Comparing Artificial Intelligence Platforms for Histopathologic Cancer Diagnosis

Andrew A. Borkowski, MD; Catherine P. Wilson, MT; Steven A. Borkowski; L. Brannon Thomas, MD; Lauren A. Deland, RN; Stefanie J. Grewe, MD; and Stephen M. Mastorides, MD

Two machine learning platforms were successfully used to provide diagnostic guidance in the differentiation between common cancer conditions in veteran populations.

Artificial intelligence (AI), first described in 1956, encompasses the field of computer science in which machines are trained to learn from experience. The term was popularized by the 1956 Dartmouth College Summer Research Project on Artificial Intelligence. The field of AI is rapidly growing and has the potential to affect many aspects of our lives. The emerging importance of AI is demonstrated by a February 2019 executive order that launched the American AI Initiative, allocating resources and funding for AI development. Prior to modern ML models, users would have to import many thousands of training images to produce algorithms that could recognize patterns with high accuracy. Modern ML algorithms allow for a model known as transfer learning, such that far fewer images are required for training.

Two novel ML platforms available for public use are offered through Google (Mountain View, CA) and Apple (Cupertino, CA). They each offer a user-friendly interface with minimal experience required in computer science. Google AutoML uses ML via cloud services to store and retrieve data with ease. No coding knowledge is required. The Apple Create ML Module provides computer-based ML, requiring only a few lines of code.

The Veterans Health Administration (VHA) is the largest single health care system in the US, and nearly 50,000 cancer cases are diagnosed at the VHA annually. Cancers of the lung and colon are among the most common sources of invasive cancer and are the 2 most common causes of cancer deaths in America. We have previously reported using Apple ML in detecting non-small cell lung cancers (NSCLCs), including adenocarcinomas and squamous cell carcinomas (SCCs); and colon cancers with accuracy. In the present study, we expand on these findings by comparing Apple and...
Google ML platforms in a variety of common pathologic scenarios in veteran patients. Using limited training data, both programs are compared for precision and recall in differentiating conditions involving lung and colon pathology.

In the first 4 experiments, we evaluated the ability of the platforms to differentiate normal lung tissue from cancerous lung tissue, to distinguish lung adenocarcinoma from SCC, and to differentiate colon adenocarcinoma from normal colon tissue. Next, cases of colon adenocarcinoma were assessed to determine whether the presence or absence of the KRAS proto-oncogene could be determined histologically using the AI platforms. KRAS is found in a variety of cancers, including about 40% of colon adenocarcinomas. For colon cancers, the presence or absence of the mutation in KRAS has important implications for patients as it determines whether the tumor will respond to specific chemotherapy agents. The presence of the KRAS gene is currently determined by complex molecular testing of tumor tissue. However, we assessed the potential of ML to determine whether the mutation is present by computerized morphologic analysis alone.

Our last experiment examined the ability of the Apple and Google platforms to differentiate between adenocarcinomas of lung origin vs colon origin. This has potential utility in determining the site of origin of metastatic carcinoma.

METHODS
Fifty cases of lung SCC, 50 cases of lung adenocarcinoma, and 50 cases of colon adenocarcinoma were randomly retrieved from our molecular database. Twenty-five colon adenocarcinoma cases were positive for mutation in KRAS, while 25 cases were negative for mutation in KRAS. Seven hundred fifty total images of lung tissue (250 benign lung tissue, 250 lung adenocarcinomas, and 250 lung SCCs) and 500 total images of colon tissue (250 benign colon tissue and 250 colon adenocarcinoma) were obtained using a Leica Microscope MC190 HD Camera (Wetzlar, Germany) connected to an Olympus BX41 microscope (Center Valley, PA) and the Leica Acquire 9072 software for Apple computers. All the images were captured at a resolution of 1024 x 768 pixels using a 60x dry objective. Lung tissue images were captured and saved on a 2012 Apple MacBook Pro computer, and colon images were captured and saved on a 2011 Apple iMac computer. Both computers were running macOS v10.13.

