Detecting nodular basal cell carcinoma in pathology imaging using deep learning image segmentation

Jeannie Ren

ABSTRACT

With over 4.3 million new cases in the U.S. every year, basal cell carcinoma (BCC), is the most common form of skin cancer. Pathologists must examine pathology images to diagnose BCC, potentially resulting in delay, error, and inconsistency. To address the need for standardized, expedited diagnosis, we created an automated diagnostic machine to identify BCC given pathology images. In MATLAB, we adapted a deep neural network image segmentation model, U-Net, to train on BCC images and their corresponding masks, which can learn to highlight these nodules in pathology images by outputting a computer-generated mask. We trained the U-Net on one image from the dataset and compared the computer-generated mask output from testing on three types of images: an image from a different region of the same image taken with the same microscope, an image from a different tissue sample with a different microscope, and an image taken with a confocal microscope. We observed good, medium and poor results, respectively, illustrating that performance depends on the similarity between test and training data. In subsequent tests using data augmentation, we achieved sensitivity of 0.82±0.07 and specificity of 0.87±0.16 on N = 6 sample sections from 3 different BCCs imaged with the same microscope system. These data show that the U-Net performed well with a relatively few number of training images. Examining the errors raised interesting questions regarding what the errors mean and how they possibly arose. By creating a surgeon interface for rapid pathological assessment and machine learning diagnostics for pathological features, the BCC diagnosis process will be expedited and standardized.

Keywords: Basal Cell Carcinoma, skin cancer, deep learning, machine learning, image segmentation, automated diagnosis

1. INTRODUCTION

This research is concerned with using deep learning, a facet of machine learning, to automate nodular basal cell carcinoma detection.

1.1 Machine learning

Artificial intelligence (AI) has flourished in today's data-driven world. It has been applied as a solution in fields ranging from self-driving vehicles, to language translation, to facial recognition, and even to spam filtering.

The subset of AI that deals with developing computer code to learn a task without explicit instruction is called machine learning (ML). The solutions to the fields mentioned above all use machine learning to implement AI. ML has its own subset called Deep Learning, which relies on incredibly powerful Neural Networks (NN). These NNs replicate the human brain's structure , where computational nodes map loosely to human neurons. Deep learning is specifically ML that happens to use NNs with more than one hidden layer.

It is well documented that particular NNs are suited to particular tasks. For instance, convolutional networks are particularly suited to image related tasks.[3] The basic idea underpinning convolutional networks involves reading an image as a matrix of pixels and then performing various operations on the pixels to extract a working model based on many training images.

A particular class of image problems are relevant to the field of automated medical diagnosis, which is the diagnosis of disease in an automated, machine-based way. Usually this involves feeding pathology images retrieved from biopsies or scans to a trained ML model that will then give a probabilistic result of a classification (e.g. determining the likelihood of the image being of normal versus cancer). It can also, however, involve image

segmentation, or the differentiation of an image into its objects or features (object recognition often falls under image segmentation), for example, determining which part of an image is the tumor. This latter question is the one we are interested in exploring in this research.

1.2 Basal cell carcinoma

Basal cell carcinoma (BCC) is the most common form of skin cancer. Originating in the basal epidermis, a highly-proliferative region in the skin, BCC is caused by long-term exposure to ultraviolet (UV) radiation from sunlight and is diagnosed using digital pathology, which in turn is analyzed by pathologists. Nodular basal cell carcinoma, known for its distinctive nodes in pathology imaging, create uniform masks and serve as an example of deep learning's applications in translational technology.

Highlighting tumors in pathology images is incredibly useful for creating a standardized, expedited process by eliminating any potential human error in the diagnosis of BCC.

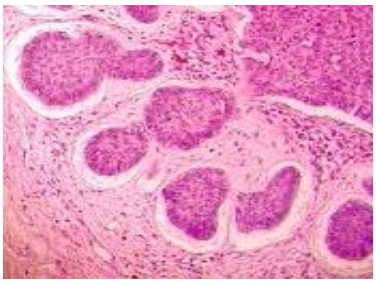
2. METHODS

2.1 Data collection

In order to train a machine learning image segmentation model, training image data and correlating ground truth masks must be used. A mask is an image that has a single value (e.g. 1 or 2) designating each pixel as in a particular region (e.g. normal or cancer), which acts as a label in segmentation algorithms.

