Deep Learning Algorithm to Predict ERG Gene Fusion Status in Prostate Cancer

Daniel Gonzalez[2,3]; Vipul Kumar Dadhania[1]; Mustafa Yousif[1]; Jerome Cheng[1]; Saravana M. Dhanasekaran[8]; Arul M. Chinnaiyan [1,4,7,8,9]; Liron Pantanowitz[1,7#]; Rohit Mehra [1,7,8#]

[1]Department of Pathology, University of Michigan Medical School, Ann Arbor, MI
[2]Department of Pathology and Laboratory Medicine, Jackson Memorial Hospital, Miami, FL
[3]Department of Pathology and Laboratory Medicine, University of Miami Miller School of Medicine, Miami, FL
[4]Department of Urology, University of Michigan Medical School, Ann Arbor, MI
[5]Department of Radiology, University of Michigan Medical School, Ann Arbor, MI
[6]Department of Radiation Oncology, University of Michigan Medical School, Ann Arbor, MI
[7]Rogel Cancer Center, Michigan Medicine, Ann Arbor, MI
[8]Michigan Center for Translational Pathology, Ann Arbor, MI
[9]Howard Hughes Medical Institute, Ann Arbor, MI

#These authors contributed equally to this work

Daniel Gonzalez, MD
Resident (PGY-4)
Department of Pathology and Laboratory Medicine
University of Miami - Jackson Memorial Hospital
daniel.gonzalez2@jhsmiami.org

VipulKumar Dadhania, MD
Assistant Professor of Pathology
Department of Pathology
University of Michigan
NCRC Bldg 36
2800 Plymouth Rd.
Ann Arbor, MI 48109-2800
vdadhani@med.umich.edu

Mustafa Yousif, MD
Clinical Lecturer
Department of Pathology
University of Michigan
NCRC Bldg. 35
2800 Plymouth Road
Ann Arbor, MI 48109
mustafay@med.umich.edu

Jerome Cheng, MD
Assistant Professor of Pathology
Department of Pathology
University of Michigan
NCRC Bldg. 35, Rm 30-1597
2800 Plymouth Road
Ann Arbor, MI 48109
jeromech@med.umich.edu

Saravana M Dhanasekaran
Associate Research Scientist
Department of Pathology
MM Ctr fr Translational Path
MM HITS OCIO Administration
Mich Ctr for Transl Pathology
5411 CCC
Ann Arbor MI 48109-5940
saravana@umich.edu

Arul Chinnaiyan, MD, PhD
Director, Michigan Center for Translational Pathology
Professor of Pathology
2976 Taubman Health Sciences Library
1135 Catherine St
Ann Arbor, MI 48109
arul@umich.edu

Liron Pantanowitz, MD MHA
Professor of Pathology
Department of Pathology
University of Michigan
NCRC Bldg 36-1221-35
2800 Plymouth Rd.
Ann Arbor, MI 48109-2800
lironp@med.umich.edu

Rohit Mehra, MD
Professor of Pathology
Michigan Medicine
University of Michigan
Department of Pathology
NCRC Bldg 36
2800 Plymouth Rd.
Ann Arbor, MI 48109-2800
mrohit@med.umich.edu

Abstract

Background: The TMPRSS2-ERG gene rearrangement contributes to the pathogenesis of prostate cancer and plays a role in tumor multifocality and metastatic potential. ERG rearrangement in prostate cancer currently cannot be reliably identified from hematoxylin and eosin (H&E) features. Current methods for detection include immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH). We sought to develop a deep learning algorithm to identify ERG rearranged prostate adenocarcinoma based on digitized slides of H&E morphology alone.

Methods: Using the Python Keras API we developed a deep learning model for distinguishing between ERG rearranged and ERG nonrearranged prostate adenocarcinoma. The model is based on MobileNetV2 convolutional neural network architecture pre-trained on ImageNet. Weights were fine-tuned using datasets of in-house whole slide images (WSI) and The Cancer Genome Atlas (TCGA) database containing ERG Positive and Negative cases. In-house WSI were scanned at 40x using a Leica Aperio AT2 whole slide scanner. WSI's were reviewed by two pathologists, annotated using QuPath v0.2.3, and exported as 224x224 pixel sized tiles in 10x, 20x, and 40x, for input into the deep learning model. A separate model was trained for each magnification. Training and test sets for the model consisted of 268 cases (763945 tiles) and 155 cases (246060 tiles), respectively. The output of the model consisted of a prediction of ERG Positive or ERG Negative for each input tile. The ERG status for each case was determined by majority vote from the tiles comprising each case.

Results: All three models showed similar ROC curves with area under curve (AUC) results ranging between 0.768 and 0.790 (Figure 1). The sensitivity and specificity of these models were as high as 0.78 (20x model) and 1.00 (10x model), respectively. Overall case accuracy ranged between 0.75 and 0.78.

Conclusions: We demonstrate that a deep learning based AI model can successfully predict ERG fusion status in the majority of cases from H&E stained digital slides. Such a model can eliminate the need for IHC or FISH testing and thereby improve turnaround time, provide economic savings, and conserve tumor tissue when assessing for prostate tumor mutational status.

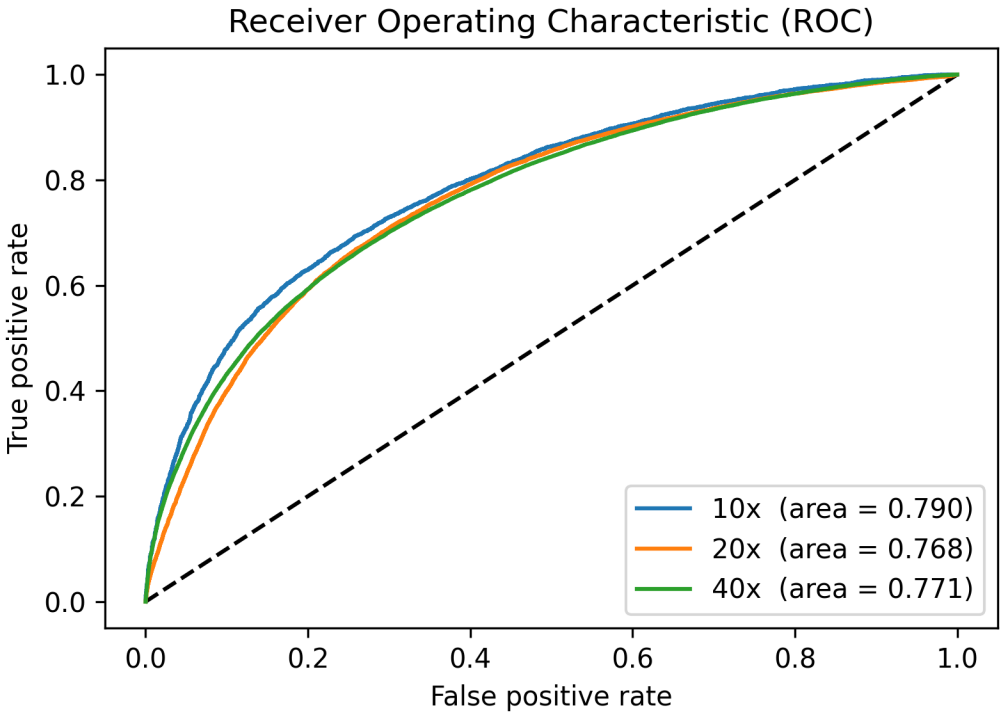

Figure 1. Receiver operating characteristics (ROC) for each model trained at 10x, 20x, and 40x magnification.