Detecting Breast Cancer Metastases in Lymph Nodes on Gigapixel Pathology Images

Fariba Dambandkhameneh
IAMLAB
Ryerson University, Toronto, Canada.
Department of Chemical and Biomedical Engineering
fdambandkhameneh@ryerson.ca

***Abstract*— The uncontrollable growth of cells in the breast tissue causes breast cancer, which is a common type of cancer in women. Lymph nodes in the axilla are small glands that filter the fluid circulating through the lymphatic system. Lymph node metastasis involvement is one of the most important prognostic factors in breast cancer. The metastasis evaluation is performed in WSIs. The grading of whole slide image is time-consuming and therefore expensive. The proposed method deals with these problem in three steps. First, a WSI is split into smaller patches. Next, different networks are applied to automatically detect tumors on patches as small as 256×256 pixels in gigapixel microscopy images sized 100,000×100,000 pixels. After choosing the best architecture, the results of each slides are merged to be classified into normal or tumor slides. The highest accuracy of the patch-level classification using VGG-19 is 99%; however, the accuracy of the slide-level classification is 76%.**

I. INTRODUCTION

Breast cancer is a common type of cancer in women which is caused with the uncontrollable growth of cells in the breast tissue Breast cancer is a common cancer type between women in the United States of America [1]. About 268,600 new cases of invasive breast cancer are diagnosed in women in the U.S. as well as about 62,000 new cases of non-invasive (in situ) breast cancer in 2019. More than half of the cases are discovered at a localized stage with a 5-year survival rate of 99%.
Lymph nodes in the axilla are small glands that filter the fluid circulating through the lymphatic system. Lymph node metastasis involvement is one of the most important prognostic factors in breast cancer [2]. Diagnosis becomes difficult if cancer has spread to the lymph nodes. The procedure to diagnose the cancer is tedious and time-consuming. Small metastases are very hard to detect, and sometimes they are missed.

Metastasis detection helps clinicians to choose a proper medication option and to obtain an indication of prognosis. In this project, we will focus on the metastasis detection of lymph node metastases, which is an essential part of a breast cancer diagnosis. In traditional pathology, tissue samples are examined through the microscope by a medical doctor who is specialized in detecting and characterizing diseases. Glass slides are prepared to process further the tissue samples. Tissue processing includes formalin fixation, paraffin embedding, cutting a couple of micrometers thin slices and placing on a glass slide, and staining. H&E staining is applied to highlight the cell nuclei that induces sharp blue and pink contrasts across various cellular structures. Whole-slide scanners are employed to digitize glass slides at high resolution.

Whole Slide Images (WSIs) are a multi-resolution pyramid image that contains multiple down-sampled versions of the original image. A common WSI is about 200000×100000 pixels on the highest resolution level with a 3-byte RGB pixel format. This indicates 55.88GB of uncompressed pixel data from a single level.
Automatically analysing histopathological WSIs in a clinically relevant setting is possible, but can not yet be performed without some form of guidance by a clinical specialist. Digitized pathology methods could effectively assist clinicians by pre-screening the WSIs. The pre-screened images could point the attention of the pathologist to the appropriate areas and ease the pathologic N stage (pN-stage) by outlining metastases. This project will focus on evaluating new methods for detection, segmentation, and classification of breast tumoral areas in WSIs for histological lymph node sections. Lymph node classification aims to distinguish between normal tissue and tumoral tissues and to perform a prognostic in breast cancer. Hematoxylin and Eosin(H&E) is the principal stain of tissue specimens for routine histopathological diagnostics. Through mi-

croscopic assessment, the pathologist selects the slides for tumor presence. The pathologist measures the size of the founded tumors in order to determine the pN-stage of the tumor.

A. Literature review

There are many methods developed for the digital pathology image analysis, from rule-based to applications of machine learning [3]. Camelyon2016 challenge is organized for improving automated breast cancer spreading in WSIs of the lymph nodes [4].
Some of the algorithms performed significantly better than pathologists under time pressure. Since Camelyon16 can not be used in clinical practice, Camelyon17 [5] approached to comprehend the idea and to be used in clinical applications. In Camelyon16, they have been focused on the classification of single WSIs, whereas Camelyon17 is focused on patient-level pN-stage prediction, including multiple WSIs per patient. In Camelyon16, WSIs are provided from two different centers; however, the Camelyon17 dataset includes WSIs from five different centers, which allows for a more realistic description of preparation and staining diversity across laboratories and scanners.
For pre-processing, all teams applied simple filtering and threshold to identify tissue regions in WSIs. Different groups applied thresholding in various color spaces such as RGB (red-green-blue), HSV (hue-saturation-value), or HSI (hue, saturation-intensity). Some participants applied the median filter, connected component analysis, and size filtering to remove small regions.
All participants of Camelyon17 extracted small image patches of metastases and normal areas from the WSIs to train their CNNs, although the patch size and pixel resolution differed considerably. The teams of Camelyon17 used CNNs. Variant common network architectures such as ResNet [6], GoogLeNet/Inception [7], VGG-Net [8], U-Net [9], DenseNet [10] are used for classification or segmentation of the patches.

