Histopathological Image Classification with Pre-Trained Deep Learning Models

Sandeep Sohal
Department of Computer Science
Yeates School of Graduate Studies
Ryerson University
Ontario, Canada
kaur.sohal@ryerson.ca

***Abstract*—The purpose of this research is to study different pre-trained deep learning algorithms to perform multi-class classification on Histopathological Image data. Transfer learning is said to reduce the complexity of the model by reusing learned parameters from a different dataset such as the ImageNet dataset. The concept of transfer learning by implementing pre-trained CNNs is examined in this research and the results are compared with a CNN built from scratch. The time is takes to train and the number of iterations it takes for the model to converge will be considered when evaluating the performance of the four proposed models. The three pre-trained CNN based models that are implemented in this research using the Keras library are as follows: VGG-16, ResNet-50, and the Inception-V3. The CNN model that is implemented in this study is 10 layers deep. An accuracy of 99% is achieved after fine-tuning the output layer of the VGG-16 model and training for only 10 epochs. This research verifies that the concept of transfer learning by using pre-trained deep learning frameworks will most certainly increase the performance of the model and yield better results with lesser training time.**

Keywords—deep learning frameworks, VGG-16, ResNet-50, Inception-V3, CNN, epochs, fine-tuning, output layer, accuracy, keras

I. INTRODUCTION

Misclassification of medical reports and images is always concerning and different studies are being conducted and published every year that try to address these issues in the medical domain. Primarily, this is due to the fact that experts involved in analyzing medical files are after all humans, and each medical expert will provide different treatment plans. Although they are trained, but different experts will interpret or diagnose a patient differently. This is where deep learning and machine learning models play an integral role and can assist medical professionals understand different medical files.

Medical images are known for being polluted with noise which makes them difficult to understand and contain intricate details that requires a lot of time to study. Thus, it can be concluded that misclassification is inevitable. There are different types of medical images including grayscale and colored images.

Histopathological images are analyzed in this paper to identify different lung and colon cancers. To obtain histopathological images a surgery is performed and biopsy is taken from a patient and this biopsy is given to the pathologist who is responsible for studying these tissues under a microscope. According to the Canadian Cancer Society approximately 25% of all cancer related deaths reported this year are due to lung cancer alone [1]. Colorectal cancer on the other hand is the 3rd most common cancer in Canada [2]. In a paper published in 2019, researchers found that pathologists disagree with the opinions of other pathologists almost 25% to 26% of the time [3]. This could potentially result in misdiagnosis. The researchers implemented deep learning models to classify histopathological images and compare their results with the results provided by 11 pathologists. The authors found that their CNN model performed better with an accuracy of 68% compared to the accuracy of the 11 pathologists which is 59.2% [3]. To know that there is a 40% chance of misclassification is troubling and this study suggests that deep learning models can be used to help pathologists analyze histopathological images.

This paper aims to address this issue by including the concept of transfer learning to classify lung and colon histopathological images. There is no doubt that transfer learning is gaining popularity and has been proven to improve the performance of deep learning models. The idea of transfer learning is essentially based on reusing learned parameters from a base model and base dataset to approximate a new problem [4]. It is this reusability of parameters that makes transfer learning immensely popular and it also reduces the computational complexity of any given problem. The time it takes for learning the optimal parameters of a model is essentially reduced. The medical imaging domain has particularly taken advantage of transfer learning and have implemented it to perform classification, segmentation, and recognition tasks to provide accurate diagnosis to patients.

The pre-trained models that are implemented in this research are trained on large datasets such as the ImageNet and the learned model parameters are used to perform the classification task in this paper. VGG-16, ResNet-50, and Inception-V3 are the three pre-trained deep learning frameworks implemented as well as a CNN model trained to learn the optimal parameters from scratch. These models are compared and analyzed. The VGG-16 model after fine-tuning achieved the highest test accuracy of 99% followed by 97% accuracy achieved using the CNN model with 10 layers after several hours of training and fine-tuning. The Inception V3 achieved 87% accuracy and the ResNet-50 achieved 85% accuracy with very minimal finetuning. These results verify that the pre-trained models outperform the CNN model due to the fact that the learning is much slower with CNNs built from scratch and it will take more time and iterations to learn the optimal parameters.

