Classifying Brain Tumor MRI Images and Lung and Colon Cancer Histopathological Images

Dan Zylkowski

DSC 680

Fall 2021

Abstract/Executive Summary

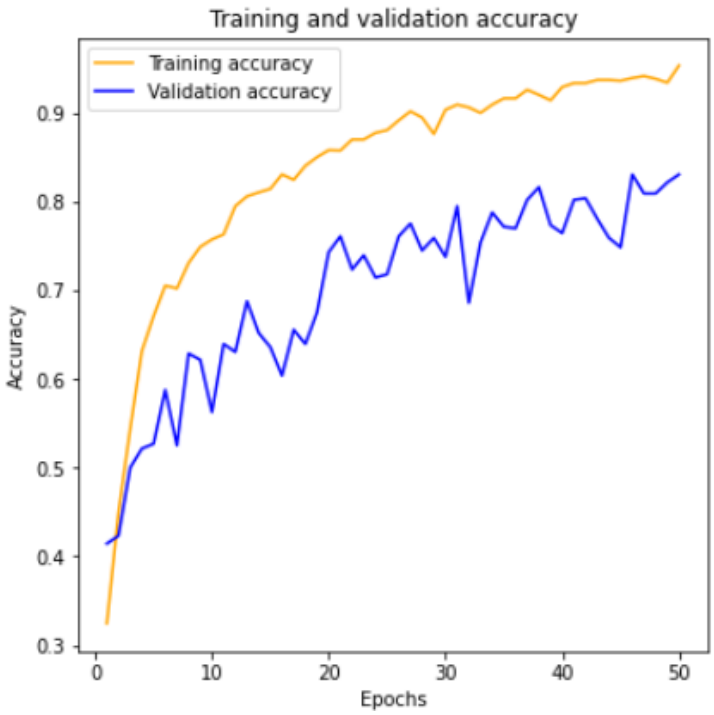
According to the American Cancer Society estimates [15], for the year 2021 Lung and Colon cancer will account for approximately 20% of all cancer diagnoses, and approximately 30% of all cancer deaths. Early detection and diagnosis of cancerous tumors are the most critical factors for the successful treatment of cancer. A tumor is an abnormal growth or mass of cells, but not all tumors are cancerous. Medical imaging is one way to identify cancerous tumors, and two common medical imaging methods used are Magnetic Resonance Imaging (MRI)[10] and histopathology [14]. Manually identifying the different types of cancerous tumors can be a difficult task.

The project uses three different image datasets. The first dataset is a brain tumor MRI image dataset [13] where each image is one of four types of tumors: glioma, meningioma, pituitary tumor, or benign. The second dataset [9] contains histopathological images of Lung tissue from three types of tumors: adenocarcinoma, squamous cell carcinoma, or benign. The third dataset [9] contains histopathological images from two types of Colon tumors: adenocarcinoma and benign. We will therefore review the results from each dataset separately.

Brain Tumor MRI image classification

For the brain tumor MRI image classification task, a deep learning convolutional neural network (CNN) was built from scratch and trained on 2,296 training images and 574 validation images. The images were distributed evenly across the 3 cancer tumor classes, while the benign tumor class contained half as many images. Accuracy was the metric chosen to optimize, as we are trying to optimize the correct classification of the images. After 50 training epochs, the validation accuracy reached 83%, but when evaluated on the test set, the accuracy was only 56% which makes it unusable for classifying tumors.

Figure 1: Baseline CNN model training and validation accuracy for Brain Cancer tumor classification



To improve on the baseline score two additional models were built using the fastai[6] python library: one on top of the pre-trained ResNet50[1] model architecture, and one on top of

the pre-trained VGG19 model architecture. Since deep-learning models are highly repurposable we can take the ResNet50 and VGG19 models that were trained on a large-scale dataset and reuse them on our dataset with only a few changes. Adding a few layers on top of the pre-trained ResNet50 model, the validation accuracy increased up to 92.6% after training only 5 epochs. Adding a few layers on top of the pre-trained VGG19 model, the validation accuracy increased up to 96.8% after training only 8 epochs.