Creating Image Classifier Models Using Apple Create ML
Apple Create ML is a suite of products that use various tools to create and train custom ML models on Apple computers. The suite contains many features, including image classification to train a ML model to classify images, natural language processing to classify natural language text, and tabular data to train models that deal with labeling...
Creating ML Modules Using Google Cloud AutoML Vision Beta

Google Cloud AutoML is a suite of machine learning products, including AutoML Vision, AutoML Natural Language and AutoML Translation. All Cloud AutoML machine learning products were in beta version at the time of experimentation. We used Cloud AutoML Vision beta to create ML modules for our project. Unlike Apple Create ML, which is run on a local Apple computer, the Google Cloud AutoML is run online using a Google Cloud account. There are no minimum specifications requirements for the local computer since it is using the cloud-based architecture.

Experiment 1
We compared Apple Create ML Image Classifier and Google AutoML Vision in their ability to detect and subclassify NSCLC based on the histopathologic images. We created 3 classes of images (250 images each): benign lung tissue, lung adenocarcinoma, and lung SCC.

Experiment 2
We compared Apple Create ML Image Classifier and Google AutoML Vision in their ability to differentiate between normal lung tissue and NSCLC histopathologic images with 50/50 mixture of lung adenocarcinoma and lung SCC. We created 2 classes of images (250 images each): benign lung tissue and lung NSCLC.

Experiment 3
We compared Apple Create ML Image Classifier and Google AutoML Vision in their ability to differentiate between lung adenocarcinoma and lung SCC histopathologic images. We created 2 classes of images (250 images each): adenocarcinoma and SCC.

Experiment 4
We compared Apple Create ML Image Classifier and Google AutoML Vision in their ability to detect colon cancer histopathologic images regardless of mutation in KRAS status. We created 2 classes of images (250 images each): benign colon tissue and colon adenocarcinoma.

Experiment 5
We compared Apple Create ML Image Classifier and Google AutoML Vision in their ability to differentiate between colon adenocarcinoma with mutations in KRAS and colon adenocarcinoma without the mutation in KRAS histopathologic images. We created 2 classes of images (125 images each): colon adenocarcinoma cases with mutation in KRAS and colon adenocarcinoma cases without the mutation in KRAS.

Experiment 6
We compared Apple Create ML Image Classifier and Google AutoML Vision in their ability to differentiate between lung adenocarcinoma and colon adenocarcinoma histopathologic images. We created 2 classes of images (250 images each): colon adenocarcinoma lung adenocarcinoma.

RESULTS
Twelve machine learning models were created in 6 experiments using the Apple Create ML and the Google AutoML (Table). To
investigate recall and precision differences between the Apple and the Google ML algorithms, we performed 2-tailed distribution, paired t tests. No statistically significant differences were found (P = .52 for recall and .60 for precision).

Overall, each model performed well in distinguishing between normal and neoplastic tissue for both lung and colon cancers. In subclassifying NSCLC into adenocarcinoma and SCC, the models were shown to have high levels of precision and recall. The models also were successful in distinguishing between lung and colonic origin of adenocarcinoma (Figures 1-4). However, both systems had trouble discerning colon adenocarcinoma with mutations in KRAS from adenocarcinoma without mutations in KRAS.

**DISCUSSION**

Image classifier models using ML algorithms hold a promising future to revolutionize the health care field. ML products, such as those modules offered by Apple and Google, are easy to use and have a simple graphic user interface to allow individuals to train models to perform humanlike tasks in real time. In our experiments, we compared multiple algorithms to determine their ability to differentiate and subclassify histopathologic images with high precision and recall using common scenarios in treating veteran patients.