II. LITERATURE REVIEW

A study conducted at Yonsei University by researchers in the Engineering and Medicine department utilized transfer learning to classify Histopathological images [4]. Their

research makes use of Google's Inception V3 deep learning model to perform the classification task. The deep learning framework was retrained on their image dataset and to avoid overfitting the authors augmented their data by rotating their images by 90, 180, and 270 degrees. They also generated more data by flipping the images from top to bottom and from left to right. Random noise was also added to avoid generalization. They used a total of 11,184 histopathological images belonging to benign and malignant classes. With transfer learning their model achieved an overall accuracy of 83% for benign class and 89% for malignant class [4].

Kaur et al. [5] conducted research which implemented VGG-16 and transfer learning to identify brain images. The authors claim that they achieved 100% accuracy, sensitivity, and specificity scores. The authors first performed a binary image classification task to identify MR images into Normal and Abnormal and further classified abnormal images into 8 different classes: Sarcoma, Meningioma, Glioma, Picks, Huntington's, Alzheimer visual Agnosia, Metastatic bronchogenic carcinoma and Herpes encephalitis. Their work was compared with other deep learning and machine learning implementations and their model outperformed other models. They proposed VGG-16 and fine-tuned the last 3 layers of the model to learn features based on their problem [5].

Another research by Srivastav et al. [6] examines the performance of VGG16 to classify chest X-rays. The problem involves classification of X-rays into normal and with pneumonia. The dataset used contains 5,856 samples of chest X-rays segmented into two classes and to augment this dataset and generate synthetic images the authors proposed to use DCGAN. With this the authors generated a fake set of chest X-ray data. The real X-ray images and fake X-ray images were then fed to the VGG-16 model to perform the classification task. Using this methodology, the authors achieved an accuracy of 94.5% [6].

Soumik et al. [7] on the other hand also implemented transfer learning based deep learning model to categorize breast histopathological images as malignant or benign. Their method made use of the popular Inception V3 model and the achieved accuracy is 99.50% [7]. They did not perform any pre-processing on their dataset and suggested that transfer learning will still achieve a good accuracy on a dataset that is not pre-processed as long as it is a good quality dataset. They used the BreakHis dataset to verify their model's performance [7].

A conference paper published in 2019 employed the ResNet-50 architecture and trained their dataset using the pre-trained weights of ResNet-50 model [8]. The dataset used in this paper contains images of blood cell slides and the authors obtained 27,558 samples of infected and uninfected images and the proposed model was used to identify presence of Malaria. They first split their dataset and preprocessed it by resizing the images to 224 by 224. The input and output layers of the proposed model were changed according to the problem and they achieved a validation accuracy of 95.4% [8].

A group of researchers performed experiments to classify ultrasonic computed tomographic images using transfer-deep learning-based models [9]. They implemented the following models: Mobile Net, NasNet, Ameobanet, SOM, and Inception V3 model and achieved the best performance of 96% using the Ameobanet [9]. They first augmented their small dataset and resized their images to 256x256 before feeding it to the proposed deep learning models for classification. The Inception V3 model achieved the 3rd highest accuracy of 91.7%. The authors discussed that by utilizing transfer-deep- learning models they were able to reduce the time required to train their model and learn the features required to perform the classification task for their specific problem [9].

III. Methodology

There are several pre-trained models readily available with the Keras applications library [10]. The models implemented in this paper include VGG-16, ResNet-50, and Inception V3. These are all state-of-the-art models that achieved a high accuracy on a large base dataset.