Figure 2: VGG19 validation accuracy results for Brain Cancer tumor classification

epoch	train_loss	valid_loss	accuracy
0	0.391056	0.322775	0.878049
1	0.310011	0.263938	0.902439
2	0.238178	0.200687	0.933798
3	0.181757	0.211982	0.926829
4	0.129686	0.236787	0.933798
5	0.095988	0.128574	0.959930
6	0.063498	0.130134	0.965157
7	0.043327	0.121494	0.968641

Colon Cancer Histopathological Image Classification

For the Colon Cancer image classification task, a CNN was built from scratch and trained on 8,000 training images and 2,000 validation images. The images were distributed evenly across the adenocarcinoma and benign classes. After 20 training epochs, the validation accuracy reached 99.8%. While there wasn't much room for improvement, additional models

were built on top of the ResNet50 and VGG19 model architectures, and the validation accuracies improved to 99.95% for both models after only 3 training epochs.

Figure 3: Baseline CNN model training and validation accuracy for Colon Cancer classification

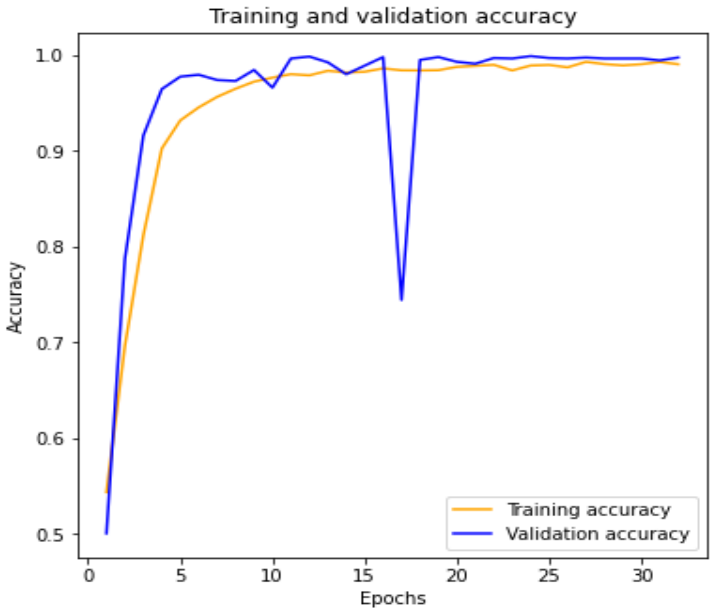


Figure 4: ResNet50 validation accuracy results for Colon Cancer classification

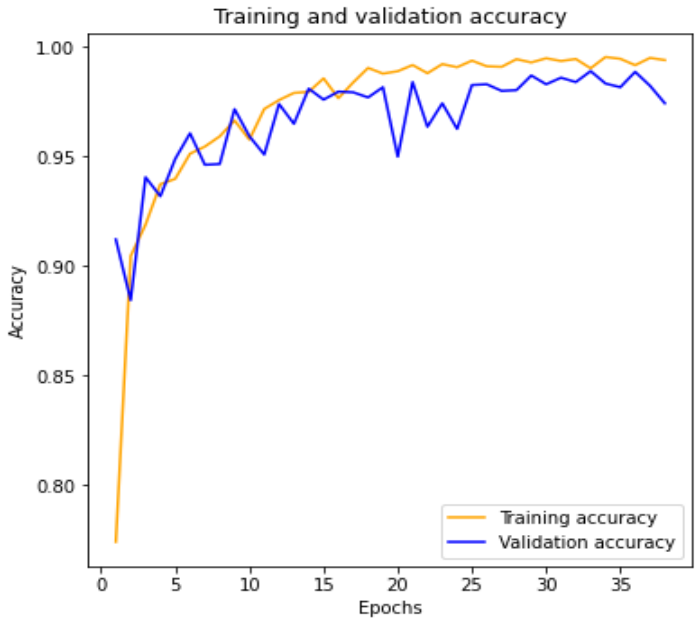
epoch	train_loss	valid_loss	accuracy	time
0	0.024488	0.005115	0.998000	02:44
1	0.016173	0.003897	0.998000	02:44
2	0.015715	0.001926	0.999500	02:45