The training image data consists of nodular basal cell carcinoma pathology images that were found on Google. See Fig. 1 for an example.

This dataset was then processed using a MATLAB algorithm that allowed the user to manually crop each image into smaller, more specific square regions of the larger, original image, which allowed for a way to increase the data by producing more high quality and varied images of distinct BCC nodes. See Fig. 2.



To standardize the training data before mask creation, the cropped, hand selected images were scaled to be 1500x1500x3 pixel JPEGs. We created a mask labeller that enabled a user to generate 1127x1127x1 square pixel gold standard masks of 0 and 1 for background and BCC, respectively. By making the images larger than the U-Net requirement of 1127x1127 pixels, possible discrepancies along the edges of the tumor masks can be validated and checked manually in an effort to preserve the shape and cropped edge of the tumors within the masks. See Fig. 3.

Each mask accounted for one tumor or continuous shape within an image. See Fig. 4 for an example. The images and masks are collected and created, respectively, and used for training with a U-Net.

Figure 1. Example image of nodular basal cell carcinoma.

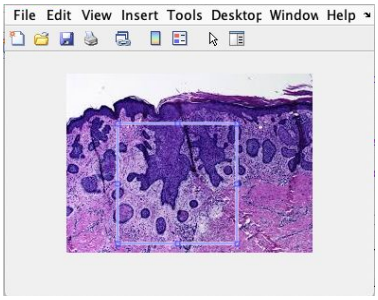


Figure 2. Manually cropped images into smaller, more specific square regions of the larger, original image to increase variety within the dataset.

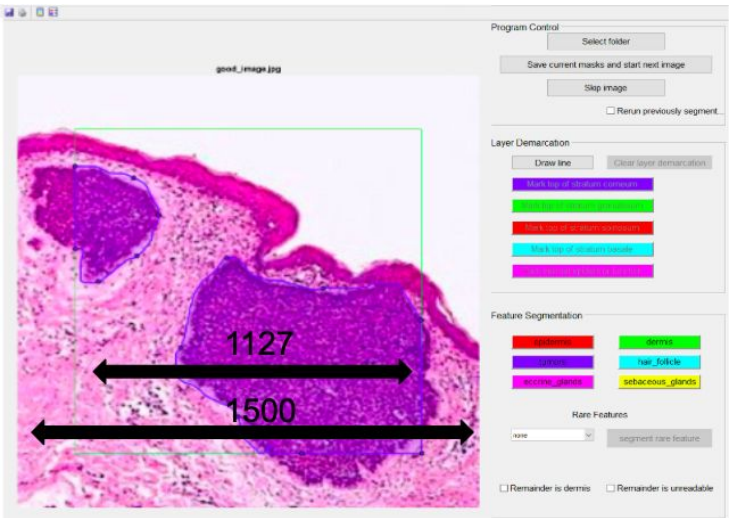


Figure 3. Mask labeller allows hand-labelled, smaller masks of 1127x1127 pixels for training.

In order to know how the result worked across variation within a single sample, or variation in cellular morphology of disease and normal, we tested a new dataset of images of larger, higher quality BCC. Each image was then cropped into n sub-images, and the model trained on n-1 sub images and tested on nth image. See Fig. 5.

