

Pei-Chen Lee(Peggy)

Steven Bethard

INFO 507 Research Method

16 April 2021

Enhance spectrally encoded confocal microscopy (SECM) on breast cancer diagnosis through deep learning models created by 124 images from 23 patients

SPECIFIC AIMS

Breast cancer is the second leading cause of cancer death in women. The chance that a woman will die from breast cancer is about 1 in 39 (about 2.6%).[1] Many breast cancer patients are treated with breast-conserving surgery(BCS) due to positive margins found in margin assessments. There are a few intraoperative margin assessment methods been used that including frozen section analysis[2] and touch preparation analysis.[3] They showed the capabilities to reduce the need for additional surgeries. However, these methods are not widely used because of technical limitations, sampling bias, and the requirements of specialist's interpretations.

Optical imaging technologies hold great promises for intraoperative margin assessment applications since they can examine optical properties and or histomorphologic features of freshly excised tissues without the need for frozen sections or other tissue preparations. One of these techniques is Spectrally encoded confocal microscopy (SECM), a high-speed reflectance confocal microscopy technology that has the potential to rapidly produce images from the entire surgical margin at subcellular resolution. [4] Executing SECM for many breast cancer patients who need breast-conserving surgery is more convenient and faster than the gold-standard method.

There are kinds of approaches that have been applied to medical image analysis, and many research articles indicated machine learning methods[5] and deep learning models[6] would be applied to differentiate pathological images. To be more specific, using machine learning, deep learning, and the latest research progress of data analysis on differentiating medical images indicate that transfer learning help produced accurate enough techniques or predictive models for medical image diagnosis. Now, we have not got a model for making diagnoses through SECM images so far. For training a model for classifying histopathology images, we usually need one thousand more images. However, it is hard for us to get enough SECM images for training a deep learning model in this study because the ground truths of the training dataset can only be provided by experienced pathologists. Moreover, these pathologists are not able to tell the details from grayscale images without accepting further training. Additionally, they still need to spend more time on making results of SECM images and gold standard histologic diagnoses on the same sample. Here, we only acquire 124 diagnostic results from well-trained pathologies in a limited time since they need to make double results on the same numbers of specimens. [7]

Our overall goal is to build up deep learning models demonstrating decent accuracy, specificity, or recall when classifying histopathology images from breast cancer patients based on using 124 SECM images. To achieve that goal, I probably need to select techniques for grayscale image preprocessing and augmentation for our research. Then, I am going to test the performance of a variety of pre-

trained models by using accuracies on validating datasets firstly. Next, I planned to expand the retrain dataset via importing H&E images, and see whether the method may improve the training effectiveness. Finally, I will select practical models among all deep learning models presenting the greatest potentials of screening or diagnosing for patients who accept breast margin assessments. Therefore, the specific aims of this project including: What are the effective ways of doing image preprocessing or data augmentation of the greyscale images? Any some methods can utilize while processing these SECM images? I planed to apply some techniques of digital image processing on SECM images. Once confirmed the approaches of performing better resolutions or data augmentation, I will utilize them in manipulating SECM images.

1. What are the effective ways of doing image preprocessing or data augmentation of the greyscale images? Any methods can be applied to SECM images? I planed to apply some techniques of digital image processing on SECM images. Once confirmed methods of performing better resolutions or data augmentation. I will utilize them in manipulating SECM images.
2. Will build a pre-trained model from many color images benefit training a model's ability to differentiate SECM images? Since almost all of the pathologists learned to make diagnoses by H&E stain images, I want to develop a pre-trained model from many H&E stain images[8], which are two colors without too many artificial effects. After that, this pre-trained model would benefit the training model's ability to differentiate SECM images.
3. What are the relationships between any classification systems and indicators from performance metrics based on 124 SECM images? For clinical professionals, they care about high accuracy in differentiating benign or malignant tissues. Additionally, if the model shows high specificity on adipose, fibrous, and the other rest of types, that can be a helpful classification system for clinical application.