Lung Cancer Histopathological Image Classification

For the Lung Cancer image classification task, a CNN was built from scratch and trained on 12,000 training images and 3,000 validation images. The images were distributed evenly across the adenocarcinoma, squamous cell carcinoma, and benign classes. After 25 training

epochs, the validation accuracy reached 98.2%. Additional models were built on top of the ResNet50 and VGG19 model architectures, and the validation accuracies improved to 99.8% and 99.9% respectively after only 4 training epochs.

Figure 5: Baseline CNN model training and validation accuracy for Colon Cancer classification



Problem Statement

The purpose of this project is to examine if a deep learning model can be created to accurately identify if tumor images can be accurately classified by cancer type, or if the tumor is benign. If successful, the model could be used as both a pre-screener and post-screener for the images. Pre-screening could be performed immediately after the image was developed (before being viewed by a doctor) and could allow for the identification of images that need immediate review. Post-screening images (after images are classified by a doctor) could ensure the classification is consistent.

Methods

Building a baseline CNN model

Using the Keras library, I built a CNN model from scratch using several convolutional and max-pooling layers. For the Brain MRI and lung cancer datasets, the problem was a multi-class classification problem, so I used the categorical cross-entropy loss function when compiling the model and then the softmax function for the final activation function. For the colon cancer dataset, the problem was a binary classification problem, so I used the binary cross-entropy loss function during compiling and the sigmoid function for the final activation function.

When using neural networks for image classification, a large amount of data is often required. This is because neural networks can contain thousands or millions of parameters that need to be trained. To help with the lack of data, I used image augmentation and dropout to synthetically add additional data to the current dataset. Common augmentation techniques include flipping, rotating, scaling, cropping, and translating current images to create new images. The neural network will consider the augmented images as being new data, and this will help train the network to learn from all different parts of the image. Augmentation was only performed on the Brain Cancer image dataset, since both the Colon and Lung Cancer image datasets were already augmented.

Building models on top of the ResNet50 and VGG19 model architecture

Using the fastai library, I froze the convolutional layers of the ResNet50 and VGG19 models and added dense layers on top of it. The idea behind this strategy is to use transfer learning to keep all the generalized skills of the ResNet50 and VGG19 models, and to train a few

dense layers on top of it using the dataset. The learning rate was set to a very low value, since we are only trying to fine-tune the model performance. After building a model on top of the ResNet50 and VGG19 architectures, I used fastai to identify images with the largest loss values. These images could be reviewed by an imaging specialist to examine if they are labeled correctly or reclassify them if they are misclassified.

Limitations

As with any use case for machine learning in medicine, we need to remember that a model's prediction is no substitute for the diagnosis of a trained medical professional. Therefore, machine learning should only be used in conjunction with the diagnosis of a medical professional, rather than as a unilateral method of diagnosis. Additionally, the models can only identify the images of cancers that they have been trained on, so any new cancers would not be correctly classified.

Discussion/Conclusion

While the fastai models built from the Colon and Lung Cancer image datasets had extremely high accuracies (99.5%+), the models built from the Brain Cancer image dataset were not as high (96.8%). The CNN models built from scratch in Keras also performed very well on the Colon and Lung Cancer image datasets, but poorly on the Brain Cancer MRI images, and this is likely due to the high resolution of the images when compared to the Brain Cancer MRI images. The results indicate that the models could be a useful tool for pre-screening or post-screening cancer images, but only if the cancer is one of the previously labeled cancer types.

