# Comparing Training Methods of Convolutional Neural Networks in Identifying Breast Cancer Metastasis in Histopathologic Scans of Lymph Node Sections

**Siddharth Kini**

*Crooms Academy of Information Technology*

## Abstract

Metastatic cancer is responsible for between 66 to 90 percent of cancer fatalities. With Digital Pathology becoming more widespread, computers can utilize convolutional neural networks to identify metastatic tissue in histopathologic scans with a high accuracy. This paper sought to compare two methods of training these convolutional networks, training from scratch compared to transfer learning with a pre-trained model. It was found that the model utilizing transfer learning had an F1 score of 92.47 percent, 1.76 percent higher than the best performing model developed from scratch.

## 1. Introduction

Cancer is a group of diseases which are characterized by the uncontrolled growth and spread of cells. If this spread is not controlled, the result can be fatal. [1] Many factors can lead to the development of cancer, including lifestyle choices, such as smoking and alcohol usage, or genetic factors, such as an inherited mutation. [1, 2] The metastases of Cancer are responsible for between 66 to 90 percent of its fatalities. [3] Metastasis involves cancerous cells detaching from their initial tumor and entering the circulatory or lymphatic systems, evading immune attack, and proliferating in a separate part of the body. [4]

With Digital Pathology becoming more widespread, algorithms can be developed to identify cancers in whole-slide images. One such algorithm is a deep learning convolutional neural network. These networks have a series of layers that are assigned weights and biases, numerical values, to classify or detect objects.

Conventionally, convolutional models have been developed from scratch, however there is an alternative to having to train an entire classifier: Transfer Learning.

Transfer learning utilizes Pre-trained models, which are classifiers that have been trained on a large dataset and can extract features well. Transfer learning is the process of training the last few layers of these models to classify new objects, by only training the prediction layers, but maintaining the layers that extract features.

*Related Works*

Utilizing convolutional neural networks to identify metastatic cancer has been done before, through the Camelyon16 and Camelyon17 competitions, and the publicly available and commonly utilized PatchCamelyon Dataset. The Winner of the 2016 competition utilized a GoogleLeNet pre-trained model.

*Objectives*

This paper seeks to compare training methods of convolutional neural networks for the task of identify metastatic tissue in histopathologic scans of lymph node sections.

## 2. Methods

The hardware used includes: Intel Xeon CPU at 2.00 GH Hyperthreaded Single Core, 13GB Ram, Nvidia P100

The software used includes: An IPython Notebook running Python version 3.7.6, with the following modules: NumPy, Pandas, MatPlotLib, Scikit Learn, Keras.

An algorithm to obtain the overall methodology used in this paper is as follow:

- Collect and Preprocess Image Data
    - Collect Image Data
    - Sample Data and Equalize Labels
    - Conduct Train, Validation, Test Split
- Development of Classifier from Scratch
    - Manually define layers of Convolutional Neural Network
    - Train multiple classifiers using different configurations of layers and parameters.
- Development of Classifier from Pre-trained Convolutional Neural Network
    - Extract features using convolutional neural network
    - Retrain Final Layers of the Classifier for the specific application

*Collect and Preprocess Image Data*

The PatchCamelyon benchmark dataset was utilized for the training of the classifiers to detect metastatic breast cancer. The dataset contains 327,680 color images, of size 96 x 96 pixels, extracted from histopathologic scans of lymph node sections. 107,655

duplicates were removed, resulting in a dataset of 220,025 images. Each image was given a binary label based on the presence of metastatic tissue. Images were labelled as having metastatic cancer if the center 32 x 32 pixel region of the image had a

