Histopathologic Cancer Detection

Varan Singh Rohila
Dept. of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, U.S.A
vrohila@stevens.edu

2nd Neeraj Lalwani
Dept. of Mathematical Sciences
Stevens Institute of Technology
Hoboken, U.S.A.
nlalwan1@stevens.edu

3rd Lochan Basyal
Dept. of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, U.S.A.
lbasyal@stevens.edu

***Abstract*—Early diagnosis of the cancer cells is necessary for making an effective treatment plan and the health safety of a patient. Nowadays, doctors usually use a histological grade that pathologists determine by performing a semi-quantitative analysis of the histopathological and cytological features of hematoxylin-eosin (HE) stained histopathological images. This research contributes a potential classification model for cancer prognosis to efficiently utilize the valuable information underlying the HE-stained histopathological images. This work uses the PatchCamelyon benchmark datasets and trains them in a multi-layer perceptron and convolution model to observe the model's performance in terms of precision, Recall, F1 Score, Accuracy, and AUC Score. The evaluation result shows that the baseline convolution model outperforms the baseline MLP model. Also, this paper introduced ResNet50 and InceptionNet models with data augmentation where ResNet50 able to beat the state-of-the-art model. Furthermore, majority vote and concatenation ensemble were evaluated and provides the future direction of using transfer learning, and segmentation to understand the specific features.**

I. Introduction

Cancer has become one of the major health concerns worldwide and early diagnosis of cancer cells plays a vital role in the effective treatment planning of the patient. Cancer screening using breast tissue biopsies aims to distinguish between benign and malignant lesions. However, manual assessment of large-scale histopathological images is a challenging task due to the variance in appearance, heterogeneous structure, and textures[1]. Such a manual analysis is laborious, time-consuming, and often dependent on subjective human interpretation. For this reason, the concept of cancer detection with the analysis of histopathological images with machine learning algorithms provides a significant direction in the research of early cancer diagnosis.

In recent years, deep learning outperformed state-of-the-art methods in machine learning and medical image analysis tasks, including classification, detection, segmentation, and computer-based diagnosis. The PatchCamelyon benchmark dataset as histopathological images has been used for model training and further computation in this research project. These datasets' images are extracted from the histopathologic scans of lymph nose sections. For implementing the machine learning model, this paper mainly focuses on two approaches: a multi-layer-perceptron model and a simple convolutional model for image classification. The model's performance is observed with the Precision, Recall, F1 Score, Accuracy, and AUC Score.

The rest of the paper is organized so that related work has been discussed in section II. Furthermore, the proposed solution includes the description of the dataset, machine learning algorithms, and implementation details presented in section III. Similarly, the model comparison and the future direction are discussed in IV and V. The purpose of this research is concluded in section VI.

II. Related Work

This section summarizes existing works in the field of histopathology cancer detection and image classification in general.

A. Histopathology Cancer Detection

Various image classification approaches have been proposed for automatic cancer detection. The authors of [2] presented the concept of transfer learning and deep feature extraction. The two models, AlexNet and Vgg16, are considered for feature extraction, and AlexNet is used for further fine-tuning. Furthermore, a support vector machine(SVM) is used for the classification of the images into benign and malignant classes. This research used the BreaKHis dataset for the experiment and calculated the accuracy scores for performance evaluation.

Other authors [3] proposed an ensemble transfer learning (ETL) framework for classifying well-differentiated, moderately differentiated, and poorly differentiated cervical histopathological images. First, the author developed a transfer learning structure based on Inception V3, Exception, VGG16, and Resnet 50. Then, an ensemble learning strategy based on weighted voting was introduced to improve classification performance. This research claims that the problem of histopathological cancer detection can be solved by adopting the concept of deep multiple instances learning that integrates deep convolutional neural networks and multi-instance learning.

Continuing with the ensemble techniques, the authors of [4] proposed an ensemble model which adapts three pre-trained CNNs, namely VGG19, MobileNet, and DenseNet. The ensemble model is used for the feature extraction, and a multi-layer perceptron classifier was used to perform the classification task. Furthermore, diverse pre-processing and CNN tuning techniques, stain-normalization, data augmentation, hyperparameter tuning, and fine-tuning were used to train the model. Author of this research-validated four publicly available benchmark datasets, i;e., ICIAR, BreakHis, PatchCameleon, and Bioimaging.