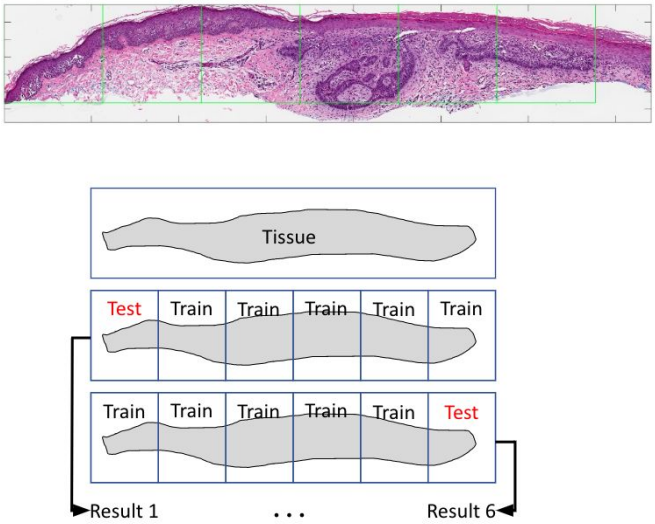


Figure 5. Image subdivision (top) and training rotation (bottom) where a result is generated for each ttest segment by training on the other segments in the same sample.

2.2 U-Net

U-Net is the state-of-the-art architecture for image segmentation in medical imaging. With two paths, the U-Net architecture begins with the contracting path (or encoder) to "capture context and a symmetric expanding path that enables precise localization."[2]

Segmentation was performed using encoder-decoder network adapted from the well-studied UNet (1). The input image and output mask size were both 1127X1127 pixels with the input being a 3-channel image and output being a single channel binary mask. Dropout set to 0.5 was used on two of the deep layers in the encoder path.

Data augmentation during training included rotation, width and height shift, shear, zoom, and horizontal flip. Image pixels were normalized to the range [0,1] for training and testing. The model was trained for 3 epochs, with each epoch consisting of 300 augmented images. This encoder consists of standard convolutional and max pooling layers found in convolutional neural networks. The second path is known as the symmetric expanding path (or decoder), which, using transposed convolutions that "enables precise localization."[2] As a result, the U-Net is a Fully Convolutional Network , which given an input image tile, is able to produce an output segmentation map that predicts each pixel's class. Using MATLAB, the U-Net was adapted to fit binary class parameters. See Fig. 6.

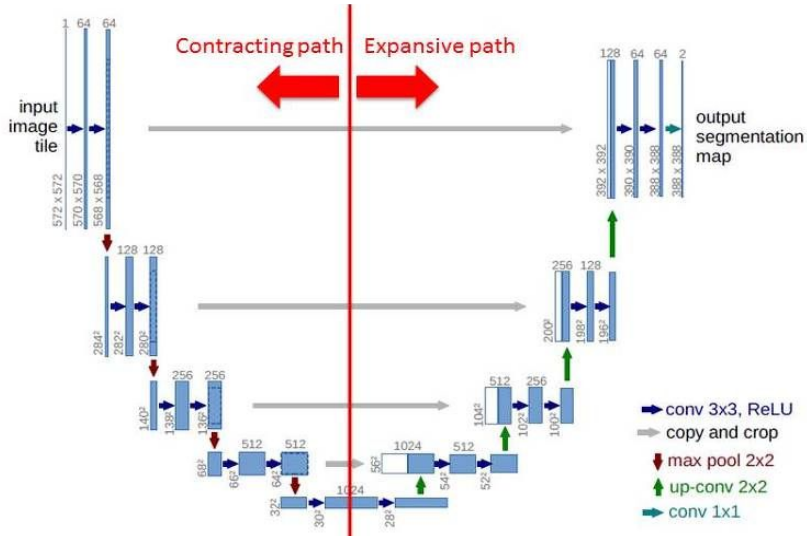


Figure 6. U-Net architecture.[2]

3. RESULTS

The accuracy of a model shows how much of the testing data the model labeled correctly. An epoch is a complete pass of all the training data through the model. It can be said that as the number of epochs increase, the model becomes more "experienced." Accuracy and loss go hand in hand in determining how "good" a model is. By optimizing a loss function, the ML model can then fine-tune the parameters at the various layers so that the model can improve. A multinomial logistic regression takes the output of all the layers in the network so far and determines which category the input is most likely to belong to, and then during training, the

regression produces a loss after comparing the computed results to the expected results. The loss shows how well the model is doing and the model tries to minimize it as it iterates through the data. The lower the loss, the better.