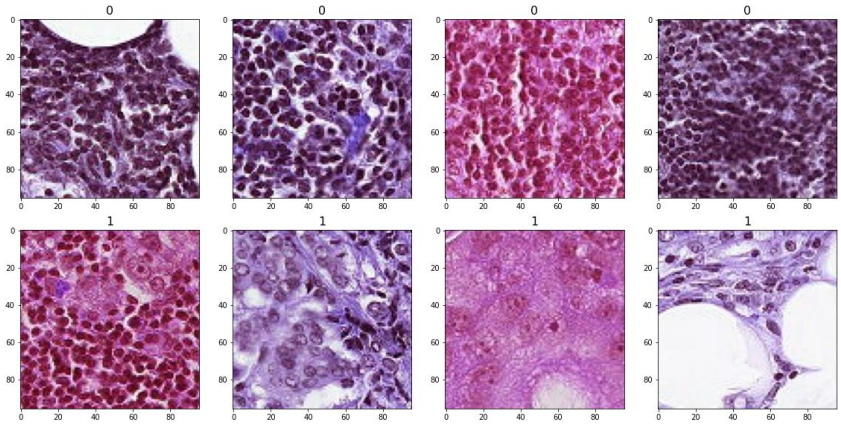


*Figure 1 Sample of Images from the Patch Camelyon dataset with their associated labels.*

metastatic tumor, and tumor tissue outside of this region did not influence the label.

The Dataset contains 130,908 images without metastatic cancer and 89,117 images with breast cancer. A stratified down sample was conducted on these images, resulting in a dataset of 120,000 images with 60,000 images for each label (See figure 2). A train-validation-test split was conducted, with 10 percent for testing, 9 percent for validation, and 81 percent for training. The image's RGB data channel was

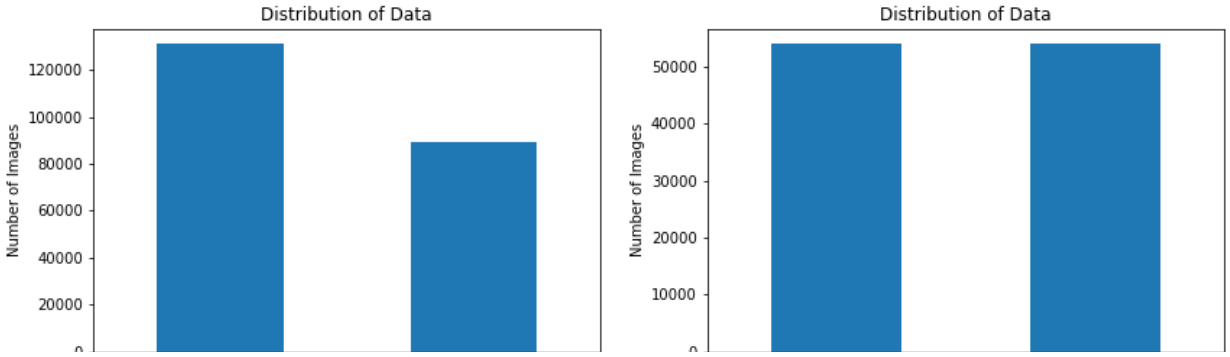


*Figure 2: Distribution of the data before and after a stratified down sample was taken from the dataset.*

resized from 0-255 to 0-1. This data was fed to the models during the training stage using a DataImageGenerator instance.

*Development of Classifier from Scratch*

Nine different model configurations were created and validated. Each model used the Keras Sequential model format, which involves adding layers onto a model. The layers used in each configuration include Conv2D, Dropout, MaxPooling2D, Dense, and Flatten. The models were created based on "sections" and the number of convolutional layers per "section". A section is defined as a series of Conv2D layers which ends in a Dropout and a MaxPooling2D layer. The last "section" before the dense connections does not include a MaxPooling2D layer. The configurations trained had 2, 3, and 4 sections with 1, 2, and 3 Convolutional, Conv2D, layers per section. Each Conv2D layer utilizes a rectified linear unit activation function. The configurations share a common Dense Layer configuration: A rectified linear unit activated Dense layer, a Dropout layer, and a Dense prediction layer, utilizing a Sigmoid activation function.

The models were compiled using the Adam optimizer, binary crossentropy loss, and accuracy as the metric. The models were trained for 15 Epochs. Training data was logged and viewed through Keras' Tensorboard functionality.

*Development of Classifier from Pre-trained Convolutional Neural Network*

A VGG19 pre-trained classifier was trained for this application using transfer learning. The model was added using the Keras Sequential model format. Every layer in the pre-trained model except the last 12 was frozen, and the last 12 layers were retrained for

the purpose of this application. The model's prediction layer consisted of the following: a Dense rectified linear activated layer, a Dropout layer, a Batch Normalization layer, and a Dense sigmoid activated layer.

