

Abstract

Manual medical image analysis by health professionals is a time-intensive process that can be automatized and optimized with the application of machine learning algorithms. We apply a convolutional neural network (CNN) on a dataset of colorectal histology images with the ultimate aim of classifying images for tumor recognition. The dataset contains 5000 smaller images that each represent one of eight tissue categories (tumor, stroma, complex, lympho, debris, mucosa, adipose, empty). These are extracted from the 10 larger images that contain multiple tissue types. We perform data augmentation and survey several popular CNNs, adjusting hyperparameters to explore effects on accuracy in classifying the various tissue types. The highest-performing model after augmentation was a VGG-16 with a test set accuracy of 94.6%. The high performance of the CNN indicates such data is well-suited for classification by machine learning algorithms such as those employed. The future of such applications in the medical field is promising and could result in significant reductions in time spent on diagnoses.

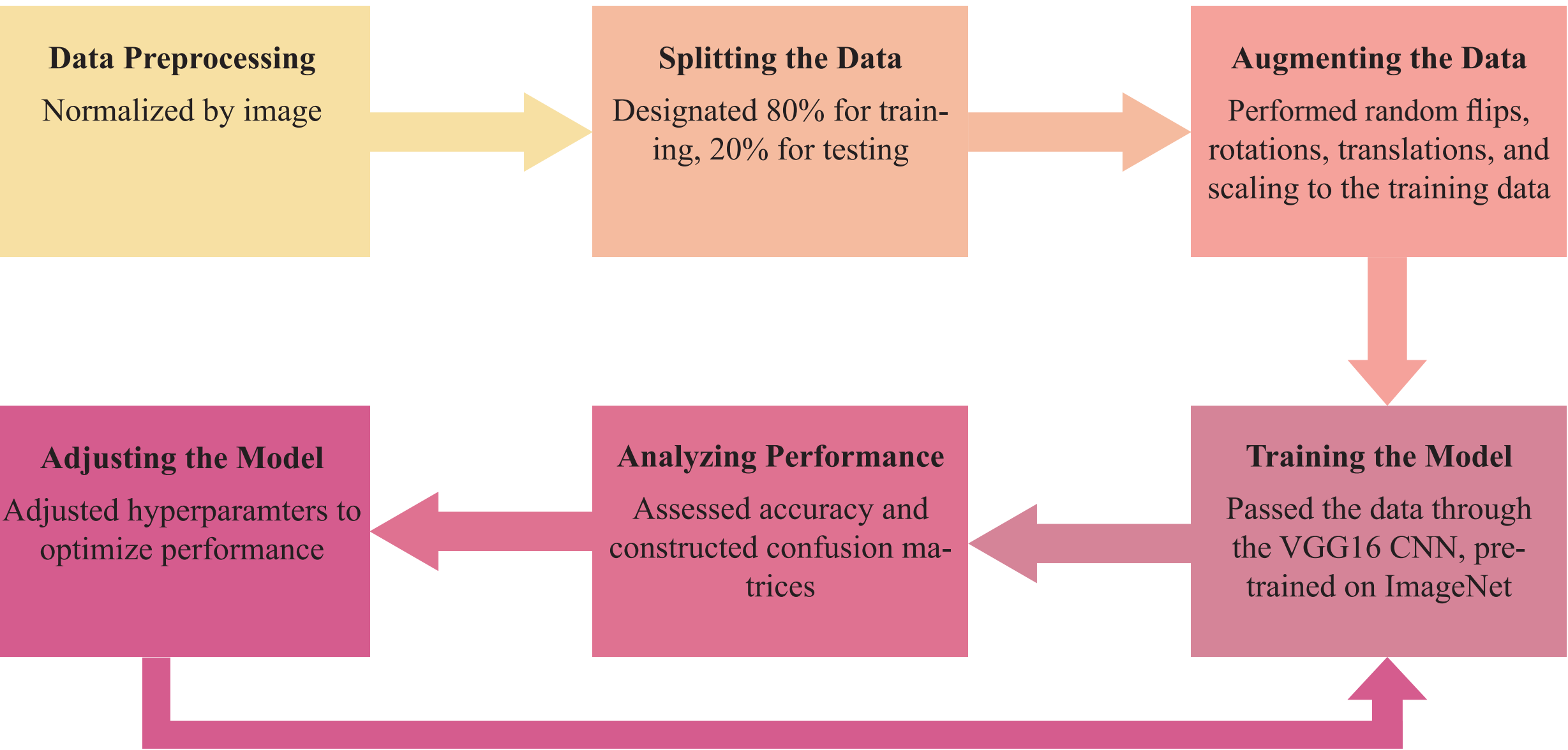
Objectives

- Accurately classify the smaller image tiles into one of the eight tissue classes
- Identify which classes are most commonly confused
- Perform segmentation of the larger images to determine the presence of the various tissue types

Methods

Data Augmentation

The initial dataset of 5000 smaller texture patches, each belonging to one of eight tissue categories, was first normalized by image before undergoing augmentation. During data augmentation, every training image batch was augmented with randomly chosen transformations that included left/right flips, up/down flips, slight rotations, translations and scaling. Each image resulted in the production of 3 new images that were the result of a combination of random transformations performed on each. In total, this resulted in the training batch tripling in size. The augmented data was then passed through VGG16 CNN model pre-trained on ImageNet.



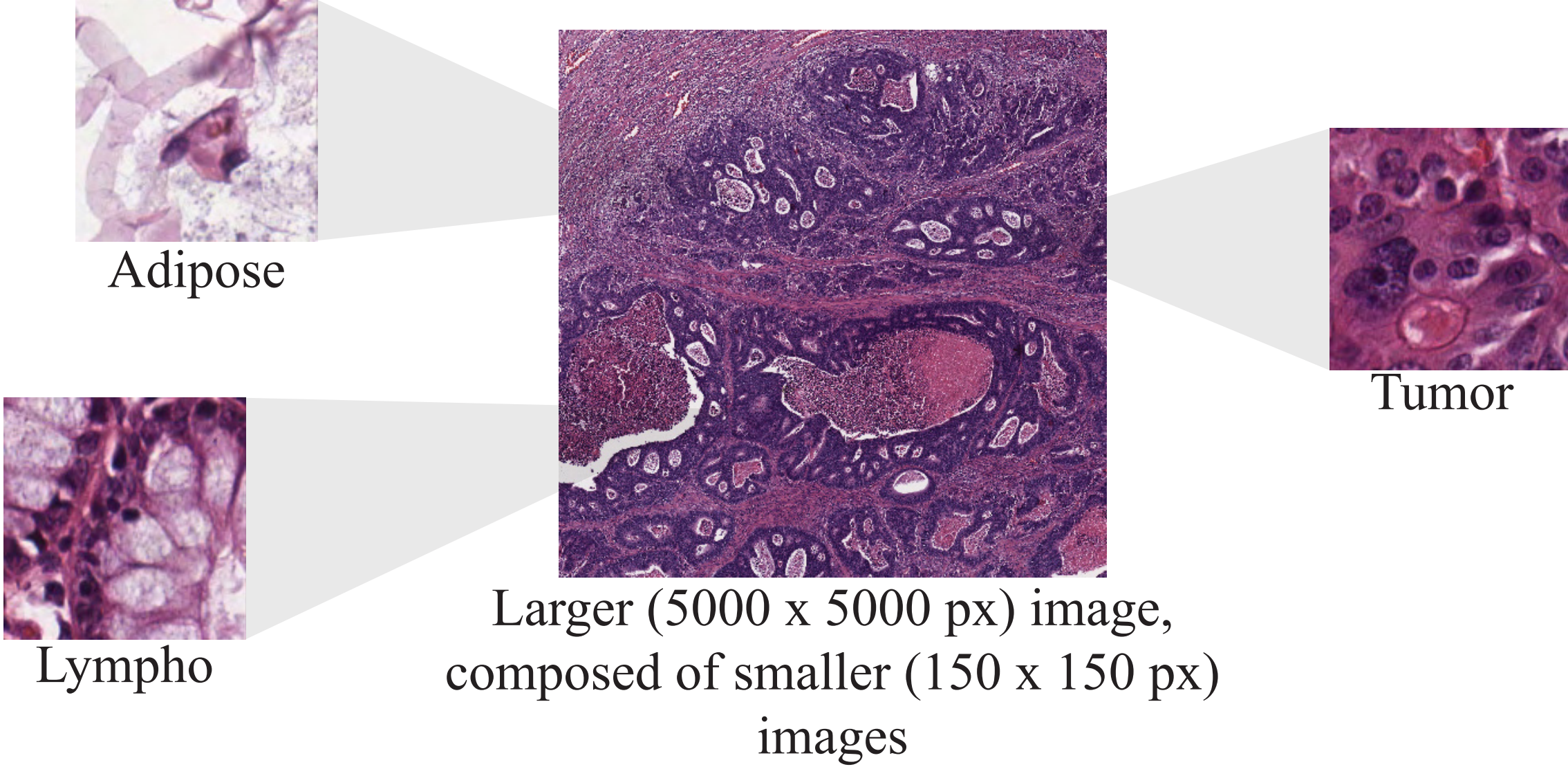
Deep Learning for Colorectal Histology Tissue Classification