When graphed against epochs, the accuracy remained around 90% throughout training, while the loss experienced less fluctuation as it approached 0. See Fig. 7.

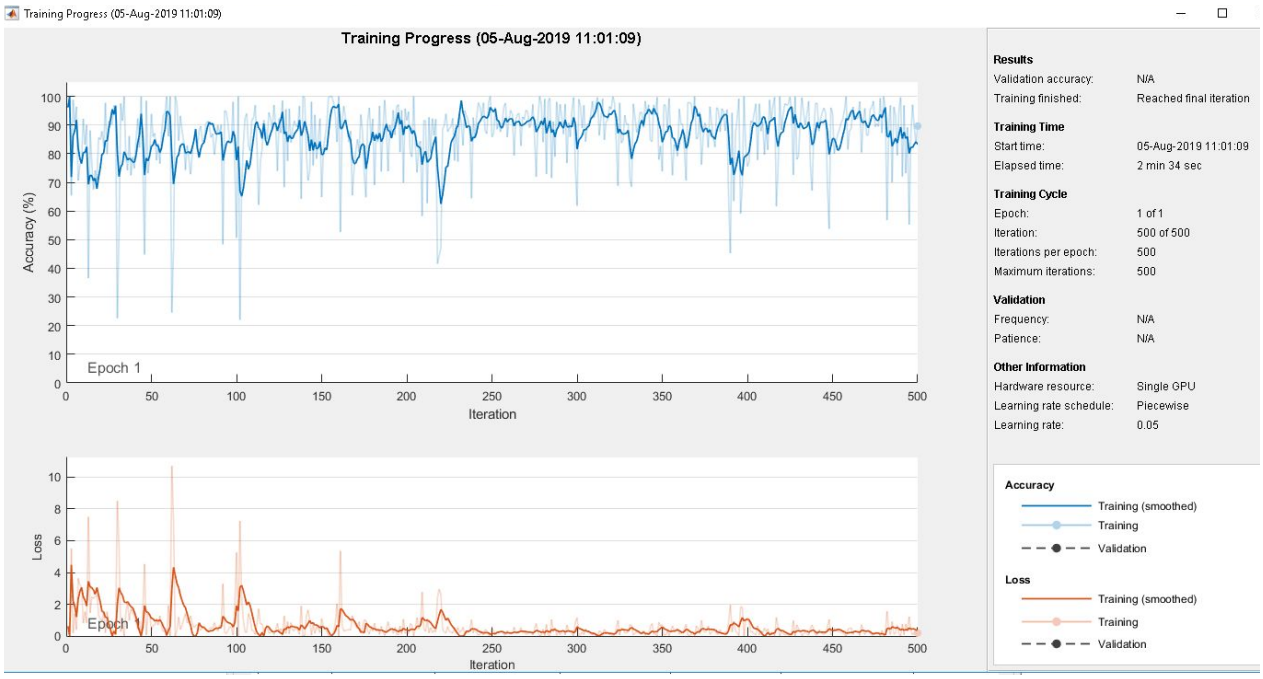


Figure 7. Accuracy and loss graphed against iteration for U-Net training (from Experiment 1).

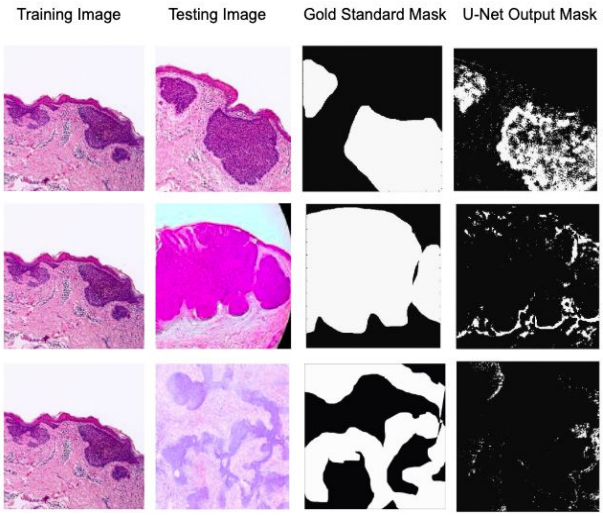


Figure 8. Given the same training image, three images were tested: one of the same origin image, one of a different microscope image, and one confocal image. (Experiment 1)