Besides, almost all teams conducted extensive data augmentation for increasing the variation in the training set. Some of them applied rotations with angles, and some used other strategies included random cropping of patches and applying transformations (e.g., scaling). Also, to make their algorithm robust to color variation generated by differences between scanners or staining protocols, most participants used patch color augmentation in the HSV, RGB, or H&E color spaces by combining noise to the color channels or additional brightness, contrast, and gamma adjustments. One of the teams used a stain normalization algorithm [11] to obtain a uniform color distribution. To increase the robustness of the classifier, strong data augmentation is applied in [12] and deep convolutional features are extracted at different scales with publicly available CNNs pre-trained on ImageNet. In [13], AlexNet [14] and VGG16[15] is applied to classify the patches. In [16], they proposed the PFA-ScanNet with synergistic learning for metastasis detection and pN-stage classification to improve the accuracy close to clinical usage. The submitted algorithms were able to detect the presence of metastases and measure the size of tumors to determine the pN-stage that is used in clinical practice.

TABLE I
DATASET FROM CAMELYON2016 AND CAMELYON 2017

Institution	Train Tumor	Train Normal	Test
Radboud UMC	85 Tumor	65 Normal	10 Test
UMC Utrecht	65 Tumor	35 Normal	10 Test
CWZ	-	-	20 Test
RST	-	-	20 Test
UMCU	-	-	20 Test
LPON	-	-	20 Test
RUMC	-	-	20 Test

II. MOTIVATIONS

The motivation of this project is to compare different algorithms to detect, segment, and classify breast cancer metastases in WSIs of lymph node sections in data from different centers, with different stain protocols and scanners. This project will focus on the detection of lymph node metastases on a patch-level and after getting high accuracy, the patches will be merged into the WSIs results and the slide level result will be shown. This procedure requires linking the detection and classification results of metastases tiles into a one WSI image.

A. Dataset

The Camelyon16 dataset consists of 400 WSIs, which are split into 270 for training and 130 for testing. Both of them include slides from two institutions (Radbound UMC and UMC Utrecht) with particular details presented in Table 1.
The ground truth for the training slides consists of a pathologist's representation of areas of metastatic cancer on WSIs of lymph nodes. The data is provided in two forms: XML files, including points of the annotated contours of the locations of cancer metastases and

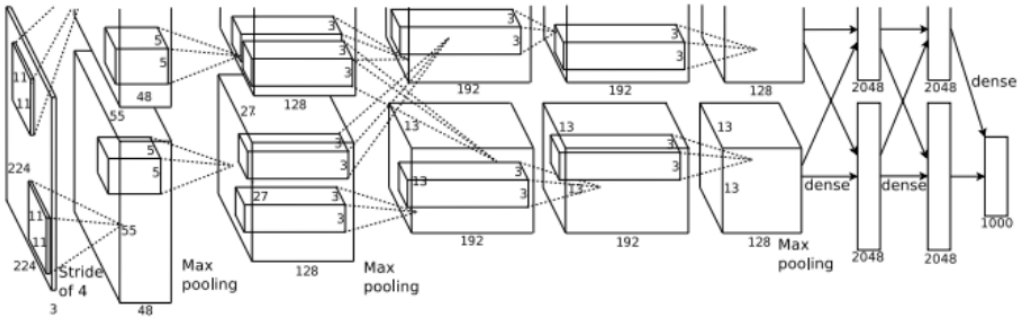

Fig. 1. Alexnet architecture (Image is inspired from [17]) .

WSI binary masks showing the location of the cancer metastasis. Additionally, 10 WSI's from 5 centers in the Camelyon17 training set are annotated by carefully outlining each lesion in the WSIs with polygons. As I don't have the result of test sets, the method is tested on the 50 annotated tumor and 50 normal slides of the training set from Camelyon2017 dataset. Additionaly, 20 slides (10 tumor, 10 normal) are used from Camelyon16 training set for testing which are not used in the training.

B. Method

Data pipeline is broke down into two main tasks: The first one, and most importantly, will determine whether patches contains tumor or not. A convolutional neural network for is used for detection tumor patches. The second task takes those results and decides whether the WSI contains metastases or not. Different architectures are tested for determining the best model.

C. Different Architectures

CNN, is a particular kind of multi-layer neural network that is designed to recognize visual patterns from pixel images with minimal pre-processing. The Transfer learning could be used to train the networks to enhance the performance of the network. Most of the networks used the networks trained on ImageNet. ImageNet is a comprehensive visual database produced for aid in visual object recognition software research. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) tried to compete to correctly classify and detect objects and scenes. The idea of transfer learning is instead of training data from scratch, we can take a network that is trained on a different domain and adapt it for our task. Different architectures have been used to train the model. The details of the architectures that are used in this project are explained here.