Analysis of the results revealed high precision and recall values illustrating the

**TABLE** Results Summary for the Apple Create ML and the Google AutoML Machine Learning Models

<table>
<thead>
<tr>
<th>Machine Learning Model</th>
<th>Classes</th>
<th>Recall</th>
<th></th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Apple, %</td>
<td>Google, %</td>
<td>Apple, %</td>
</tr>
<tr>
<td>Model 1</td>
<td>Lung normal</td>
<td>100.0</td>
<td>100.0</td>
<td>98.0</td>
</tr>
<tr>
<td></td>
<td>Lung adenocarcinoma</td>
<td>86.0</td>
<td>85.7</td>
<td>97.7</td>
</tr>
<tr>
<td></td>
<td>Lung squamous cell carcinoma</td>
<td>98.0</td>
<td>89.9</td>
<td>89.1</td>
</tr>
<tr>
<td>Model 2</td>
<td>Lung normal</td>
<td>100.0</td>
<td>100.0</td>
<td>98.0</td>
</tr>
<tr>
<td></td>
<td>Non-small cell lung cancer</td>
<td>98.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 3</td>
<td>Lung adenocarcinoma</td>
<td>90.0</td>
<td>82.4</td>
<td>91.8</td>
</tr>
<tr>
<td></td>
<td>Lung squamous cell carcinoma</td>
<td>92.0</td>
<td>92.3</td>
<td>90.2</td>
</tr>
<tr>
<td>Model 4</td>
<td>Colon normal</td>
<td>98.0</td>
<td>95.8</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Colon adenocarcinoma</td>
<td>100.0</td>
<td>100.0</td>
<td>98.0</td>
</tr>
<tr>
<td>Model 5</td>
<td>Colon adenocarcinoma KRAS+</td>
<td>72.0</td>
<td>88.2</td>
<td>69.2</td>
</tr>
<tr>
<td></td>
<td>Colon adenocarcinoma KRAS-</td>
<td>68.0</td>
<td>50.0</td>
<td>70.8</td>
</tr>
<tr>
<td>Model 6</td>
<td>Lung adenocarcinoma</td>
<td>96.0</td>
<td>91.3</td>
<td>96.0</td>
</tr>
<tr>
<td></td>
<td>Colon adenocarcinoma</td>
<td>96.0</td>
<td>100.0</td>
<td>96.0</td>
</tr>
</tbody>
</table>

*FIGURE 3 Results Using Google AutoML to Differentiate Lung Adenocarcinoma From Colon Adenocarcinoma With Precision and Recall Data for Colon Adenocarcinoma Label (Experiment 6)*
models’ ability to differentiate and detect benign lung tissue from lung SCC and lung adenocarcinoma in ML model 1, benign lung from NSCLC carcinoma in ML model 2, and benign colon from colonic adenocarcinoma in ML model 4. In ML model 3 and 6, both ML algorithms performed at a high level to differentiate lung SCC from lung adenocarcinoma and lung adenocarcinoma from colonic adenocarcinoma, respectively. Of note, ML model 5 had the lowest precision and recall values across both algorithms demonstrating the models’ limited utility in predicting molecular profiles, such as mutations in KRAS as tested here. This is not surprising as pathologists currently require complex molecular tests to detect mutations in KRAS reliably in colon cancer.

Both modules require minimal programming experience and are easy to use. In our comparison, we demonstrated critical distinguishing characteristics that differentiate the 2 products.

Apple Create ML image classifier is available for use on local Mac computers that use Xcode version 10 and macOS 10.14 or later, with just 3 lines of code required to perform computations. Although this product is limited to Apple computers, it is free to use, and images are stored on the computer hard drive. Of unique significance on the Apple system platform, images can be augmented to alter their appearance to enhance model training. For example, imported images can be cropped, rotated, blurred, and flipped, in order to optimize the model’s training abilities to recognize test images and perform pattern recognition. This feature is not as readily available on the Google platform. Apple Create ML Image classifier’s default training set consists of 75% of total imported images with 5% of the total images being randomly used as a validation set. The remaining 20% of images comprise the testing set. The module’s computational analysis to train the model is achieved in about 2 minutes on average. The score threshold is set at 50% and cannot be manipulated for each image class as in Google AutoML Vision.