Experiment 1 consisted of using one square region image to train the U-Net. Three different test images were used: one square region image from the same tissue, one square region image from a different tissue and microscope, and one square region image from a confocal microscope. See Fig. 8. The results show that when the training image and the testing image are from the same tissue and photographed using the same microscope, the U-Net output mask is the most accurate, generally matching the gold standard masks and identifying the tumor regions with relatively high precision. However, as the tissue and microscope begin to differ, the model is not able to create as accurate of a mask, instead, it is only able to outline the tumor. When a completely different type of microscope is used, in the case of the third test, a confocal microscope, the model fails to identify any clear shape or region.

As a result of these tests, we pursued the test with the most favorable and consistent outcome: the model trained and tested on the same tissue and microscope, known in the paper as Experiment 2. The single training image was

augmented into five additional, slightly different images, using various techniques such as cropping and flipping to increase the diversity in the data. The U-Net architecture, with a data-augmentation-focused training strategy, is known to perform well with limited data.[2]

The accuracy of the same model's test labels are significantly higher, this time producing test labels (output masks) that clearly identified the tumor completely. See Fig. 9.

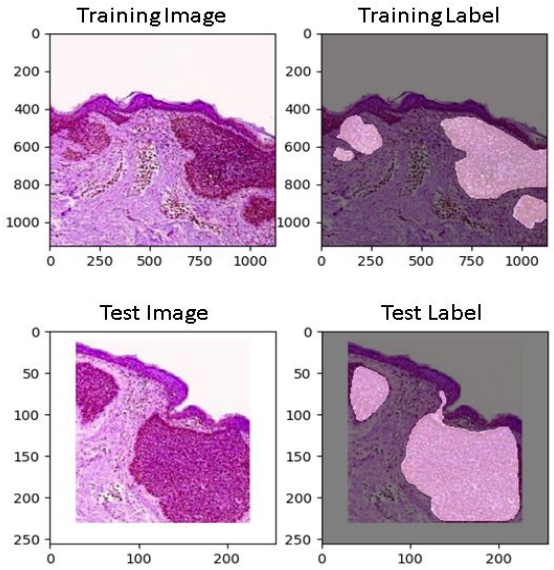
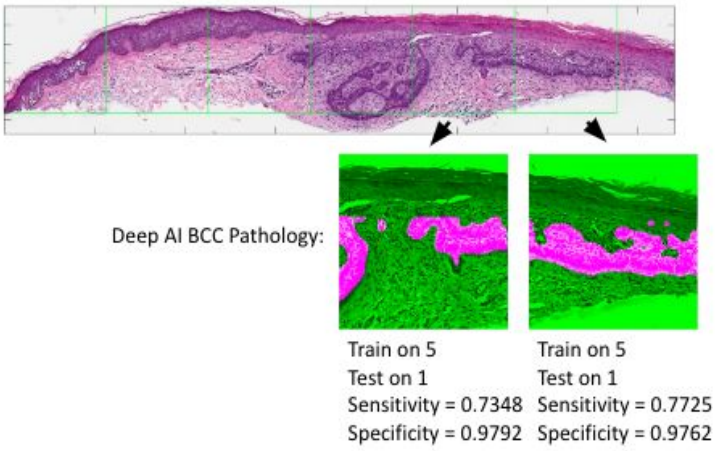


Figure 9. Post-data augmentation results (Experiment 2).

In an attempt to further the accuracy without data augmentation, we used a cropping approach.

Figure 10. Example trained on five images, tested on 1.