1) AlexNet: In 2012, AlexNet [17] has defeated all the previous models and overcame the challenge by decreasing the top-5 error from 26% to 15.3%. The second-place top-5 error rate, which was not a CNN variation, was around 26.2%. It includes five convolutional layers and three fully connected layers. ReLU is applied after every convolutional and fully connected layer. Dropout is employed before the first and the second fully connected layer. The image size of the architecture should be 227×227 instead of 224×224, as it is pointed out by Andrei Karpathy in his famous CS231n Course. This network is similar to architecture as LeNet [19], But it was more in-depth as it had more filters per layer and with accumulated convolutional layers. It consisted of 11×11, 5×5,3×3, convolutions, max pooling, dropout, data augmentation, ReLU activations, SGD with momentum. ReLU activations are attached after every convolutional and fully-connected layer. AlexNet was designed by the SuperVision group, consisting of Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever. AlexNet has 62.3 million parameters. We can also see convolution layers, which accounts for 6% of all the settings, consumes 95% of the computation. The architecture is represented in Figure 1. The last fully-connected layer is fed to a 1000-way softmax, which produces a distribution over the 1000 class labels. The kernels of the second, fourth, and fifth convolutional layers are connected only to the kernel maps in the previous layer. The kernels of the third convolutional layer are connected to all kernel maps in the second layer. The neurons in the fully-connected layers are connected to all neurons in the previous layer.

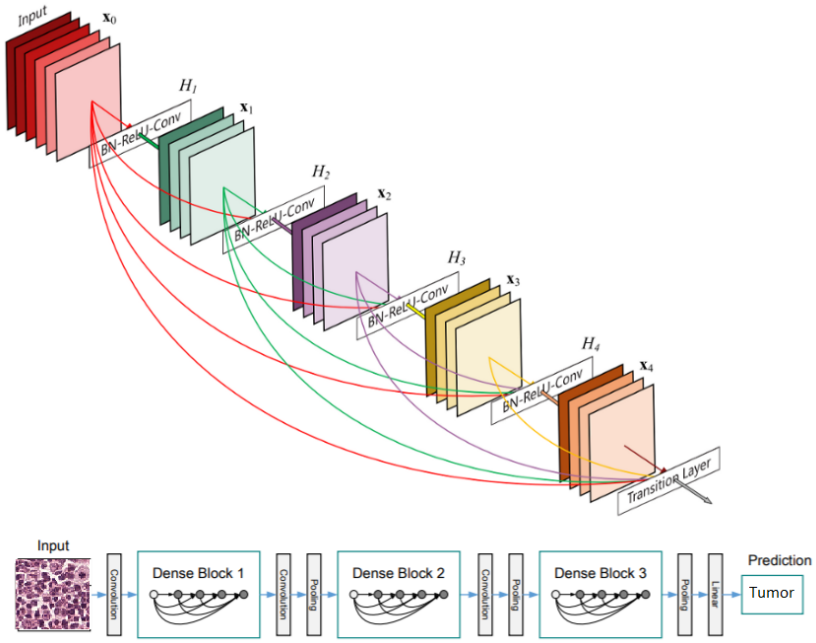

Fig. 2. Original DenseNet architecture from [18]. The top figure depcits an example dense block. The bottom figure depicts a full DenseNet architecture. [18]

2) DenseNet: DenseNet [18] reached state of the art results on classification datasets (CIFAR, SVHN, ImageNet) using fewer parameters. DenseNet uses residuals that can be deeper than other networks, and it is easy to optimize. DenseNet is formed of dense blocks. In these blocks, the layers are densely connected together. Each layer takes the input of all previous layers output feature maps. This excessive use of residual forms deep supervision because each layer takes more supervision from the loss function. A single layer has Batch Normalization, ReLU activation, and 3×3 convolution. The pre-activation mode was more effective than the typical post-activation mode. A zero padding is recommended before the convolution in order to have a fixed size. DenseNet concatenates all the feature maps instead of summing the residual like in ResNet. Concatenating feature maps of different sizes would be impracticable. Therefore, the feature maps of each layer have the same size in every dense block. However, down-sampling is essential in CNN. Transition layers between two dense blocks assure this role. A transition layer consists of Batch Normalization, 1×1 convolution, and average pooling. As 3×3 convolution can be upgraded with a bottleneck, it has few parameters. A layer of a dense block with a bottleneck consists of Batch Normalization, ReLU activation, 1×1 Convolution bottleneck producing, Batch Normalization, ReLU activation, and 3×3 Convolution. This helps the network have hundred, if not thousand, layers. DenseNet architecture is shown in Figure 2.