Ameer Syedibrahim, Barbara Xiong,
Nicole Moiseyev, Xinghong Tang, Yuqin Shen

Methods (continued)

Segmentation of the Larger Images

To apply our model to unlabeled test sets of 10 larger images of size 5000 x 5000 px, we slice those larger images into smaller images with size 150 x 150 px with a slicing window of 150 px as step size and save them in sequence. There are images that are not in the shape of 150 x 150 px. We exclude them from our prediction because they cannot fit in the input data form. Therefore, for each larger image, we have 1089 sliced smaller images in total.



Results

When compared to the pre-augmented VGG16 model, which had a 90.68% accuracy, the augmented VGG16 model produced a 94.6% accuracy. Thus, we observe that data augmentation aided the model’s ability to perform multi-label classification.

Further, the confusion matrix produced by the model with an augmented batch of images depicts how the model rarely conflated labels between images. This was a significant improvement from our pre-augmented model which appeared to conflate ‘STROMA’ and ‘COMPLEX’ and ‘STROMA’ and ‘DEBRIS’ quite a bit. See Figure 1.

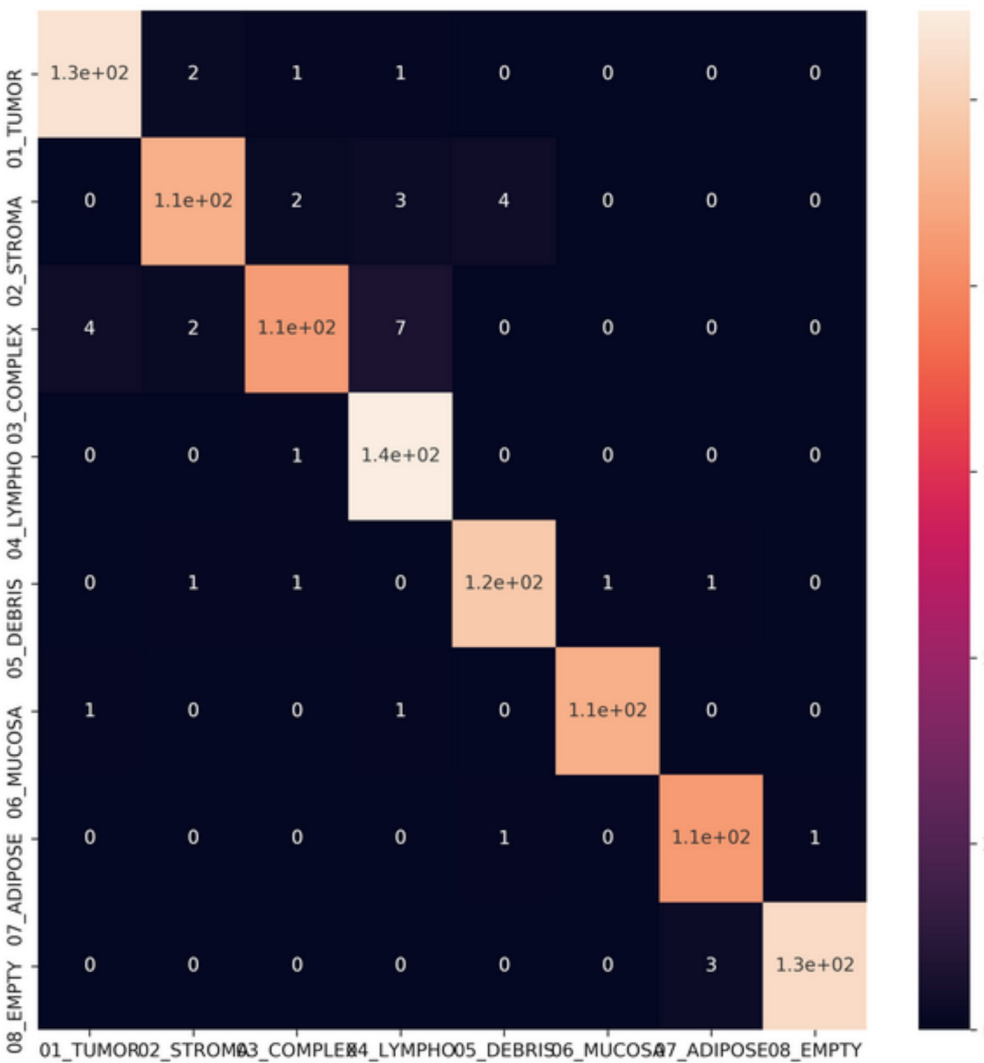


Figure 1: VGG16 classification confusion matrix.

Results (continued)

Table 1: Results of Various Models before Data Augmentation.

Model Type	# of Epochs	Image Size	Adam LR	Validation Accuracy
ResNet50	50	128 x 128	1E-4	81.38%
ResNet50	100	128 x 128	1E-4	82.82%
ResNet50	200	128 x 128	1E-4	84.62%
ResNet50	200	128 x 128	1E-5	86.72%
ResNet50	250	128 x 128	1E-5	87.20%
VGG16	50	128 x 128	1E-5	90.68%

Table 2: Results of VGG16 after Data Augmentation.

Model Type	# of Epochs	Image Size	Adam LR	Validation Accuracy
VGG16	50	160 x 160	1E-5	94.62%
VGG16	50	160 x 160	1E-5	94.6%

* Early stopping was invoked on all models.

We expect that for the predictions of sliced smaller images, they can form a 2D prediction matrix with dimension (33,33), ruling out the images with size not equal to 150 x 150 near the edge. The shape of the output matrix is aligned with that of the original image. For a better visualization of our result, we convert the matrix into a color map with 8 different colors indicating 8 different classes. We expect to see clusters of colors, indicating the clusters of different tissue classes. Refer to Figure 2.