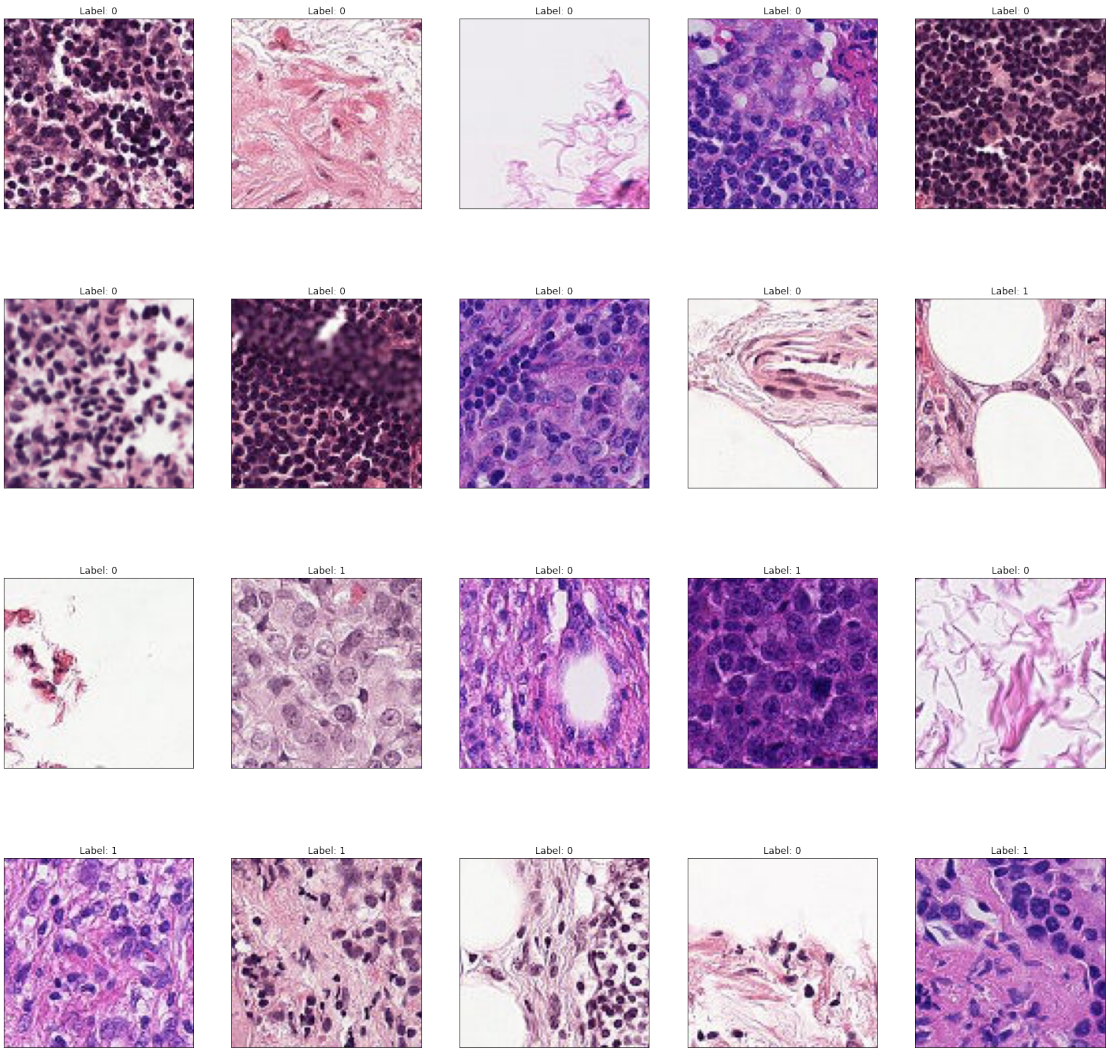


Fig. 1: Examples From the Dataset.