3) SqueezeNet: As you can see in Figure 3, "squeeze" layers are convolution layers which are made up of only 1×1 filters, and the expand layers are convolution layers with a mix of 1×1 and 3×3 filters. By decreasing the amount of filters in the "squeeze" layer filling into the "expand" layer, they will reduce the number of connections entering these 3×3 filters, therefore decreases the number of parameters in a CNN while attempts to preserve accuracy. Also, tries maximizing accuracy on a limited budget of parameters. The authors of [15] name this specific architecture the "fire module," and it works as the basic building block for the SqueezeNet architecture. Using SqueezeNet, reduces the size of the model 50 times compared to AlexNet, while reaching or surpassing the top-1 and top-5 accuracy of AlexNet.

4) VGGNet: VGGNet [8] is designed by VGG (Visual Geometry Group) from the University of Oxford, Though VGGNet is the first runner-up of the ILSVRC 2014 in the classification task, which has significantly improved AlexNet and GoogLeNet. VGG-11 (Conv1) reaches 9.4% error rate, which indicates the extra three 1×1 Conv layers improve the classification accuracy. 1×1 Conv increases the non-linearlity of the decision

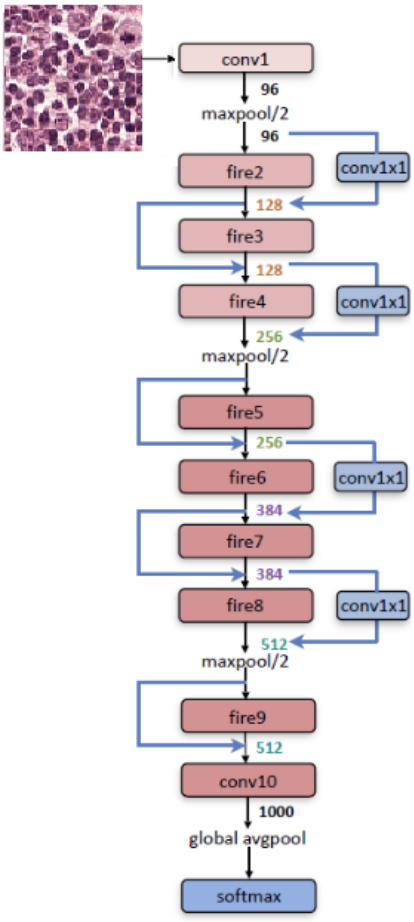

Fig. 3. SqueezeNet with complex bypass connections between the remaining fire modules. Image is taken from [14].

function. It also performs the projection mapping in the same high dimensionality without changing the dimensions of input and output. This technique is essential in the GoogLeNet and ResNet. VGG-16 obtains an 8.8% error rate, which indicates the deep learning network is still improving by adding several layers. VGG-19 achieves 9.0% error rate, which means adding the too much layers does not improve performance of deep learning networks. Thus, authors stopped adding layers. VGG-19 architecture is shown in Figure 4. The input size is fixed to 224×224 RGB image. The image is carried through a stack of convolutional layers, where the small receptive field of 3×3 were used. In one of the configurations, they additionally utilize 1×1 convolution filters, which can be a linear transformation of the input channels. The convolution stride is 1 pixel; the spatial padding of convolution Layer input is such that the spatial resolution is preserved after convolution, i.e., the padding is 1-pixel for 3×3 convolution layers. Spatial pooling is offered out by five max-pooling layers that accompany some of the convolution layers. Max-pooling is conducted over a 2×2 pixel window, with stride 2.

5) ResNet: In ILSVRC 2015, the so-called Residual Neural Network (ResNet) by Kaiming He et al. [6], introduced a new architecture with "skip connections" and batch normalization. Skip connections are also identified as gated units or gated recurrent units and have a powerful similarity to recent successful elements applied in Recurrent Neural Network (RNN)s. There are some techniques that are prepared to train a neural network with 152 layers and still have lower complexity than VGGNet. It reaches a top-5 error rate of 3.57% that beats human-level performance on ImageNet dataset. AlexNet has parallel two CNN lines trained on two GPUs with cross-connections, GoogleNet has inception modules, ResNet has residual connections. ResNet solves the vanishing gradient problem as the network is too deep, and the gradients of the loss function are measured easily. The results in the weights are never updating their values, and therefore, no learning is being implemented. The gradients can flow directly through the skip connections backwards from later layers to initial filters. Small changes are applied on ResNet50, in previous versions, the shortcut connections skipped two layers, but now they used three layers, and also, there was 1×1 convolution layers. ResNet architecture is shown in Figure 5.

III. Method

A. Preprocessing

A WSI is usually huge, while the tissue regions are only parts of it. The algorithm should detect tissue regions that are likely to contain cancer metastasis to reduce computation time and resources. First, black space is removed with thresholding in the grayscale image. Then, white background is removed with the Otsu algorithm in H and S channels in HSV space. Otsu can give us the threshold which used to binarize image according to the pixel statistics between foreground and background. Finally, the median filter is applied to remove very small regions. This process can eliminate about 80% of slides. We can see an example of preprocessing as shown in Figure 6.