A. VGG-16

Simonyan et al. [11] first introduced the VGG-16 model in their paper entitled 'Very Deep Convolutional Networks for Large-Scale Image Recognition' which was published in 2015. The purpose of this research was to beat the state-of-the-art models at performing image classification on the ImageNet dataset. This dataset is the base dataset utilized for the purpose of transfer learning. The proposed model achieved an over-all test accuracy of 92.7% on this large dataset [11]. The dataset is composed on 14 million images belonging to 1,000 categories. The default input size of the image is 224 x 224 for this architecture and the image should be RGB [11]. There is a way to pre-process a grayscale input data and expand the 1 channel to 3 channels using the keras library, but this is not necessary since the images used in this research are RGB images. The input and output layers should be designed according to the problem. This deep learning model consists of 16 layers which include: 13 convolution layers and 3 dense layers in the output. The kernel is 3 x 3 and the 3 dense layers were fine-tuned to achieve the best performance. The max pooling layers are 2x2.

Figure I [11]: VGG-16 Architecture

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

The input image is in the shape of (224,224,3) and the last 2 fully connected dense layers employ the ReLU activation function with the first dense layer having 200 neurons and the second layer having 100 neurons. The last

dense layer has 5 neurons since there are 5 classes to be predicted and SoftMax activation function. There are a total of 19,753,093 model parameters of which 14,714,688 are non-trainable parameters which were learned during the training of the VGG-16 model on the ImageNet dataset and these parameters are transferred and reused for the problem of histopathological image classification. The trainable models which correspond to the parameters in the dense layer are 5,038,405 which are learned through back-propagation.

B. ResNet-50

The ResNet-50 model is another complex deep neural network that consists of 49,311,621 parameters of which 25,723,909 are trainable parameters and 23,587,712 are non-trainable. The trainable parameters of this model compared to the previously discussed model are 5 times more and this means that the training time will be affected as well. Also, it can also be noted that the model requires more iterations to learn the features but since it is a deeper network it will learn more intricate features that are necessary in the medical imaging domain.

The 'Deep Residual Learning for Image Recognition' is a paper that first discussed the ResNet-50 model architecture and it is a 50 layers deep network [12]. It was also trained on the ImageNet dataset and is capable of classifying up to 1,000 classes. The default input is also (224,224,3) for this model. Similar to VGG-16 the ResNet-50 model contains several 3x3 convolution layers and several 2x2 max-pooling layers. The input images are also zero-padded and from (224,224,3) input shape the image is transformed to (230,230,3) input shape before it is fed to the first convolutional layer [12].

The output of the last convolutional layer is flattened and used as an input to the dense layers. Similar to the VGG-16 model there are three dense layers added to fine-tune the model. The first one having 256 neurons with ReLU activation function, second dense layer has 128 neurons with ReLU as well, and the last dense layer has 5 neurons and employs SoftMax activation function. These neurons were altered and for this model, the best recorded performance was achieved with the specified parameters.

Figure II: ResNet-50 Dense Layers

```
conv5_block3_out (Activation)  (None, 7, 7, 2048)   0          ['conv5_block3_add[0][0]']

flatten (Flatten)              (None, 100352)       0          ['conv5_block3_out[0][0]']

dense (Dense)                  (None, 256)          25690368   ['flatten[0][0]']

dense_1 (Dense)                (None, 128)          32896      ['dense[0][0]']

dense_2 (Dense)                (None, 5)            645        ['dense_1[0][0]']

==================================================================================================
Total params: 49,311,621
Trainable params: 25,723,909
Non-trainable params: 23,587,712
```

C. Inception V3

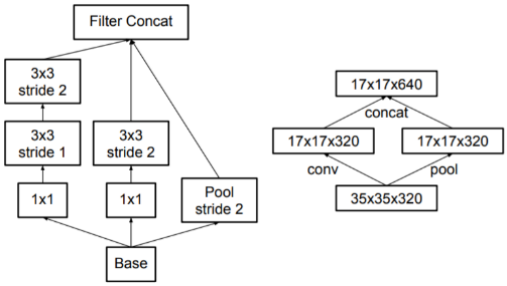
Another popularly used deep learning model for medical image classification is the Inception-V3 model and it is also trained on the ImageNet dataset and is able to perform classification task on more than 1,000 classes. The original paper that discusses this model was published under CVPR in 2016 [13]. It is entitled 'Rethinking the Inception Architecture for Computer Vision'. Szegedy et al. [13] proposed a 48 layers deep model which by default accepts input image size of 299 x 299.