LITERATURE REVIEW

There are kinds of approaches that have been applied to medical image analysis. The application of machine learning and deep learning in the early diagnosis of diseases combines the latest research progress of big data analysis of medical images, especially the classification and segmentation of medical images.

Researchers utilized Convolutional Neural Networks (CNNs) extra features from images and machine-learning algorithms for differentiating histological and microscopical elements. First, they chose the parameters of denoising and histogram equalization methods to preprocess the images. Then, the machine-learning algorithm can imitate the clinician's qualitative and visual process of analyzing reflectance confocal microscopy (RCM) mosaics at the dermal-epidermal junction (DEJ) of skins. They divide the mosaics into localized areas of processing and capture the textural appearance of each area using dense Speeded Up Robust Feature (SURF). Using these features, they train a support vector machine (SVM) classifier that can distinguish between meshwork, ring, clod, aspecific, and background patterns in benign conditions and melanomas. The results show classification with 55 – 81% sensitivity and 81 – 89% specificity in distinguishing these patterns. [5]

Another paper used pre-trained CNNs on a traditional classifier which automatically differentiates between healthy tissues and cancerous samples. For this purpose, a two-phase model was for the automated classification of histopathology images. In the first phase, pre-trained CNNs such as VGG16, VGG19, Xception, and ResNet50 were designed to extract features from histopathological images. In the second phase, these extracted features would be further utilized by machine learning methods. For example, SVM and Logistic Regression (LR) applied on different magnification factors (including 40x, 100x, 200x, and 400x). In sum, they indicated the ResNet50 network achieved maximum accuracy for LR when comparing to SVM in magnification factor. [6]

These are earlier machine learning algorithms of SR, SVMs, K-Nearest Neighbours(KNNs), Decision Trees, etc. Nevertheless, traditional machine learning cannot comprehend the complexity of such healthcare-oriented problem statements owing to the complexity and importance of the subject. For

example, content-aware image restoration (CARE) networks generate results of content-aware image restoration based on deep learning extends the range of biological phenomena observable by microscopy. The authors demonstrated the first application of CARE on improving common analysis tasks of live-cell imaging and nuclei segmentation. The resulting CARE network performed well even on extremely noisy, previously unseen live-imaging data from fluorescence microscopy images. [9] In most researches, it is difficult to obtain large-scale medical annotation sets, and transfer learning is an effective method to solve the problem caused by small data. The major bottleneck is how to train a deep CNN model with a limited amount of training data. Could it be possible to use transfer learning and fine-tuning in histopathology image analysis to reduce the effort of manual data labeling? Several people mainly work on detecting fundus diseases in transfer learning.[10] To find the potential factors between the accuracy and the type of primary models, the process using the pre-trained models to complete transfer learning, such as CaffeNet, GoogleNet, VGG19, and AlexNet.[11] Another article addressed this question quantitatively by comparing the performances of transfer learning and learning from scratch for cell nuclei classification. They evaluate four different CNN architectures trained on natural images and facial images. These models including AlexNet, GenderNet, GoogLeNet, and VGG-16. A maximum of 88.03% accuracy was achieved with fine-tuned VGG-16 model from size 500×500 of whole slide images from 9 patients. [10] Different medical imaging techniques, the advancement of image acquisition devices have reduced the challenge of data collection with time. Therefore, we are in an age where there has been rapid growth in medical image acquisition as well as running challenging and interesting analysis on them. Accordingly, I will try to import the approaches mentioned above and report practical results based on the SECM images.

RESEARCH STRATEGY

B. SIGNIFICANCE

B1. Adoption of new technology to a diagnosis of breast margin assessment can decrease morbidity and mortality. Since 2007, breast cancer death rates have been steady in women younger than 50, but have continued to decrease in older women. From 2013 to 2018, the death rate decreased by 1% per year. These decreases are believed to be the result of finding breast cancer earlier through screening and increased awareness, as well as better treatments.[1] Improving new examination techniques to diagnose breast margin assessment is important because those are indicators for doctors selecting different treatments. So that, New technologies for a diagnosis of breast margin assessment can lower morbidity and mortality. Breast-conserving surgery (BCS) is now the standard of care for most women with early-stage breast cancer. As many as 20% of patients still require additional surgery to interpret the state of resection margin even though the development of approaches for assessing resection margins keeps going. Several new techniques for intraoperative marginal assessment have been discovered, but they are accompanied by patients' anxiety, inconvenience, and extra costs. Compared to conventional clinical methods, SECM brings out a safer, applicable, and an affordable technology of novel ones.[12]