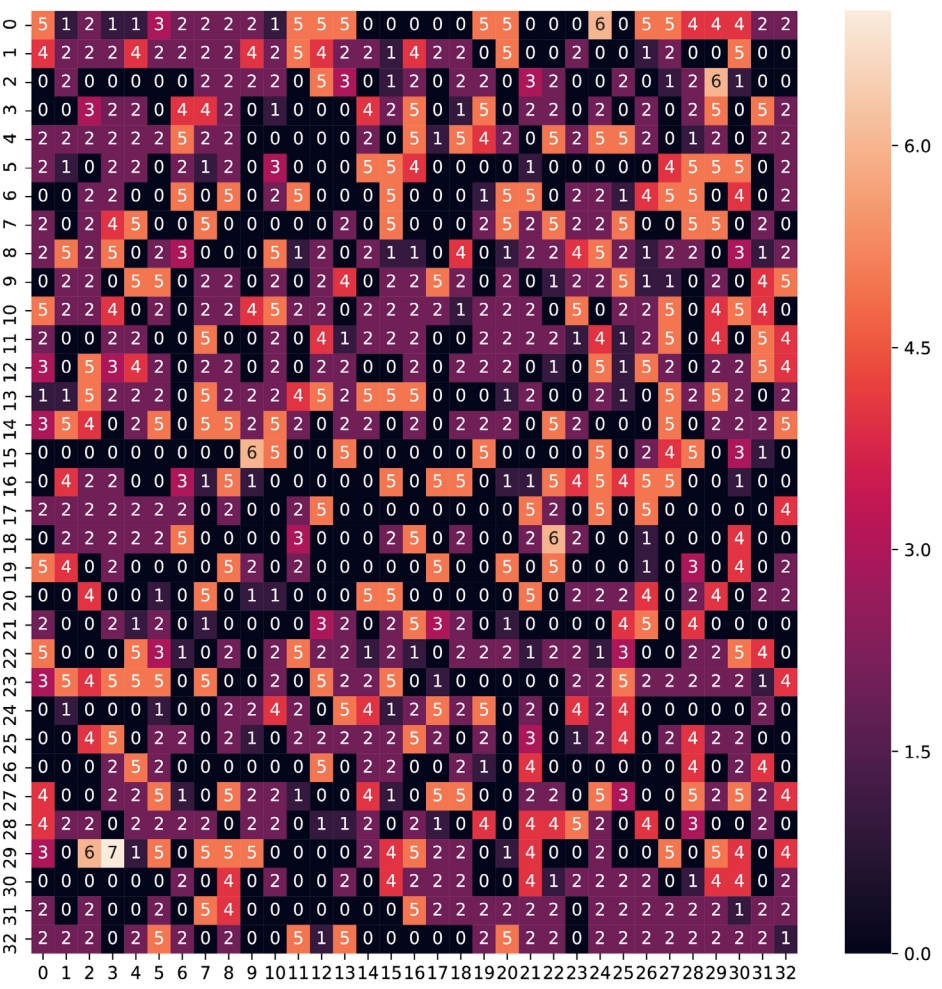


Figure 2: VGG-16 Large image prediction matrix.

Conclusion

VGG16 achieves an accuracy of 90.68%. Data augmentation specifically works well for our dataset and helps to increase the accuracy of VGG16 to 94.6%. The clustering color map of the predictions on the larger image suggests that our classification model actually works well on the new dataset, indicating a way to classify multiple classes (tissues and tumors) within a single large image.

References

Kather, J. N. et al. Multi-class texture analysis in colorectal cancer histology. Sci. Rep. 6, 27988; doi: 10.1038/srep27988 (2016).

Ponizio, Francesco et al. Colorectal Cancer Classification using Deep Convolutional Networks - An Experimental Study. BIOIMAGING (2018).

Acknowledgements: This work was supported by the Duke Plus DS (+DS) Program. We gratefully acknowledge Professor David Carlson for his continued guidance and support throughout the project. We thank Duke Research Computing for the use of the Duke Compute Cluster for high-throughput computation. The Duke Data Commons storage is supported by the National Institutes of Health (1S10OD018164-01). Health Data Science at Duke is supported by the Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002553.

Student Showcase: +DS Projects in Medicine and Health Data Science Internship Program

April 25, 2019, Duke University, Gross Hall