Google AutoML Vision is open and can be accessed from many devices. It stores images on remote Google servers but requires computing fees after a $300 credit for 12 months. On AutoML Vision, random 80% of the total images are used in the training set, 10% are used in the validation set, and 10% are used in the testing set. It is important to highlight the different percentages used in the default settings on the respective modules. The time to train the Google AutoML Vision with default computational power is longer on average than Apple Create ML, with about 8 minutes required to train the machine learning module. However, it is possible to choose more computational power for an additional fee and decrease module training time. The user will receive e-mail alerts when the computer time begins and is completed. The computation time is calculated by subtracting the time of the initial e-mail from the final e-mail.

Based on our calculations, we determined there was no significant difference between the 2 machine learning algorithms tested at the default settings with recall and precision values obtained. These findings demonstrate the promise of using a ML algorithm to assist in the performance of human tasks and behaviors, specifically the diagnosis of histopathologic images. These results have numerous potential uses in clinical medicine. ML algorithms have been successfully applied to diagnostic and prognostic endeavors in pathology.23–28
Pathologists often use additional tests, such as special staining of tissues or molecular tests, to assist with accurate classification of tumors. ML platforms offer the potential of an additional tool for pathologists to use along with human microscopic interpretation. In addition, the number of pathologists in the US is dramatically decreasing, and many other countries have marked physician shortages, especially in fields of specialized training such as pathology. These models could readily assist physicians in underserved countries and impact shortages of pathologists elsewhere by providing more specific diagnoses in an expedited manner.

Finally, although we have explored the application of these platforms in common cancer scenarios, great potential exists to use similar techniques in the detection of other conditions. These include the potential for classification and risk assessment of precancerous lesions, infectious processes in tissue (eg, detection of tuberculosis or malaria), inflammatory conditions (eg, arthritis subtypes, gout), blood disorders (eg, abnormal blood cell morphology), and many others. The potential of these technologies to improve health care delivery to veteran patients seems to be limited only by the imagination of the user.

Regarding the limited effectiveness in determining the presence or absence of mutations in KRAS for colon adenocarcinoma, it is mentioned that currently pathologists rely on complex molecular tests to detect the mutations at the DNA level. It is possible that the use of more extensive training data sets may improve recall and precision in cases such as these and warrants further study. Our experiments were limited to the stipulations placed by the free trial software agreements; no costs were expended to use the algorithms, though an Apple computer was required.

CONCLUSION

We have demonstrated the successful application of 2 readily available ML platforms in providing diagnostic guidance in differentiation between common cancer conditions in veteran patient populations. Although both platforms performed very well with no statistically significant differences in results, some distinctions are worth noting. Apple Create ML can be used on local computers but is limited to an Apple operating system. Google AutoML is not platform-specific but runs only via Google Cloud with associated computational fees. Using these readily available models, we demonstrated the vast potential of AI in diagnostic pathology. The application of AI to clinical medicine remains in the very early stages. The VA is uniquely poised to provide leadership as AI technologies will continue to dramatically change the future of health care, both in veteran and nonveteran patients nationwide.

Acknowledgments

The authors thank Paul Borkowski for his constructive criticism and proofreading of this manuscript. This material is the result of work supported with resources and the use of facilities at the James A. Haley Veterans’ Hospital.

Author disclosures

The authors report no actual or potential conflicts of interest with regard to this article.
Artificial Intelligence and Cancer Diagnoses

Disclaimer
The opinions expressed herein are those of the authors and do not necessarily reflect those of Federal Practitioner, Frontline Medical Communications Inc., the US Government, or any of its agencies.

References

462 • FEDERAL PRACTITIONER • OCTOBER 2019
mdedge.com/fedprac

Artificial Intelligence and Cancer Diagnoses
Artificial Intelligence and Cancer Diagnoses