B2. Neural network models can observe details in images to aid pathology diagnosis via reducing time on getting pathology reports and further training pathology professionals. In routine clinical-pathological diagnoses, doctors usually wait 24-36 hours to get a report for an H&E stained tissue. Also, pathologists need to accept more training for distinguishing greyscale images, and especially it is hard for humans to see the differences from non-colored smears. [13] Therefore, we suggested a better way to utilize neural network models supporting detailed observations among greyscale images and saving turnaround time for generating reports. So, a margin assessment of breast cancer could complete within one day

B3. With computer learning algorithms to classify the images from a new medical imaging technique, patients would get high-quality medical reports. Applied the latest models for classifying medical images is crucial for reducing the review rate of pathology diagnosis. In fact, pathology laboratories spend most of the effort on processing benign tissue because malignant-

like tissue usually around <20%. In some cases, intelligent imaging and deep learning provide better answers in medical image classifications by reducing the chances of misdiagnosis, which could cause by artifacts or diagnosis biases. Our work focus on applied the most helpful model to classify the images from a new medical imaging technique that can remarkably reduce workload by tissue processing and frequencies of smear reviews. Furthermore, clinical staff would be able to have sufficient time and effort to focus on fewer samples and enhance the quality of reports.

C. INNOVATION

Image preprocessing, data augmentation, machine learning, and neural network model methods are now widely used in classifying medical images from kinds of optical examiners. In previous studies with small training datasets, the widely used methods are self-created small models or the technique of whole slide images. By contrast, we firstly present that many pre-trained models with many parameters can be applied to classify histopathology images via less than 124 training images. For the data augmentation, we newly import H&E images, which is the gold standard method while retraining the models for improving the training processing of building models. This is the first-time attempt for building a neural network model to classify SECM images to enhance the workflow of diagnosing breast margin pathology tissues.

D. RESEARCH DESIGN

D1. Study Design

This application is a secondary data analysis of clinical images from professor Dongkyun's research of spectrally encoded confocal microscopy for diagnosing breast cancer in excision and margin specimens. Specifically, to construct the predictive models for these SECM images, some pre-trained models will be tested on feature extractions and classifying abilities that used transfer learning, coupled with the parameter adjustments for retraining and fine-tuning models. To overcome the limitation of sample size, a dataset of H&E images will include while retraining models. Based on the metric of the results from kinds of training data, testing data, and model settings, I may provide the most helpful model for classifying SECM images for clinical usage based on high accuracies, recall, or precision.

D2. Sample

For the proposed application, the full database of 124 SECM images representing benign or malignant breast tissues from 23 patients. For each SECM image either having one of the six normal/benign categories (adipose, fibrous, ducts/glands, inflammation, proliferative, and unspecified normal/benign) or one of the seven malignant categories (invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC), unspecified invasive carcinoma (IC), ductal carcinoma in situ (DCIS), lobular carcinoma in situ, unspecified carcinoma in situ, and unspecified malignant). The dataset also including a match H&E stained image for each SECM images. (Table 2) [7] The Institute of Electrical and Electronics Engineers (IEEE) provides a free resource 7909 H&E images those are generated from breast tissue biopsy slides, stained with H&E. The samples are collected by surgical (open) biopsy, prepared for histological study, and labeled by pathologists of the Prognostics and Diagnostics (P&D) Lab. The dataset currently contains four histological distinct types of benign breast tumors: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA); and four malignant tumors (breast cancer): ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC).[8]