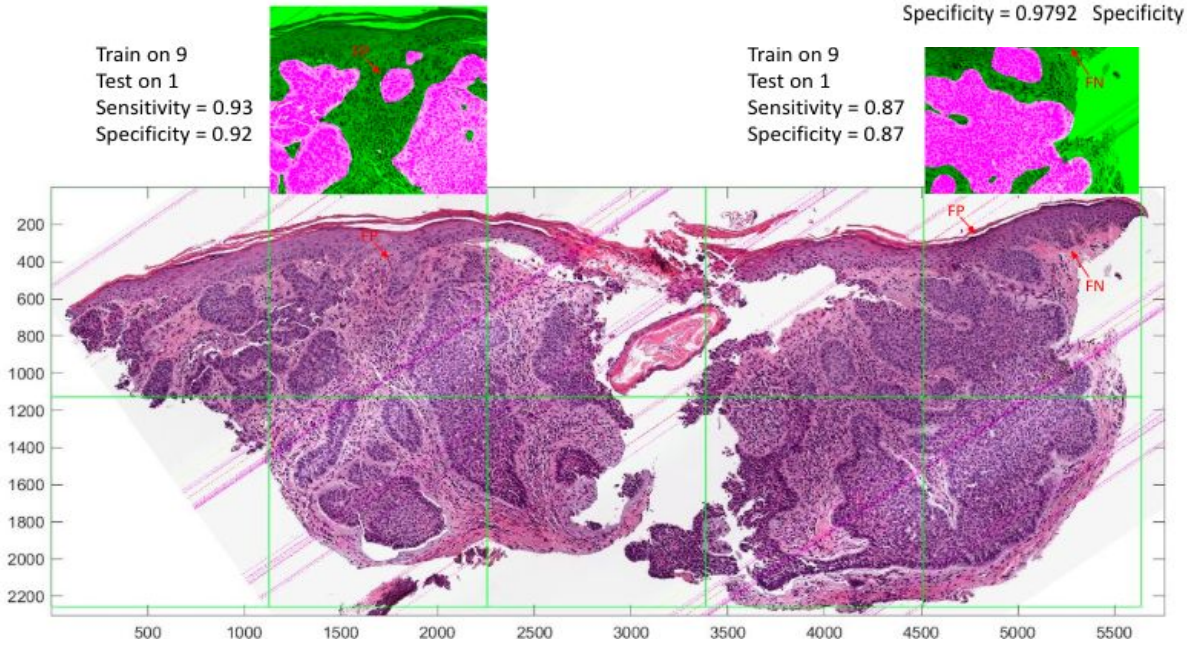


Figure 11. Example trained on 9 images, tested on 1.