B. Patch Extraction

The large size of WSIs (typically 100,000 by 100,000 RGB pixels) causes memory error in processing. To memory management, error handling and also to gain fine-grained features, tissue regions

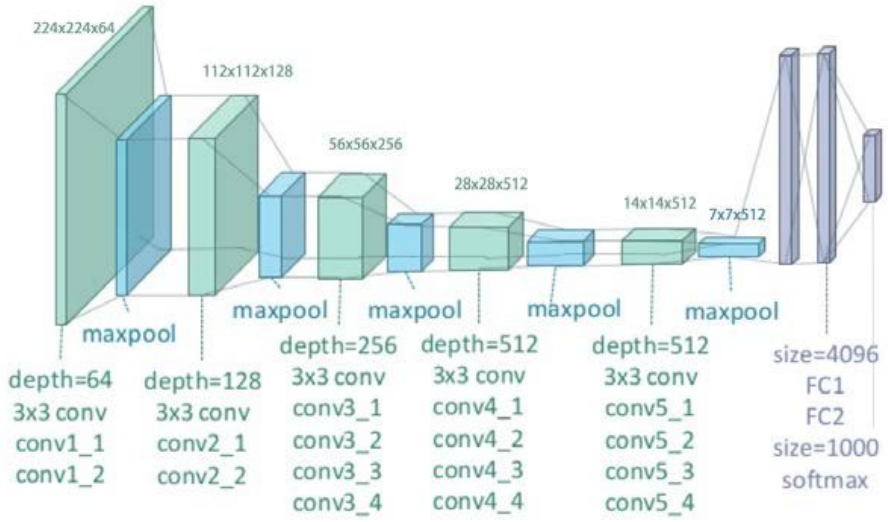

Fig. 4. VGG-19 architecture. The image is taken from [20].

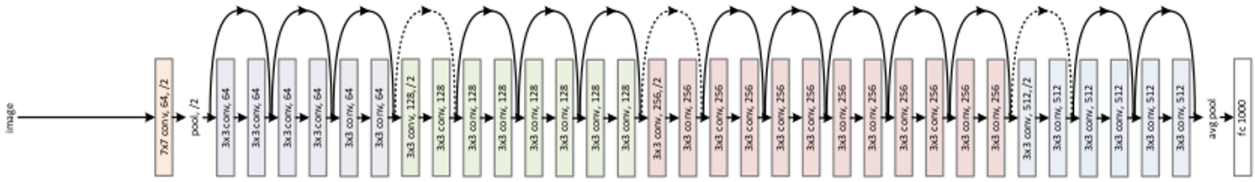

Fig. 5. Illustration of a typical residual block of ResNet (50 layers), where each layer consists of Convolutional (conv), Rectified Linear Unit (RELU) and Batch Normalization (Batch Norm) layers.

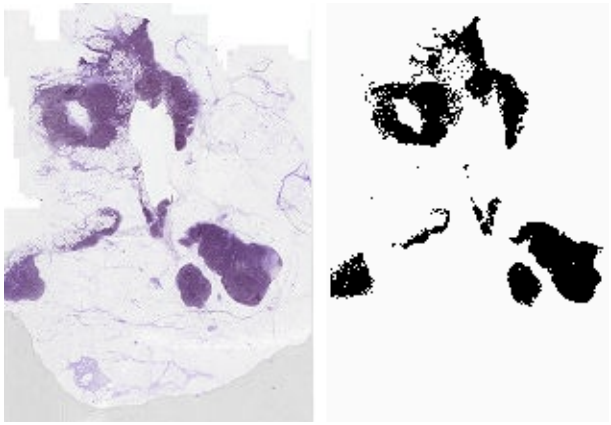

Fig. 6. Visualization of tissue regions detected in image preprocessing. Left: An original WSI. Right: Detected tissue regions delineated with the black regions.

are on WSI are divided into 256×256 patches. The patch is labeled as a tumor if over 75% pixels in the patch are annotated as a tumor. With these patches, we obtain enough samples to train the CNN for tumor pattern recognition. Figure 7 shows the sample of the WSI tiling.

C. Counstruction Dataset

As the area of the normal regions is more than the area of the tumor parts in each slide extracting specific random patches from tumor areas makes the dataset unbalanced. For normal patches, 25k patches are randomly extracted from both tumor and normal slides. Half of the normal patches are extracted from normal parts of tumor slides, another half are extracted from the normal slides. For tumor patches, 25k non-overlapped patches are selected from tumor slides. Overall, the model will be trained with 50k tumoral and normal patches with labels 1 and 0.

The training and validation sets are splitted into 80% /

Fig. 7. A sample of the WSI tiling report plot. Each box represents a 256 $\times$256 patches of the tumor region.