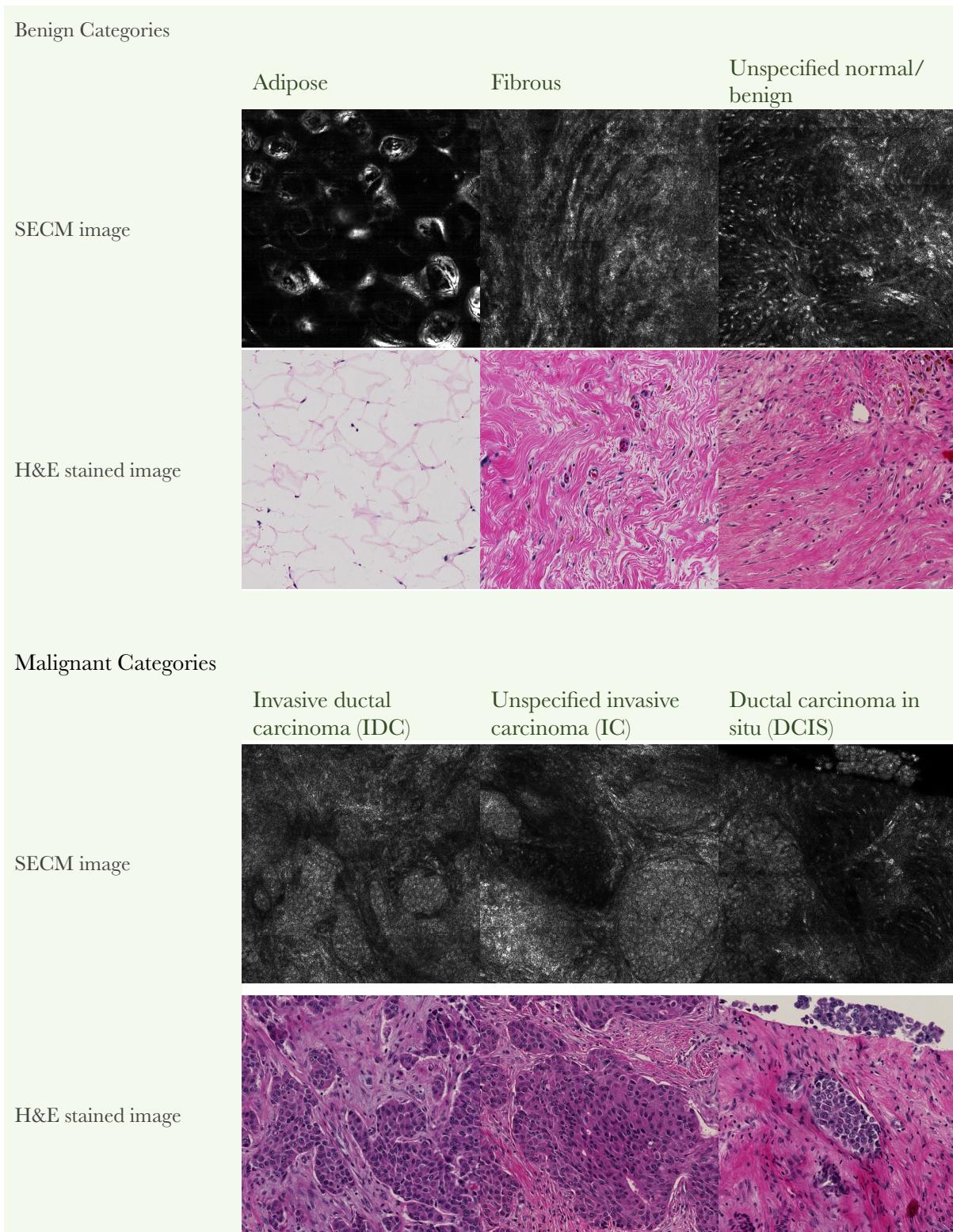


Table 1. Part of the images from the dataset.