Lastly, this study [5] suggests a deep ensemble model for the binary classification of breast histopathology pictures of benign and malignant lesions based on image-level labeling. The training, validation, and test sets of the BreaKHis dataset are randomly assigned. The numbers of benign and cancerous samples are then balanced using data augmentation techniques. VGG16, Xception, ResNet50, and DenseNet201 are chosen as base classifiers based on their performance in transfer learning and the complementarity between networks. Image-level binary classification in the ensemble network model is accurate to 98.90%. This approach is experimentally evaluated on the same dataset with the most recent MLP models to confirm its capabilities. In classification tasks, the Ensemble model has a 5-20% advantage, highlighting its extensive capabilities. Just as a few doctors should be referred to before reaching any conclusion, these works showcase the importance of ensemble learning in combining the prowess of different models.

To understand the scope and landscape of the issue at hand, [6] reviewed the detection methods of histopathological cancer cells and predicted the future development trends to guide follow-up research. This paper focuses that machine vision overcomes the disadvantages of traditional detection methods in cancer detection and can help pathologists improve detection accuracy. Furthermore, the author presented the workflow of the machine vision detection system, image acquisition, pre-processing techniques, segmentation, feature extraction, and classification. The system depicted the microscope-mounted digital cameras or scanners used to obtain histopathological images. The pre-processing of the images can be performed through enhancement and color normalization techniques. The author reviewed the various segmentation techniques, threshold segmentation, active contour, clustering, and watershed. After segmentation, feature extraction can be performed with shape features, HSV, and Gray symbiotic. In the last step of the machine vision detection system, the author discussed the supervised and unsupervised classification for finding Benign and Malignant. At every stage, couple of unique techniques can be applied for better overall results.

B. Image Classification

Image classification refers to extracting useful information from an image and then classifying it based on certain at-

tributes. The most widely common image classification task is the classification of cats and dogs. Earlier versions of image classification models found even such trivial problems challenging. The introduction of convolution operations and deep neural networks advanced state-of-the-art of image classification. Today, new architectures are evaluated on the ImageNet dataset [7]. It has 1000 categories for image classification. Recent state-of-the-art image classification models are:

- **ResNet** Deep convolution models have the issue of vanishing gradients where the gradient becomes so small during backpropagation and weights aren't updated. To counter that issue, authors of [8] introduce skip connections in their network, which adds the information from previous layers to the current layer, thus reducing the problem of vanishing gradient descent.
- **DenseNet** Inspired by the skip connections introduced in [8], authors of [9] use those skip connections by concatenation the previous layer with the current layer. This has an increased performance from the addition skip layers of Resnet.
- **InceptionNet** Complexity of convolution models increase as the number of filters increase. The authors of [10] introduce 1x1 convolution operation that helps change the network's filter size. This reduces complexity and offers greater control of the architecture.
- **ViT or Vision Transformer** Transformer models have proven successful in Natural language problems. The authors of [11] utilize the transformer model in image classification. To create the input embeddings, they divide the image into small image patches, project them in linear space and then encode their position embeddings.

III. Our Solution

A. Description of Dataset

The dataset we considered for to evaluate our approach is PatchCamelyon benchmark dataset [12]. It consists of 220,025 color images (96 x 96px) extracted from histopathologic scans of lymph node sections. Each image is annotated with a binary label indicating the presence of metastatic tissue. A positive label indicates that the center 32x32px region of a patch contains at least one pixel of tumor tissue. Tumor tissue in the outer region of the patch does not influence the label. The images from the data are shown in 1. The dataset is divided into the train, and test sets, with the test set being 25% of the dataset. The data statistics are tabulated in Table I.

Statistics	Train Set	Test Set
Positive Labels	66,837	22,280
Negative Labels	98,181	32,727
Total	165,018	55,007

TABLE I: Dataset Statistics

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
flatten (Flatten)            (None, 27648)             0
_________________________________________________________________
dense (Dense)                (None, 768)               21234432
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 769
=================================================================
Total params: 21,235,201
Trainable params: 21,235,201
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2: MLP Baseline Model.