20% training/validation split for the final classifier.

D. Data augmentation

Images are obtained from two different centers and the classification of the WSIs has been determined to be a problem that needs a robust artificial intelligence approach capable of generalizing well. The variation of H&E stained color due to the various chemical preparation process performed by each center was addressed with an extensive color augmentation method. We can improve our CNN model by adding more data using a proper image augmentation approach. The idea behind image augmentation is that taking existing images from our training dataset and implementing some image transformation operations to them, such as rotation, shearing, translation, zooming, and so on, to produce new, altered versions of existing images. Due to these random transformations, we will not get the same images, and it will increase accuracy to feed in these new images to our model during training.

E. Patch-based classification

The patch-based classification model is the most important part of the system. This model takes patches as input and output the probabilities that it contains tumor regions. CNNs have proven to be exceptional at the task of image recognition. To get a spatially invariant, one of the recent model is used by using data augmentation. The open-source VGG-19, ResNet50, SqueezNet, DenseNet, and AlexNet, architectures are employed. Most of these architecture was designed for the ImageNet competition, which included millions of every-day images, however, the architecture and filters have proven to be adaptable to different imaging applications. Transfer learning is used for this project. Transfer learning is about borrowing CNN architecture with its pre-trained parameters from someone else. When we train our data on the top of the pre-trained parameters, we can easily reach to the target accuracy. To get the WSI probability heat-map, tissue regions were segmented as before and tiled into 224$\times$224 pixel-squared regions. These tiles were passed to the re-trained classifier and probabilities for each tile were collected. The framework of the classification is shown in Figure 8

F. Post processing

The results of small patches should be merged to determine the size and area of the tumor slides. Although this project does not focus on identifying pN-stage but the result should be shown as a WSI. One of the slides from St. Michael's Hospital is tested by this algorithm. The merged patches are merged in the Figure 9. The red areas show the patches with high probability of being tumor and the blue areas show the high probability of being normal. The probability of the green ares are more than blue areas and less than yellow areas, and the probability of the yellow areas are less than red parts and more than blue and green parts.

IV. Results

A. Patch-based classification

Patch-based classification is evaluated by using different networks. The accuracy of models using different architectures is shown in Table II. The highest accuracy of the patch-level classifier is 0.997 on the validation dataset using VGG-19 architecture. So, VGG-19 is selected to perform on the next step which is slide-based evaluation. VGG-19 is applied for training model for slide-based evaluation . The accuracy and loss function results are shown in Figures 10, 11, 12, 13.

B. Slide-based classification

For slide-based evaluation, coefficient matrix is calculated on the test set which consists of 100 slides from Camelyon2017 and 20 slides from Camelyon2016. The

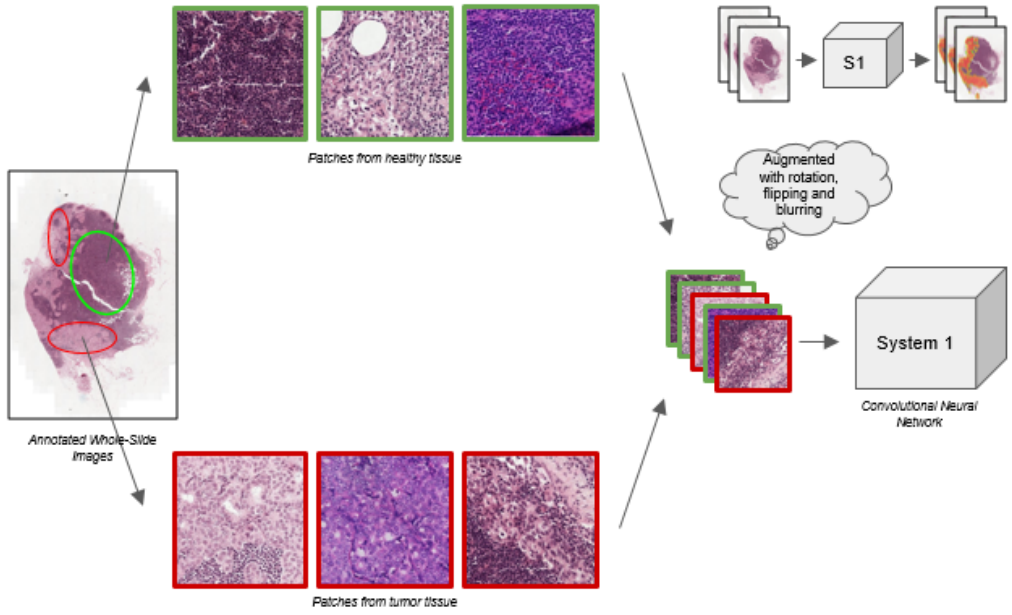

Fig. 8. Classification of the slides by extracting normal and tumor patches from training data and train the model using the specific CNN model.

