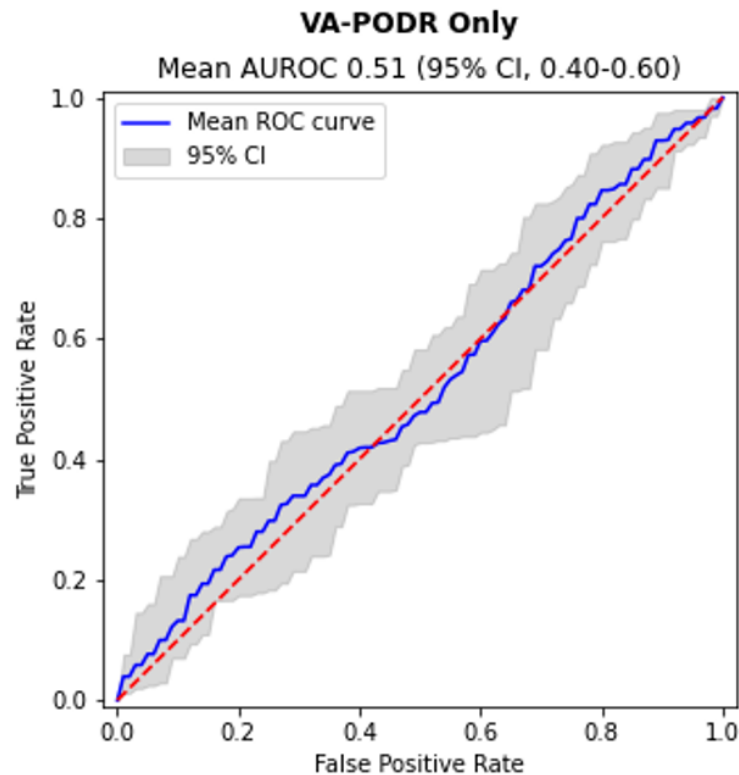
TCGA Only Model

- After pretraining on TCGA data, we evaluated our model's performance on VA-PODR with no additional training on VA-PODR data (TCGA Only Model)

- This also performed no better than our model trained only on VA-PODR (mean AUROC 0.53, 95% CI 0.41–0.62; P=0.40).

TCGA Only
Mean AUROC 0.53 (95% CI, 0.41-0.62)

TCGA Pretrained Model

- 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; P=0.02).

TCGA Pretrained, VA-PODR Tuned
Mean AUROC 0.67 (95% CI, 0.57-0.78)
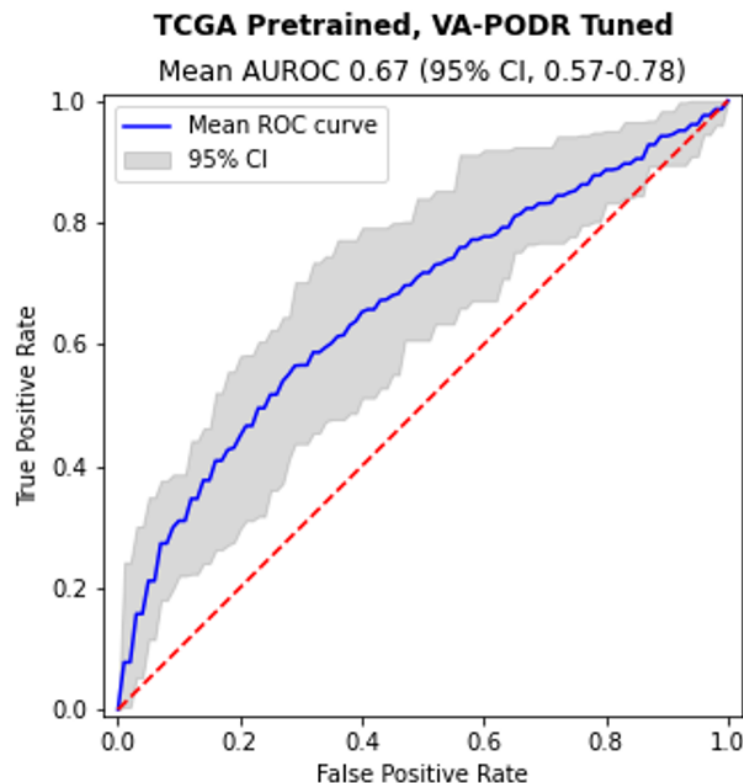
ImageNet Pretrained Model

- 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; P=0.19).

ImageNet Pretrained, VA-PODR Tuned
Mean AUROC 0.61 (95% CI, 0.45-0.73)
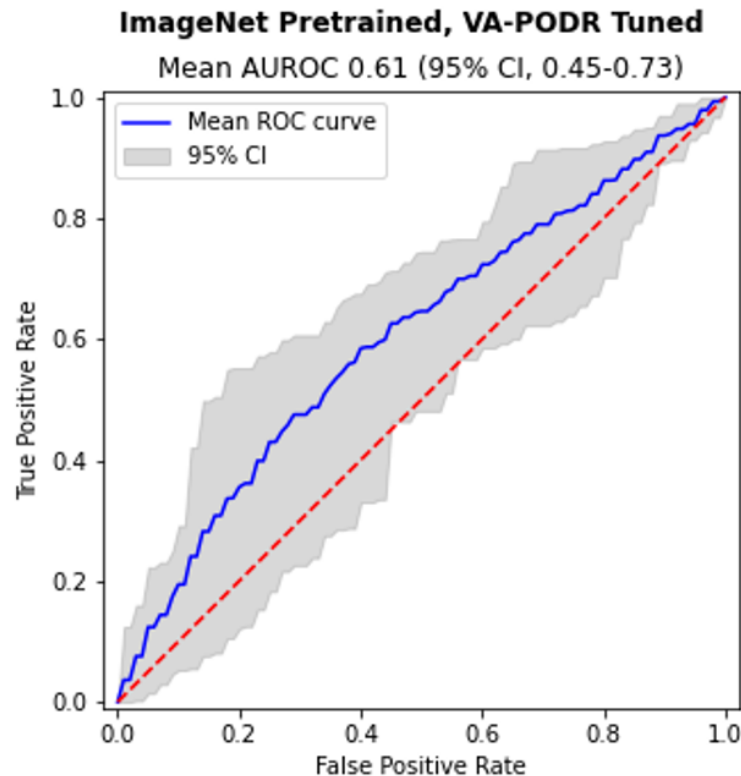
Comparison to Random Forests Baseline

Finally, 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).

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, 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.

Conclusion

- 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.

Acknowledgement

- Mathew Clark, University of Chicago

- Christopher Meyer, University of Chicago

- Robert Grossman, University of Chicago

- Danne Elbers, VA Boston Healthcare System

- Karen Pierce-Murray, VA Boston Healthcare System

- Mary Brophy, VA Boston Healthcare System

- Nathanael Fillmore, VA Boston Healthcare System

- Rafael Fricks, VA Boston Healthcare System

- Jaime Ramos-Cejudo, New York University

- Gil Alterovitz, National Artificial Intelligent Institute, Department of Veterans Affairs

- Luigi Rao, Walter Reed National Military Medical Center