



The significance of low power versus high power microscopy in the training phase of various machine learning/artificial intelligence platforms

Ananya Datta-Mitra, Anupam Mitra, Guofeng Gao, Andrew D Jones, John P Graff, Hooman H Rashidi
Department of Pathology & Laboratory Medicine, University of California Davis Medical Center, Sacramento, California

Introduction

Accurate histopathologic interpretation is the foundation of pathologic diagnosis and is an essential component of high quality patient care. Pathologists play a critical role in the interpretation of histopathology data. However, the biggest challenges encountered in pathology laboratories throughout the world are shortages of skilled pathologists, lack of subspecialty expertise and problems with accuracy and reproducibility i.e. subjectivity amongst pathologists in interpretation of the histopathology data.

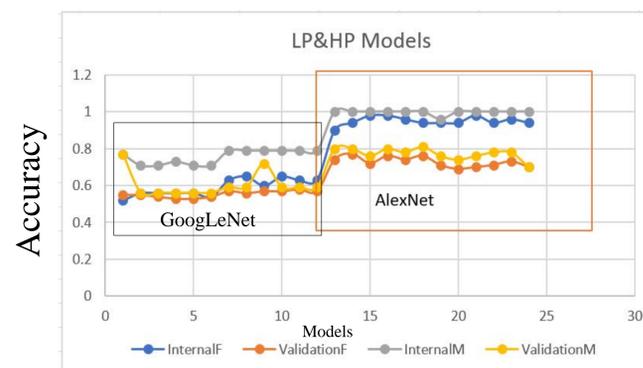
We propose that Machine learning platforms (MLP) in pathology will be an important tool that will not only benefit in accelerating a diagnosis but also increase its accuracy and correct treatment to the patients. We are developing models that would emulate a pathologist's histologic impression. Generally, a pathologist approaches a slide initially from a low power magnification (LP) that formulates a differential diagnosis followed by a high power magnification (HP) to formulate the precise diagnosis. We aim to evaluate the performance of machine learning techniques to both high and low power histologic images, specifically in the training phase of the generation of these ML/AI models.

Background

Presently, numerous 'practical' delays like generation of an H&E slide and delivery to a pathologist for review, followed by appropriate ancillary tests leads to delays in reporting a diagnosis, eventually leading to a significant delay in patient care. So the implementation of machine-learning platform may allow us to render a preliminary differential diagnosis that can aid and expedite essential followup immunohistochemistry and molecular studies before the slides reach the pathologist's microscope. This will allow a revision of the current workflow and lead to an expedited pathologic diagnosis by the pathologists. Furthermore training of the machine-learning platform with expert confirmed cases may also help to identify subtle morphologic features that can predict prognosis or presence or absence of molecular alterations.

Methods

Our aim in this study is to assess the significance of models that are trained on LP only images or HP only images of lymphoma versus normal lymph nodes utilizing 1000 images. We compared these two models groups to LP and HP image sets as well as a mixed HP and LP image set. We have tested two separate learning approaches in image analysis. Initially, we captured a mixture of HP (100x) and LP (40x) images from publicly available digital pathology image banks. To train these models, we have utilized Google's Tensorflow platform (one of the more popular ML platforms) which enabled us to train our data through two well established deep neural networks (AlexNet and GoogleLeNet).

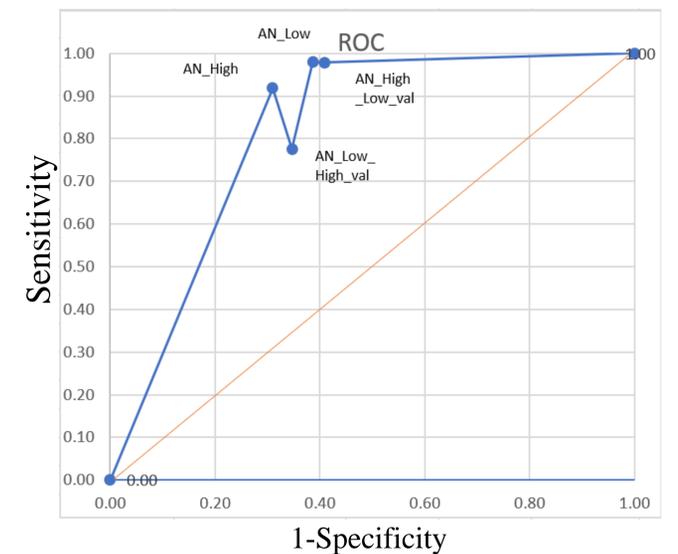


F: Final accuracy and M: maximum accuracy

Results

Comparing the results of the AlexNet and GoogleLeNet models detecting lymphoma at high and low magnification showed decent model performance with AlexNet and poor model performance with GoogleLeNet (Figure 1). When looking at the AlexNet model validation performance with LP and HP images sets showed little difference in accuracy. However when the LP and HP models are validated on a mixed set of LP and HP images both models were shown to be similar and performed reasonably well, but with slightly less accuracy on mixed images than the previous models.

Results



High_power_0 Image Classification Model

Summary

- Top-1 accuracy: 77.44%
- Top-5 accuracy: 100.0%

Confusion matrix

	LYMPHOMA	Normal LN and tonsil	Per-class accuracy
LYMPHOMA	45	26	63.33%
Normal LN and tonsil	4	58	93.55%

Discussion

We believe that our findings will help in the validation process of future ML platform applications and ultimately be instrumental in enhancing pathology practice workflow. This current study shows that more work is needed in lymphoma recognition. With the top models having accuracy above 75%, which in the field of machine learning is acceptable, but as the ROC curve and confusion matrix illustrate the false positive rate is likely not useful in clinical practice. Although the recent studies for computer-based machine-learning platforms are quite promising, the field is still in its infancy. Application of this new technology into diagnostic workflows will not only allow the H&E slide to have a pivotal role in future image analysis but also enables pathologists to continue to practice their dominant role as microscopic examiners and overseers of future machine-based interpretations.