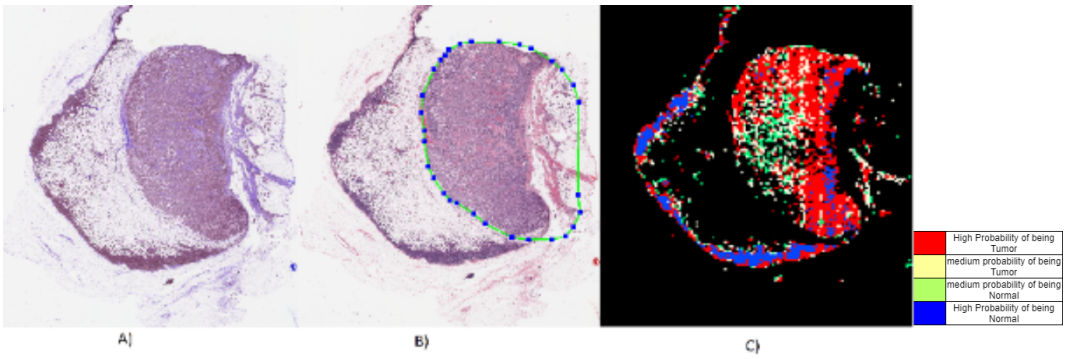

Fig. 9. A) A sample WSI from the SMH dataset B) The outline of the tumor region as annotated by the pathologist is in green. C) The results of the deep learning algorithm, red denotes areas with high probability of tumor cells and blue denotes areas with low probability of tumor cells.

confusion matrix shows that the method detects 55 tumor slides among 60 tumor slides, which shows that the true-positive value is acceptable. However, among 60 normal slides, 23 slides are recognized as tumor slides. The false-positive value is not satisfactory. The accuracy, sensitivity and Specificity are calculated with formula of the equations in 1, 2, 3. According to the confusion matrix, which is shown in Table III, accuracy, sensitivity, specificity precision are 76%, 91%, 61%, 70%, respectively, which means that the algorithm works well to find the tumor slides but it does not work well to detect normal slides.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

TABLE II

PATCH BASED CLASSIFICATION RESULTS

Network	Pre-trained	Accuracy
ResNet50	ImageNet	0.982
VGG-19	ImageNet	**0.997**
squeezNet	ImageNet	0.914
DenseNet	ImageNet	0.986
AlexNet	ImageNet	0.918

TABLE III

SLIDE-LEVEL CLASSIFICATION CONFUSION MATRIX.

	Predicted positive	Predicted negative	
Actual positive	TP = 55	FP = 23	Sensitivity = 91%
Actual negative	TN = 37	FN = 5	Specificity = 61%
	Precision = 70%	Negative Predictive = 88%	Accuracy = 76%

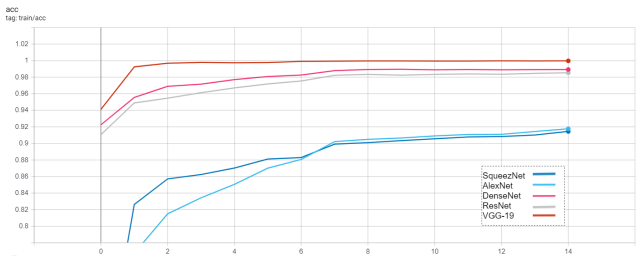

Fig. 10. The training accuracy on different architectures.

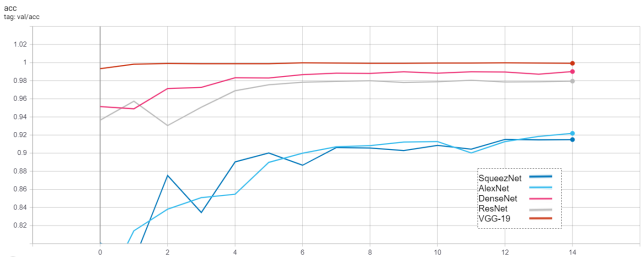

Fig. 11. The validation accuracy on different architectures.