The models were compiled using the Adam optimizer, binary crossentropy loss, and accuracy as the metric. The models were trained for 15 Epochs. Training data was logged and viewed through Keras' Tensorboard functionality.

## 3. Results

To evaluate the models, the best performing model developed from scratch was compared against the VGG19 transfer learning model. They were tested against 12,000 images split earlier from the Patch Camelyon dataset. Their predictions and the actual labels on these images were plotted on a confusion matrix, from which multiple statistics were derived. A confusion matrix contains the following categories: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). For the purposes of this paper the two models were compared with the following statistics:

- Accuracy: $\frac{tp + tn}{tp+tn+fp+fn}$
- Recall: $\frac{tp}{tp+fn}$
- Precision: $\frac{tp}{tp+fp}$
- $F_1$ Score: $\frac{2tp}{2tp+fp+fn}$

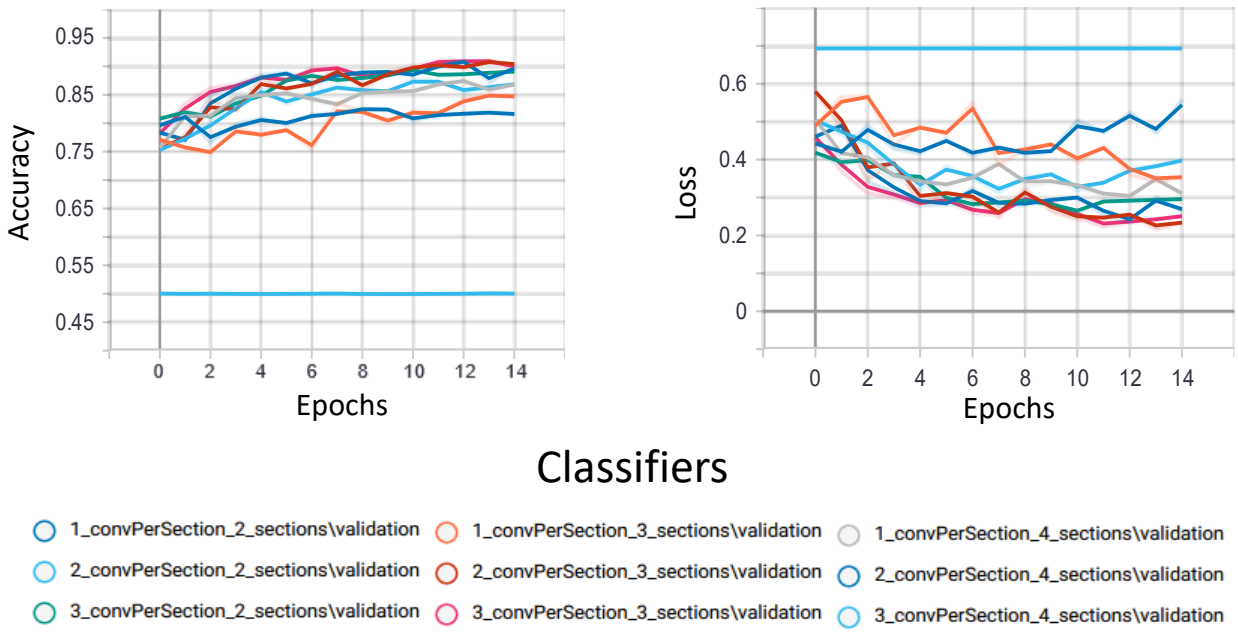


*Figure 3: Accuracy and Loss graphs of the 9 Classifiers trained from scratch.*

| Validation Accuracy and Loss of the 9 Classifiers | | |
|---|---|---|
| | Validation Accuracy | Validation Loss |
| 1 Conv Per Section 2 Sections | 81.58 | 0.5599 |
| 2 Conv Per Section 2 Sections | 86.96 | 0.4013 |
| 3 Conv Per Section 2 Sections | 89.12 | 0.2962 |
| 1 Conv Per Section 3 Sections | 84.68 | 0.3543 |
| 2 Conv Per Section 3 Sections | 90.31 | 0.2355 |
| 3 Conv Per Section 3 Sections | 89.78 | 0.2527 |
| 1 Conv Per Section 4 Sections | 87.03 | 0.3019 |
| 2 Conv Per Section 4 Sections | 90.00 | 0.2638 |
| 3 Conv Per Section 4 Sections | 50.02 | 0.6932 |