B. Machine Learning Algorithms

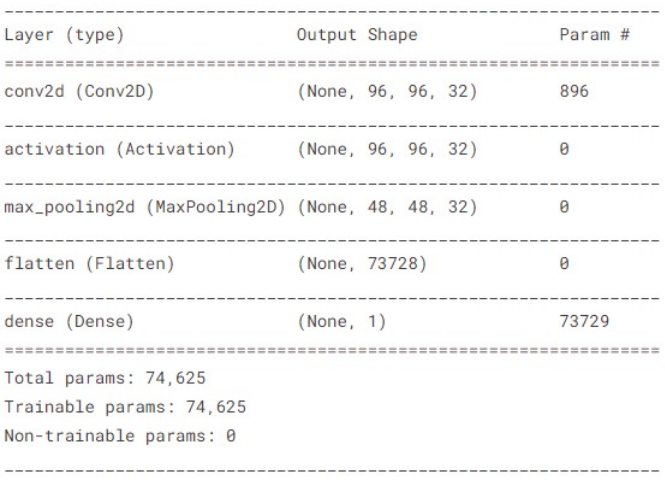
For our baseline models, we consider a multi-layer perceptron model and a simple convolutional model for image classification. The input of both these models is (96px, 96px, 3), where 3 corresponds to the RGB image channels. Briefly describing the baseline models below:

- **Multi-layer Perceptron Model**: Simple ANN models have been proven to perform well on image tasks such as MNIST Handwriting Benchmark Dataset. Hence, we consider a single-layer MLP model with 768 hidden nodes with relu activation. A single node with sigmoid activation outputs the predictions.
- **Convolution Model**: A convolution is the simple application of a filter to an input that results in activation. Repeated application of the same filter to an input results in a map of activations called a feature map, indicating the locations and strength of a detected feature in an input, such as an image. We consider a single-layer convolutional model with 32 filters, (3, 3) filter size with relu activation. After that, Max Pooling is applied with (2, 2) filter size. After flattening and adding a single output node with sigmoid activation, the output is calculated.
- **ResNet50 Model**: We utilize data augmentation with the ResNet50 model. To the output of the ResNet50 model, we add a Global Max Pooling layer and Global Average Pooling Layer. Finally, these three outputs are concatenated and attached to a single output node. A dropout layer of 0.2 is also added to the model.
- **Inception Model**: Similar to the ResNet50 model, inception model is trained and evaluated.
- **Majority Vote Ensemble Model**: We combine the ResNet50 and Inception model with majority voting. This is a hard ensemble.
- **Concatenation Ensemble Model**: We combine the ResNet50 and Inception models using concatenation and train them together as a joint Neural Network. This is a soft ensemble.

The number of parameters of both the models is shown in figure 2 and figure 3, respectively.

Models	Precision	Recall	F1 Score	Accuracy	AUC Score
Baseline MLP	0.677	0.567	0.617	0.715	0.774
Baseline Convolution	0.791	0.731	0.760	0.812	0.875
ResNet50	**0.950**	**0.932**	**0.941**	**0.952**	**0.988**
InceptionNet	0.929	0.920	0.925	0.925	0.983
Majority Vote	0.905	0.969	0.936	0.946	-
Concatenation Ensemble	0.922	0.927	0.925	0.939	0.982
Model proposed in [4]	0.957	0.952	0.955	0.946	-