V. DISCUSSION

This project shows the study using CNNs in histopathology of breast malignancy. Results express that the deep learning framework obtains the cell morphologies related to breast cancer staging and distinguishes histopathology slides with tumor cells. The analysis framework demands no human intervention, and the results are beneficial for establishing a system for tumor pathology analysis. All of the networks were trained on the ImageNet data, which shows no similarity to histopathology images. However, I explained that the method generated excellent classification performance when trained with an adequate amount of training data. This shows the generalizability of pre-trained convolutional neural networks and parameter fine-tuning through thousands of pathology training images. As anticipated, when the number of training cases is extensive, CNNs work best. Whole-slide histopathology of cancers presents a suitable opening for convolutional neural network applications, as one slide generally shows thousands of tumor cells. Also, different types of tumor cell diversity can be represented in several parts of the same WSI. The profusion of tumor cells and their varieties per slide provided sufficient data for training the network for recognizing the lymph node with breast cancer metastasis. Further studies can explore the utility of computer vision methods in the histopathology diagnosis of primary breast cancer and the classification of other clinically important phenotypes. Future studies are needed to investigate the utility of implementing an automated image classification system in the clinical setting. CNN's achieved better internal than external performance. When models were tested on internal data, they performed better than testing on new pooled data from other centers. So, the methods described here need to be further validated in different cohorts with larger sample sizes. The procedure would typically require a thorough microscopic assessment by pathologists. Therefore, the generalization of the automated computer aided device aided method would reduce the workload of pathologists and reduce the subjectivity in diagnosis in the evaluation of the pN-stage in breast cancer patients. Overall our study demonstrated the utility of convolutional neural networks in classifying lymph node histopathology images. The machine learning system presented here can provide decision support to pathologists and reclassify patients with ambiguous histopathology presentations. The method

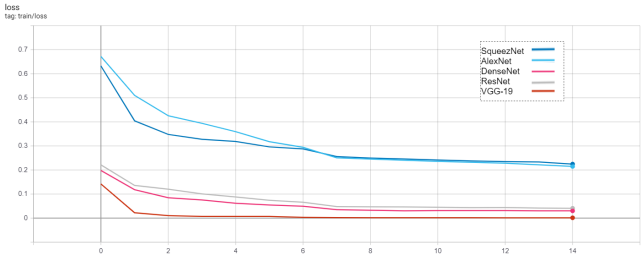

Fig. 12. The train loss on different architectures.

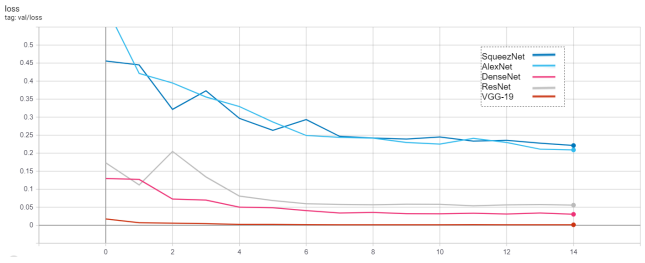

Fig. 13. The validation loss on different architectures.

was developed and evaluated using data from a few centers. Recent work has suggested that image classification CNNs may not generalize to new data as well as previously believed. CNN's achieved better internal than external performance. Testing the method with different centers, and different stain protocols and scanners, decrease the accuracy of the system. Though, it is difficult to say what is better given the high inter and intra agreement within breast cancer grading. I will focus on generalization both on segmentation and classification methods.

REFERENCES

[1] Rui-min Ma, Fan Yang, Du-ping Huang, Min Zheng, and Yi-luan Wang. The prognostic value of the expression of smc4 mrna in breast cancer. *Disease Markers*, 2019, 2019.

[2] Case Survival Time. Seer cancer statistics review 1975-2008. 2011.

[3] Stephanie Robertson, Hossein Azizpour, Kevin Smith, and Johan Hartman. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Translational Research*, 194:19–35, 2018.

[4] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.

[5] Peter Bandi, Oscar Geessink, Quirine Manson, Marcory Van Dijk, Maschenka Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee, Kyunghyun Paeng, Aoxiao Zhong, et al. From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE transactions on medical imaging*, 38(2):550–560, 2018.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[11] Babak Ehteshami Bejnordi, Geert Litjens, Nadya Timofeeva, Irene Otte-Höller, André Homeyer, Nico Karssemeijer, and Jeroen AWM van der Laak. Stain specific standardization of whole-slide histopathological images. *IEEE transactions on medical imaging*, 35(2):404–415, 2015.

[12] Alexander Rakhlin, Alexey Shvets, Vladimir Iglovikov, and Alexandr A Kalinin. Deep convolutional neural networks for breast cancer histology image analysis. In *International Conference Image Analysis and Recognition*, pages 737–744. Springer, 2018.

[13] Chongyang Cui, Shangchun Fan, Han Lei, Xiaolei Qu, and Dezhi Zheng. Deep learning-based research on the influence of training data size for breast cancer pathology detection. *The Journal of Engineering*, 2019(23):8729–8732, 2019.

[14] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

[15] Hussam Qassim, David Feinzimer, and Abhishek Verma. Residual squeeze vgg16. *arXiv preprint arXiv:1705.03004*, 2017.

[16] Zixu Zhao, Huangjing Lin, Hao Chen, and Pheng-Ann Heng. Pfa-scannet: Pyramidal feature aggregation with synergistic learning for breast cancer metastasis analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 586–594. Springer, 2019.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[19] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[20] Yufeng Zheng, Clifford Yang, and Alex Merkulov. Breast cancer screening using convolutional neural network and follow-up digital mammography. In *Computational Imaging III*, volume 10669, page 1066905. International Society for Optics and Photonics, 2018.