The third figure illustrates the model architecture and the fourth figure expands upon the introduced inception layers in this model. The inception layer on the left is slightly more expensive in terms of computation compared to the proposed inception layer on the right side. These layers are a combination of 1x1, 3x3, and 5x5 convolutional layers. The purpose of this is to retain spatial information and be able to classify images under any case of distortion.

Figure III [13]: Inception V3 Model Architecture

type	patch size/stride or remarks	input size
conv	3×3/2	299×299×3
conv	3×3/1	149×149×32
conv padded	3×3/1	147×147×32
pool	3×3/2	147×147×64
conv	3×3/1	73×73×64
conv	3×3/2	71×71×80
conv	3×3/1	35×35×192
3×Inception	As in figure 5	35×35×288
5×Inception	As in figure 6	17×17×768
2×Inception	As in figure 7	8×8×1280
pool	8 × 8	8 × 8 × 2048
linear	logits	1 × 1 × 2048
softmax	classifier	1 × 1 × 1000

Figure IV [13]: Inception Layer Expanded



The output layer proposed in this research only consists of 1 dense layer with 5 neurons for classification and it employs SoftMax activation function. The total number of parameters in this model are 21,813,029 and only 34,432 are non-trainable. The other 21,778,597 are all trained and learned according to the problem.

D. Convolutional Neural Network (CNN)

The CNN model proposed in this paper is built from scratch and therefor there are 0 non-trainable parameters. This means that there is no transfer learning happening.

Figure V: CNN Model Architecture

```
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 224, 224, 3)]     0

conv2d_5 (Conv2D)            (None, 222, 222, 256)     7168

conv2d_6 (Conv2D)            (None, 220, 220, 200)     461000

conv2d_7 (Conv2D)            (None, 218, 218, 425)     765425

max_pooling2d_3 (MaxPooling  (None, 72, 72, 425)       0
2D)

conv2d_8 (Conv2D)            (None, 70, 70, 256)       979456

max_pooling2d_4 (MaxPooling  (None, 23, 23, 256)       0
2D)

conv2d_9 (Conv2D)            (None, 21, 21, 115)       265075

max_pooling2d_5 (MaxPooling  (None, 7, 7, 115)         0
2D)

flatten_1 (Flatten)          (None, 5635)              0

dense_5 (Dense)              (None, 115)               648140

dense_6 (Dense)              (None, 115)               13340

dense_7 (Dense)              (None, 256)               29696

dense_8 (Dense)              (None, 115)               29555

dense_9 (Dense)              (None, 5)                 580
```

The CNN model implemented is only 10 layers deep and has 3,199,435 parameters that are learned during training.

Equation I [14]: Categorical Cross Entropy Loss Function

$$\text{Loss} = -\sum_{i=1}^{\substack{\text{output}\\ \text{size}}} y_i \cdot \log \hat{y}_i$$

The categorical cross entropy loss function is used to calculate the error of the model and is also used to evaluate the performance of the training, test, and validation sets against every iteration.

Additional evaluation metrics observed are the F1 score, Recall, Accuracy, and Precision. The Adam optimizer is used and for all the cases excluding the CNN model the optimizer is given the following supporting parameters: learning rate= 0.001, beta 1 and beta 2 were changed to 0.9. The F1 score is given a threshold of 0.5. The CNN model makes use of the generic Adam optimizer and the parameters of this optimizer are not altered.