*Figure 4: Table of Accuracy and Loss Values of the 9 Classifiers*

| Classifier Statistics on Testing Data | | |
|---|---|---|
| | VGG19 | 2Conv3Sections |
| Accuracy (%) | 92.60 | 90.82 |
| Precision (%) | 90.87 | 89.72 |
| Recall (%) | 94.13 | 91.73 |
| F1 Score (%) | 92.47 | 90.71 |

*Figure 5: Classifier Statistics of the two models*

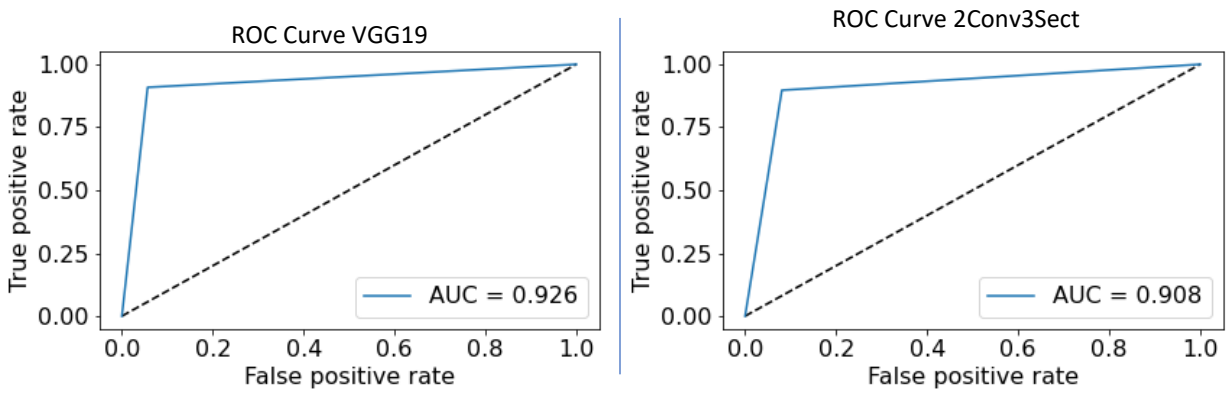


*Figure 6: Receiver Operating Characteristic Curve of the Two Models*

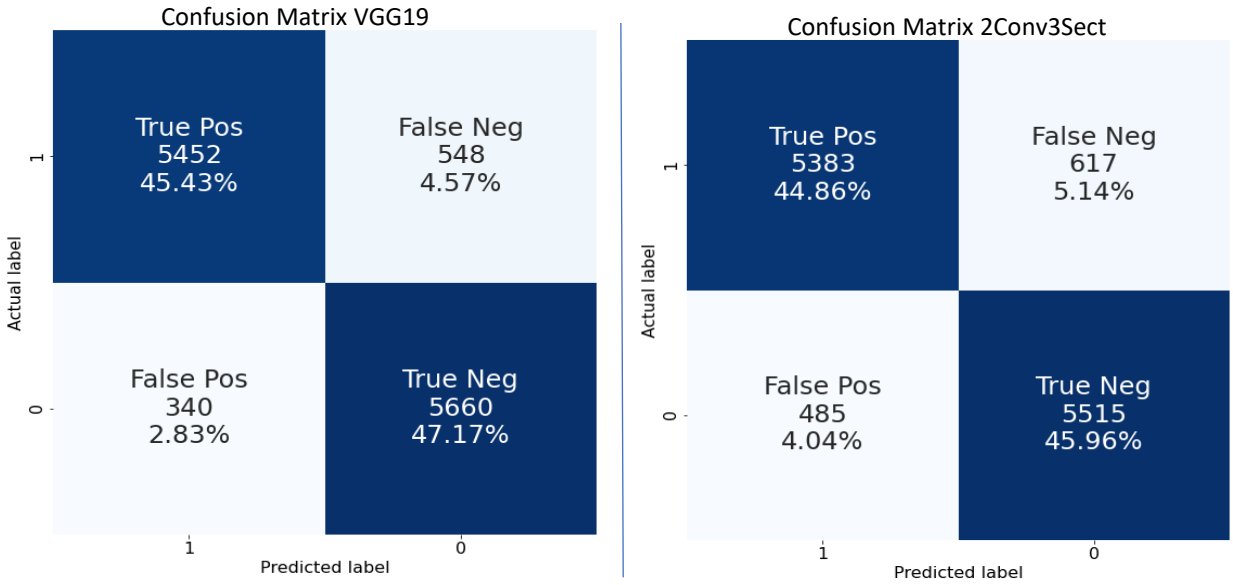


*Figure 7: Confusion Matrix of the Two Models*

## 4. Discussion

To find the best performing model developed from scratch, the training and validation data was reviewed and the model with the highest validation accuracy was saved (See Figure 3). The model with the highest accuracy had two convolutional layers per section with three sections, which will be referred to as "2Conv3Sect" in this paper. This model had an accuracy of 90.31 and a loss of 0.2355 on the validation data and was selected for test evaluation.

The Model was tested against 12,000 images from the histopathologic cancer dataset. The VGG19 model took 12m and 48s to predict the values for the images. The

2conv3sect took 40s. This shows that when predicting, the model trained from scratch with less layers predicted 12 minutes and 8 seconds faster than the transfer learning model. On the dataset, the VGG19 model had an accuracy of 92.59% and a loss of 0.2003 while the 2Conv3sect had an accuracy of 90.81% and a loss of 0.2308. This demonstrates that the transfer learning model had a higher accuracy, which is especially important for this application.

The Receiver Operating Characteristic Curves of the models were also plotted (See figure 7). These demonstrate the diagnostic ability of binary classifiers, by showing the correlation between true positive rate and false positive rate. A classifier with no predictive value is denoted by the dashed line. Both classifiers performed well as seen by their sharp positive curves. The model trained from scratch had an Area Under Curve (AUC) score of 90.8 percent, while the VGG19 model had an AUC score of 92.6 percent.