References:

1. Boesch, G. (2021, August 29). *Deep residual networks (ResNet, RESNET50) - guide in 2021*. viso.ai. Retrieved October 3, 2021, from https://viso.ai/deep-learning/resnet-residual-neural-network/.
2. Brownlee, J. (2019). *Better Deep Learning.*
3. Brownlee, J. (2019, August 19). *Crash course in Convolutional Neural Networks for machine learning*. Machine Learning Mastery. Retrieved October 3, 2021, from https://machinelearningmastery.com/crash-course-convolutional-neural-networks/.
4. Chollet François. (2018). *Deep learning with python*. Manning Publications Co.
5. Chollet, F. (n.d.). *The keras blog*. The Keras Blog ATOM. Retrieved October 3, 2021, from https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html.
6. Fast Ai. · Making neural nets uncool again. (n.d.). Retrieved October 3, 2021, from https://www.fast.ai/.
7. Kaushik, A. (2020, February 26). *Understanding the VGG19 architecture*. OpenGenus IQ: Computing Expertise & Legacy. Retrieved November 1, 2021, from https://iq.opengenus.org/vgg19-architecture/.
8. Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H.-J., & Kim, N. (2019, December). *Deep learning in medical imaging*. Neurospine. Retrieved November 1, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6945006/.
9. Larxel. (2020, April 14). *Lung and colon cancer histopathological images*. Kaggle. Retrieved November 14, 2021, from https://www.kaggle.com/andrewmvd/lung-and-colon-cancer-histopathological-images.
10. *MRI for cancer*. American Cancer Society. (n.d.). Retrieved November 13, 2021, from https://www.cancer.org/treatment/understanding-your-diagnosis/tests/mri-for-cancer.html.
11. Osinga, D. (2018). *Deep learning cookbook practical recipes to get started quickly*. O'Reilly.
12. Radhakrishnan, P. (2019, October 26). *What is transfer learning?* Medium. Retrieved November 1, 2021, from https://towardsdatascience.com/what-is-transfer-learning-8b1a0fa42b4.
13. Sartaj. (2020, May 24). *Brain tumor classification (MRI)*. Kaggle. Retrieved November 14, 2021, from https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri.
14. Wikimedia Foundation. (2021, August 7). *Histopathology*. Wikipedia. Retrieved November 13, 2021, from https://en.wikipedia.org/wiki/Histopathology.
15. *Cancer facts and statistics Powerpoint presentation*. (n.d.). Retrieved November 19, 2021, from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-statistics-presentation-2020.pptx.