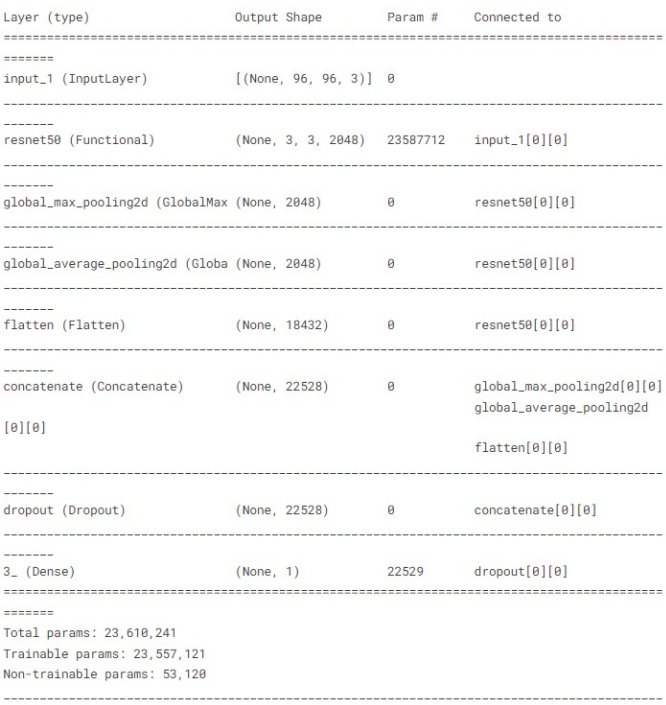
TABLE II: Baseline Results



Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 96, 96, 32)	896
activation (Activation)	(None, 96, 96, 32)	0
max_pooling2d (MaxPooling2D)	(None, 48, 48, 32)	0
flatten (Flatten)	(None, 73728)	0
dense (Dense)	(None, 1)	73729

Total params: 74,625
Trainable params: 74,625
Non-trainable params: 0

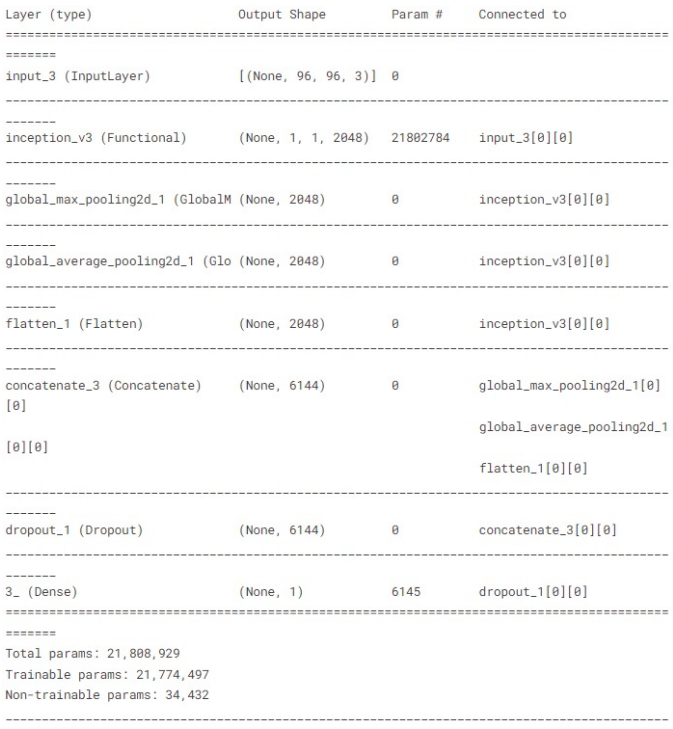
Fig. 3: Convolution Baseline Model.



Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 96, 96, 3)]	0	
resnet50 (Functional)	(None, 3, 3, 2048)	23587712	input_1[0][0]
global_max_pooling2d (GlobalMax	(None, 2048)	0	resnet50[0][0]
global_average_pooling2d (Globa	(None, 2048)	0	resnet50[0][0]
flatten (Flatten)	(None, 18432)	0	resnet50[0][0]
concatenate (Concatenate)	(None, 22528)	0	global_max_pooling2d[0][0] global_average_pooling2d[0][0] flatten[0][0]
dropout (Dropout)	(None, 22528)	0	concatenate[0][0]
3_ (Dense)	(None, 1)	22529	dropout[0][0]

Total params: 23,610,241
Trainable params: 23,557,121
Non-trainable params: 53,120

Fig. 4: ResNet50 Model.



Layer (type)	Output Shape	Param #	Connected to
input_3 (InputLayer)	[(None, 96, 96, 3)]	0	
inception_v3 (Functional)	(None, 1, 1, 2048)	21802784	input_3[0][0]
global_max_pooling2d_1 (GlobalM	(None, 2048)	0	inception_v3[0][0]
global_average_pooling2d_1 (Glo	(None, 2048)	0	inception_v3[0][0]
flatten_1 (Flatten)	(None, 2048)	0	inception_v3[0][0]
concatenate_3 (Concatenate)	(None, 6144)	0	global_max_pooling2d_1[0][0] global_average_pooling2d_1[0][0] flatten_1[0][0]
dropout_1 (Dropout)	(None, 6144)	0	concatenate_3[0][0]
3_ (Dense)	(None, 1)	6145	dropout_1[0][0]

Total params: 21,808,929
Trainable params: 21,774,497
Non-trainable params: 34,432

Fig. 5: InceptionNet Model.