To summarize, the model that was trained with the VGG19 model using transfer learning performed better than the best performing model trained from scratch, 2Conv3Sect.

## 5. Conclusions

This paper presented a comparison of pre-trained models and models trained from scratch for the purpose of identify metastatic breast cancer in histopathologic scans of lymph node sections. It was seen that a simple pretrained model, VGG19, without fine tuning outperformed a model from scratch by an F1 score difference of 1.76 percent. Further research on this topic can improve the performance of both classifiers and can provide more insight into the differences between pretrained models and models developed from scratch, including time taken to train, time taken to predict, and the effect of the number of trainable parameters in a model.

## 6. Acknowledgements

The author thanks Dr. Josue N. Urbina (Crooms Academy of Information Technology) and Amar G. Kini for their guidance throughout the project, and Pravitha Ramanand for proofreading this report. The author would like to thank Sentdex for their deep learning tutorials.

## 7. References

**[1]** American Cancer Society. Cancer Facts & Figures 2019. Atlanta: American Cancer Society; 2019.

**[2]** Dorj, U.-O., Lee, K.-K., Choi, J.-Y., & Lee, M. (2018). The skin cancer classification using deep convolutional neural network. Multimedia Tools and Applications, 77(8), 9909–9924. https://doi.org/10.1007/s11042-018-5714-1

**[3]** Dillekås, H., Rogers, M. S., & Straume, O. (2019). Are 90% of deaths from cancer caused by metastases? Cancer Medicine, 8(12), 5574–5576. https://doi.org/10.1002/cam4.2474

**[4]** Metastatic Cancer: When Cancer Spreads. Accessed 12/17/2020 from: https://www.cancer.gov/types/metastatic-cancer

**[5]** Matthew D Zeiler, & Rob Fergus. (2013). Visualizing and Understanding Convolutional Networks.

**[6]** Torrey, L., & Shavlik, J. (2010). Transfer Learning. In Olivas, E. S., Guerrero, J. D., Martinez-Sober, M., Magdalena-Benedito, J. R., & Serrano López, A. J. (Ed.), Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques (pp. 242-264). IGI Global. http://doi:10.4018/978-1-60566-766-9.ch011