D3. Research Procedures

The design of the workflow for building a predictive model, shown in Figure 1. We initially separated the images by different patients that could avoid models recognizing a patient's patterns instead of the certain patterns from each category, so I will split the 124 images into train and test sets, ensuring that

each patient is only in one set. Second, we applied some image processing techniques to reduce the noises as much as possible. Next, evaluating the accuracy provided by pre-trained models, including VGG16, VGG19, ResNet50, and ResNet152. Besides directly testing abilities of transfer learning on these models, I also set some models to do retraining firstly by using H&E images. After that, I will explore the levels of fine-tuning for classifying models. Here, our initial focus will be differentiating images from benign/malignant or reviewed by human/not reviewed needs. Then, I expected to evaluate the models by rates of accuracies or specificity plus recall. Here the proposed models are able to learn differences between categories of images and then classify SECM images from new patients with a trustworthy categorization. Furthermore, to demonstrate the best performance of the best of the most representing categorizing system, I planned to try leave-one-out cross-validation to calculate the average precision and average recall based on emphasizing the other tissues.

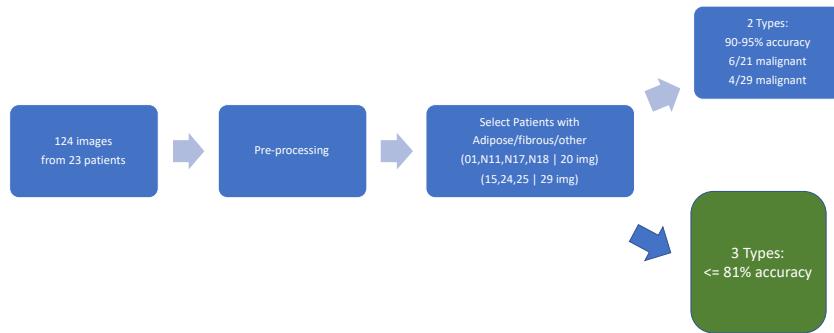


Figure 1. The first workflow for building predictive models

E. PRELIMINARY WORK

E1. Image Processing

The noise and horizontal lines existing in images may interfere with deep learning models recognizing images. After attempting to remove noises in pre-processing images by different adjustments, I have got the best quality of images while applying Gaussian Blur, color calibration, and other image preprocessing techniques. (Figure 2) I chose to utilize Gaussian filtering since it could remove most of the noise without affecting edges so much. However, I got similar or lower accuracies in results by using pre-processing techniques, so the final results in this study are based on images without any particular processing techniques.

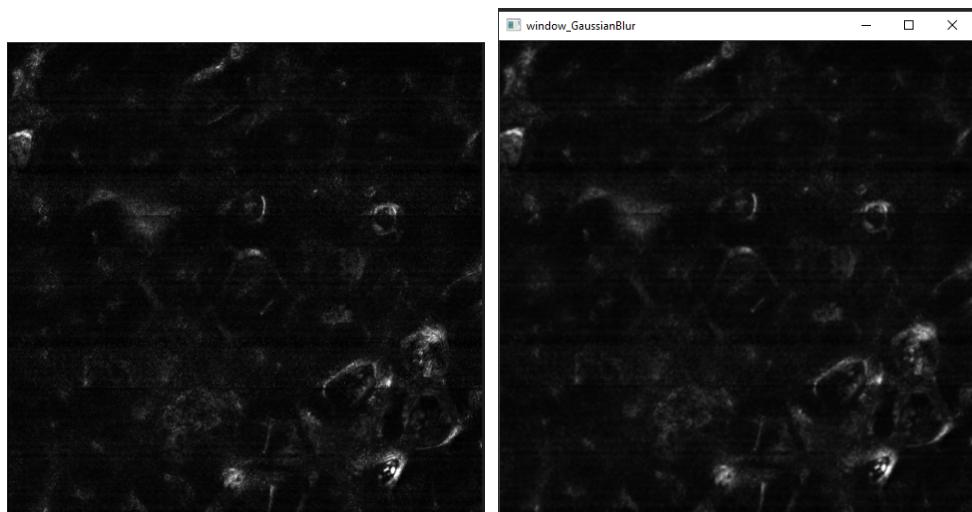


Figure 2. An example of a image without pre-processing(left) and with pre-processing(right)