C. Implementation Details

The models are trained on the train set for 50 epochs and then evaluated on the test set. Adam is the optimizer for both models, with a 0.0005 learning rate for the baseline MLP and 0.001 for the baseline convolution model. For the ResNet50, the Inception model, and the ensemble model, are trained with a 0.00003 learning rate. Early stopping with patience 5 is used to avoid overfitting.

For data augmentation, random vertical and horizontal flips are applied to the training data with a re-scale of 1/255.

IV. Comparison

The models are tested on these metrics:

- **Precision**: Precision refers to the number of true positives divided by the total number of positive predictions.
- **Recall**: The recall is the measure of our model correctly identifying True Positives.

- **F1 Score**: It is the harmonic mean of Recall and Precision.
- **Accuracy**: Accuracy is the fraction of predictions our model got right.
- **AUC Score**: AUC score represents the degree or measure of separability.

The evaluation results are tabulated in table II. The baseline convolution model outperforms the baseline MLP model in all metrics confirming the prowess of simple convolution networks in image classification. Moreover, due to the weight-sharing property of convolutions, the number of trainable parameters is considerably less than the baseline MLP model.

V. Future Directions

In this work, we evaluate the performances of baseline models on the Patch-Cameleyon dataset. The classification performances can be further increased by using the following:

- **Transfer Learning**: Transfer learning refers to the technique of applying already learned knowledge from one application area to another with shared weights and retraining some layers of the model.
- **Segmentation**: In medical image applications, segmentation extracts useful information about the subject, i.e., the tissue, which can help the model to specific features and ensures the model's focus.

VI. Conclusion

Automatic detection of cancer tissues can help doctors identify tumor tissues and act accordingly. This paper proposes and evaluates baseline models for detecting cancer tissue in lymph nodes. The baseline convolution model achieves an accuracy of 81.2%, which is quite good for a single-layer model indicating bigger, better models equipped with other techniques such as data augmentation, transfer learning, and segmentation can achieve state-of-the-art performance on this benchmark dataset. As we advance with that idea, ResNet50 and InceptionNet models were trained with data augmentation. Finally, their majority and Concatenation ensemble were also evaluated on the dataset. Extensive experimentation showcases that the ResNet50 model trained with proper data augmentation is able to beat the state-of-the-art model. In the future, transfer learning and segmentation techniques can be utilized to create a more general and efficient model.

References

[1] Li, Chao, et al. "Weakly supervised mitosis detection in breast histopathology images using concentric loss." Medical image analysis 53 (2019): 165-178.

[2] Deniz, Erkan, et al. "Transfer learning based histopathologic image classification for breast cancer detection." Health information science and systems 6.1 (2018): 1-7.

[3] Udendhran, R., B. Sreedevi, and G. Sneha. "A Unified and Semantic Model Approach for Histopathologic Cancer Detection Based on Deep Double Transfer Learning." 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE, 2022.

[4] Kassani, Sara Hosseinzadeh, et al. "Classification of histopathological biopsy images using ensemble of deep learning networks." arXiv preprint arXiv:1909.11870 (2019).

[5] Yuchao Zheng, Chen Li, Xiaomin Zhou, Haoyuan Chen, Hao Xu, Yixin Li, Haiqing Zhang, Xiaoyan Li, Hongzan Sun, Xinyu Huang, Marcin Grzegorzek. "Application of Transfer Learning and Ensemble Learning in Image-level Classification for Breast Histopathology, Intelligent Medicine" (2022) ISSN 2667-1026.

[6] He, Wenbin, et al. "Progress of Machine Vision in the Detection of Cancer Cells in Histopathology." IEEE Access 10 (2022): 46753-46771.

[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248–255).

[8] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[9] Huang, G., et al. "Densely Connected Convolutional Networks. arXiv. 2016 doi: 10.48550." arXiv preprint arXiv.1608.06993 1608.

[10] Szegedy, Christian, et al. "Going deeper with convolutions. 2014." arXiv preprint arXiv:1409.4842 10 (2014).

[11] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[12] Veeling, Bastiaan S., et al. "Rotation equivariant CNNs for digital pathology." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2018.