Appendix

Figure 6: Confusion Matrix for Colon Cancer image classification VGG19 model

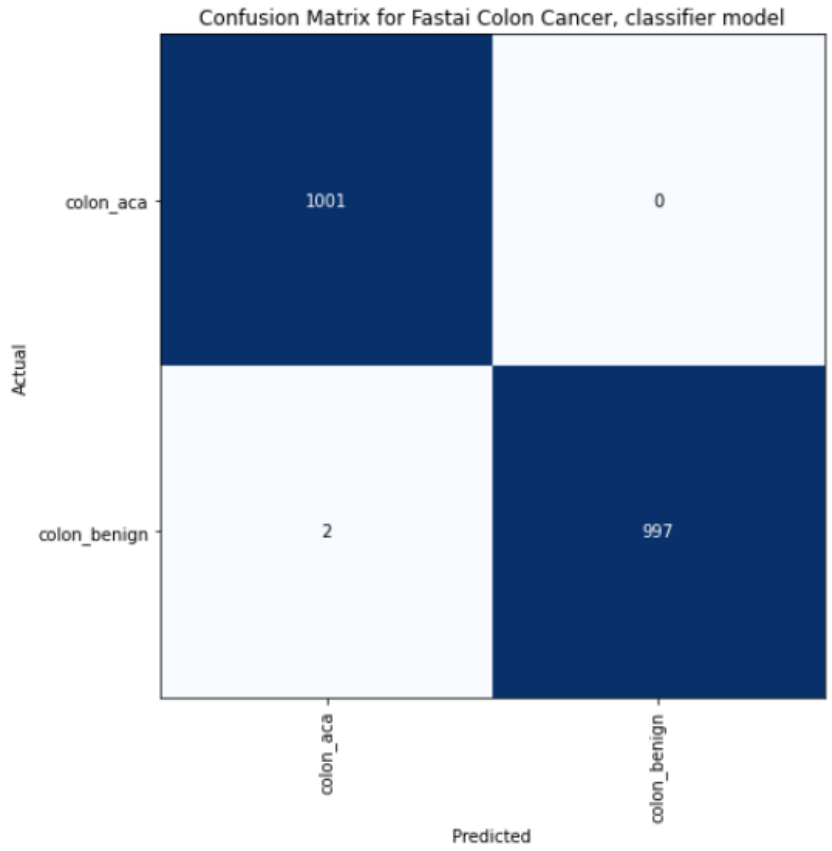
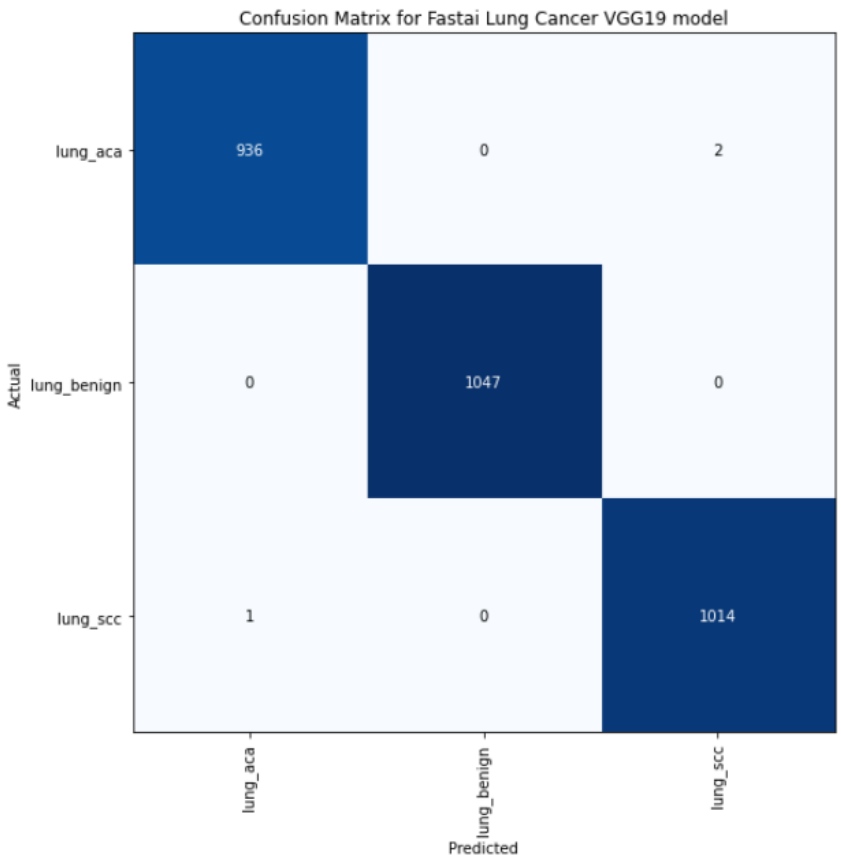


Figure 7: Confusion Matrix for Lung Cancer image classification VGG19 model


Confusion Matrix for Fastai Lung Cancer VGG19 model
lung_aca
lung_benign
lung_scc
936
0
2
0
1047
0
1
0
1014
Actual
lung_aca
lung_benign
lung_scc
Predicted