E2. Training Data and Validation Data

The images from the same patient can not both be used in training data and validation data because test data should be entirely hidden from training processing. Specifically, the models ideally recognized the different morphology of each category instead of the characteristics of individual patients. Therefore, I selected the validation data from dataset at the beginning, which should include three categories (adipose/fibrous/other) and including around 8 images in each. Under the resource limitations or criteria restrictions, I can only get 2 groups of images as validation datasets, and each of them including 21 and 29 images. In this experiment, the data of training argumentation data implemented by image rotations and mirroring since the way helps models be more resilient to camera roll and increases the numbers of training images. Due to the total number of images, I did not preserve the test dataset in this experiment, but only take the models with the highest accuracy to do further analysis. In the leave-one-out cross-validation, the validation data will loop through from the first patient to the last one and the training data is composed of the whole dataset without images from the patient.

E3. Result

According to the previous research articles, the VGG16 and VGG19 have good performances for classifying histopathological images. Based on previous research results and features of the latest pre-trained models, I decided to test the ability of transfer learning from VGG16, VGG19, ResNet50V2, and ResNet152V2 through accuracy on validation data. The result demonstrates high accuracy when categorizing adipose and fibrous into benign/ malignant tissue into malignant tissues, which performed at least 80% accuracy, and most of them are more than 90%. On the other side, the models showed 60-80% of classifying adipose, fibrous, and other tissues. While retraining pre-trained models via H&E images, I efficiently got similar results that approaching 90% accuracy for two categories, and around 60% accuracy for three categories. So far, the best model for doing these two classifying tasks is RestNet152V2, so I decided to pass this kind of model to do leave-one-out cross-validation. (Table 2)

	2 CLASSIFICATIONS	3 CLASSIFICATIONS
VGG16	<90%	<60%
VGG19	<90%	<61%
ResNet50V2	81%	71%
ResNet152V2	95%	81%
VGG16 + retrain by H&E	<90%	<61%
VGG19 + retrain by H&E	<90%	<61%

Table 2. Accuracies of models in 2 or 3 categories.

For getting these values from the best epoch so far, I tested the ResNet152V2 from 5 to 40 epochs with 5 epochs as intervals on each testing data(total patients is 23 = 23 test data). Then, I have got the best performance of fifth epochs are the average precision is 0.85, the average recall is 0.998, and the average accuracy is 0.874. To elaborate, the model showed the 85 corrected predictions on 100 predictions, and there are 85 other tissues selected from overall 86 images of the other tissue type. The

average accuracy(87.4%) calculated using each patient and their image numbers, which is very close to the accuracy of the two classifications, which is around 91%. All the programming information can be access through this Github link: https://github.com/pe791006ggy/Capstone_BreastCancer.git.

F. SUMMARY

F1. Findings and Discussion

Image preprocessing, data augmentation, and machine learning model methods are now widely used in classifying medical images from kinds of optical examiners. In previous studies with small training datasets, the widely used methods are self-created small models or the technique of whole slide images. By contrast, we first presented that many pre-trained models with many parameters can be applied to classify histopathology images via 124 training images from 23 patients. We also found all techniques of image preprocessing should not be applied in the experiment, such as Gaussian Blur and color calibration. That is, the images that become clear for human eyes are not always good for deep learning training because the techniques cause information loss and they did not provide benefits for predicted models.

Due to the size of the dataset, we newly import H&E images, which is the gold standard method while retraining the models for improving the training processing of building models. Our results indicated that the models retrained by H&E images advance are better than models with only random initial parameters because the training processes of the former one are more predictable within fewer iterations. Therefore, I suggested when training data for building models to recognize histopathology images is less than usual, people could consider expanding training datasets via H&E images.