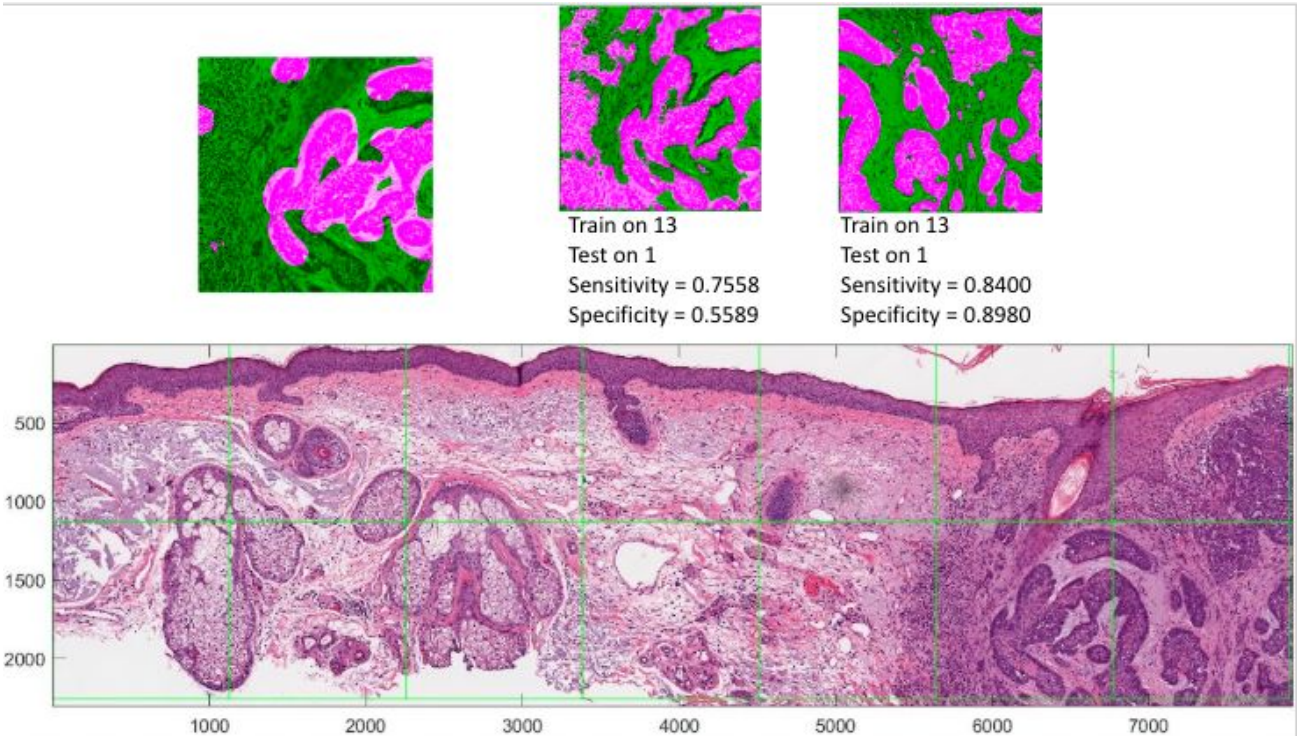


Figure 12. Example trained on 13 images, tested on 1.

CONCLUSIONS

As the most common form of skin cancer, basal cell carcinomas are a natural choice for a rapid, widespread starting point for automated skin cancer. Pathologists must examine pathology images to diagnose BCC, potentially resulting in delay, error and inconsistency. To address the need for standardized, expedited diagnosis, we created an automated diagnostic machine to identify BCC given pathology images. We collected a dataset of nodular BCC pathology images from Google and created gold standard masks using a MATLAB labeller. In MATLAB, we adapted a deep neural network image segmentation model called U-Net to train on the dataset and their corresponding masks, which can learn to highlight these nodules in pathology images by outputting a computer-generated mask.

To display the performance of the model, we executed three different experiments, each trained the U-Net on images from the dataset and compared the computer-generated mask output from testing.

In Experiment 1, to display the versatility of the model, the U-Net trained on one image from the dataset and compared the computer-generated mask output from testing on three types of images: an image from the same tissue taken with the same microscope, an image from a different tissue sample with a different microscope, and an image taken with a confocal microscope. The U-Net model outputs an increasingly inaccurate mask as testing images deviate more from the training image. However, the ability of the model to identify even outlines and parts of BCC is promising in terms of versatility as testing images begin to span across image types. In Experiment 2, as data augmentation was utilized, the accuracy of the output masks rose dramatically in the test images that were from the same tissue and microscope and produced masks that were comparable to the manually created gold standard masks. In Experiment 3, images trained on five smaller, cropped images and tested on the final cropped image yielded sensitivities between .73 and .77 and specificities around .98. As the number of cropped images increased to ten, the sensitivity increased to .87 and the specificity decreased to .87. When the number of images increased to fourteen, there was more variance in the sensitivities and specificities, some leading to significant false positive errors. When different images were trained, there occurred a false negative error.

This is the first step in creating a standardized ML approach to BCC diagnosis. The next steps involve multi-class segmentation, such as differentiation between dermis and background or other skin cancers. This is a surgeon interface that supports rapid pathological assessment and machine learning diagnostics for pathological features, expediting and standardizing the BCC diagnosis process.

REFERENCES

[1] “Basal Cell Carcinoma.” Mayo Clinic, Mayo Foundation for Medical Education and Research, 13 Mar. 2018, www.mayoclinic.org/diseases-conditions/basal-cell-carcinoma/symptoms-causes/syc-20354187.

[2] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[3] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).