Most importantly, the results have revealed the deep learning models presenting abilities to classify SECM images. Moreover, the accuracies provided by VGG16, VGG19, and ResNet152V2 on differentiating adipose, fibrous, and other tissues could be even better than judgments from the experienced pathologist, which is around 80%. However, the accuracies of categorizing SECM images into adipose tissue, fibrous tissues, or other tissues are lower than the two categories system. That seems not intuitive because the morphologies of adipocytes and fibrous tissues are very different from other tissues for medical professionals. Therefore, the rationality for models to distinguish tissues is unclear because models can categorize adipose and fibrous into benign, but they cannot identify the characters of adipose and fibrous. The models may rely on some criteria to do classifying, but those criteria are not as same as humans look at while classifying medical symptoms, or these predicted models always have higher chances to guess the correct labels under fewer categories.

Next, the results of leave-one-out cross-validation for evaluating ResNet152V2 convinced us with its capability of helping auto-labeling and auto-selecting adipose/fibrous tissues through the model had 0.85 of the average precision and 0.998 of the average recall for distinguishing images without human-reviewed needed. Although the data resource cannot prove the model is general enough, the model tends not to put images that need human reviewing into the other groups, hence, the model still appropriate in medical applications based on it can reduce the workload for pathologists. Considering the size of datasets and the single data resource, data bias, single data resource, particular tissues, image noises, hardware, I believe the models may present more classifying abilities if the size of training data is bigger and more balancing.

F2. Future Directions and Potential Impact

For assisting marginal assessments of breast cancer based on image recognition models, the best model of this study is ready for prospective testing in a larger cohort of participants and it can be used to reduce tissues that need human-reviewed. I believe the models could present higher accuracies on any tissue via two methods. The first approach is to increase the size of SECM data and let the data be more balancing. The second method is importing H&E stained images of all categories into the retraining process. We anticipate this temptation will optimize the processing while introducing new

examinators for histopathology, enabling cancer patients to get better treatment within a shorter time. In the long-term, the rationalities behind predictive models may lead to more precise diagnosis standards, which will reduce costs, provide more benefits, and potentially reduce the burden of cancer patients.

Bibliography

1. *Cancer Facts and Figures 2021*. American Cancer Society, 2021(2021).
2. Boughey, J.C., et al., Impact of analysis of frozen-section margin on reoperation rates in women undergoing lumpectomy for breast cancer: evaluation of the National Surgical Quality Improvement Program data. *Surgery*, 2014. **156**(1): p. 190-7.
3. Esbona, K., Z. Li, and L.G. Wilke, Intraoperative imprint cytology and frozen section pathology for margin assessment in breast conservation surgery: a systematic review. *Ann Surg Oncol*, 2012. **19**(10): p. 3236-45.
4. Schlachter SC, K.D., Gora MJ, et al, Spectrally encoded confocal microscopy of esophageal tissues at 100 kHz line rate. *Biomedical optics express*, 2013. **4**,**9**: p. 1636-1645.
5. Kose, K., et al., A machine learning method for identifying morphological patterns in reflectance confocal microscopy mosaics of melanocytic skin lesions in-vivo. *SPIE BiOS*. Vol. 9689. 2016: SPIE.
6. Karan Gupta, N.C., Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN. *Procedia Computer Science*, 2020. **167**: p. 878-889.
7. Brachtel, E.F., et al., Spectrally encoded confocal microscopy for diagnosing breast cancer in excision and margin specimens. *Lab Invest*, 2016. **96**(4): p. 459-67.
8. Spanhol, F.A., et al., A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans Biomed Eng*, 2016. **63**(7): p. 1455-62.
9. Weigert, M., et al., Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods*, 2018. **15**(12): p. 1090-1097.
10. Neslihan Bayramoglu, J.H., Transfer Learning for Cell Nuclei Classification in Histopathology Images. *ECCV Workshops*, 2016.
11. Cai, L., J. Gao, and D. Zhao, A review of the application of deep learning in medical image classification and segmentation. *Ann Transl Med*, 2020. **8**(11): p. 713.
12. Dumitru, D., M. Douek, and J.R. Benson, Novel techniques for intraoperative assessment of margin involvement. *Ecancermedicalscience*, 2018. **12**: p. 795.
13. Rubin, L.R., et al., Using color and grayscale images to teach histology to color-deficient medical students. *Anat Sci Educ*, 2009. **2**(2): p. 84-8.