IV. Dataset and Pre-Processing

The dataset used for this research project is publicly available and can be obtained from Kaggle[15]. A total of 25,000 histopathological images are provided that belong to 5 different categories of lung and colon tissue types. The 5 classes are as follows: Lung benign tissue, Lung adenocarcinoma, Lung squamous cell carcinoma, Colon adenocarcinoma, and Colon benign tissue. These images were augmented using and Augmentor package to increase the samples to 25,000. The images are 768 x 768 presented in RGB format and for VGG-16, ResNet-50, and CNN the size is reduced to 224x224. For Inception-V3 the size is changed to 299x299. Very minimal pre-processing is done since the publishers of this dataset had already augmented it. The dataset was reduced to 11,250 images for training, 625 for testing, and 625 for validation. A total of 2,250 images from each category were used for training. This was done to reduce the complexity of training the models. After performing exploratory data analysis, it can be observed that the dataset is clean, consistent, and balanced. Training and test samples are similar and the noise is reduced in this dataset. From this analysis it can be predicted that the models proposed will produce good results.

V. Machine Requirement

All the experiments related to this research were performed on google colab pro which provides faster GPU, and longer run times. Tesla P100 is the GPU available with the colab pro subscription and was used to speed up the training of the models.

VI. Experiments and Results

This section goes over the results obtained from each experiments performed.

A. *VGG-16*

The overall accuracy achieved with the VGG-16 model is 99% and this is with the least number of trainable parameters compared to the other two pre-trained deep learning models. This model has 19,753,093 parameters and the other pre-trained models employ more than 20,000,000. The sixth figure show the accuracy and loss plots of the model after training it for 10 epochs. The graph shows that in the first epoch with batch size of 40 the model achieved an impressive training accuracy of 93% but it is unusual because the validation accuracy is greater. This is mostly due to the fact that our training, testing, and validation sets are inconsistent and the training set contains majority of the samples. Over time it can be observed that the model adjusts and learns the training data very well and the accuracy achieved in the tenth epoch is 99.70% and validation accuracy is 98.88%. After training and monitoring the validation sets, the model is ready for prediction with sklearns .predict function by taking the test set and in the seventh figure the performance of the model is observed. Overall, the model achieved an accuracy of 99% and the confusion matrix shows that there are very less false positives and negatives. The calculated error of the model is 0.014399.

Figure VI: VGG-16 Accuracy and Loss

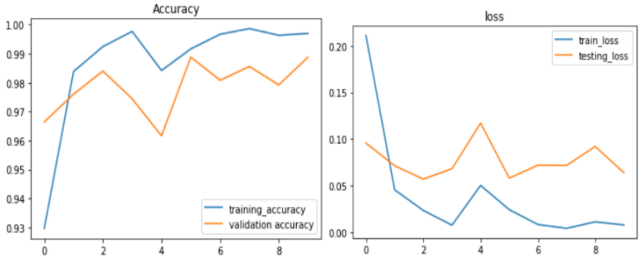
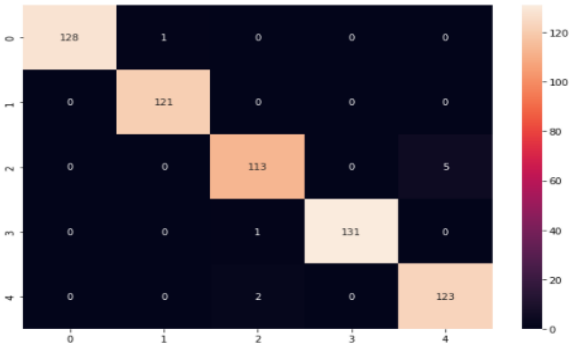


Figure VII: VGG-16 Performance on Test Data

	precision	recall	f1-score	support
0	1.00	0.99	1.00	129
1	0.99	1.00	1.00	121
2	0.97	0.96	0.97	118
3	1.00	0.99	1.00	132
4	0.96	0.98	0.97	125
accuracy			0.99	625
macro avg	0.99	0.99	0.99	625
weighted avg	0.99	0.99	0.99	625

Figure VIII: VGG-16 Confusion Matrix



B. *ResNet-50*

As mentioned previously ResNet-50 is a complicated model with the greatest number of trainable parameters from the chosen pre-trained models for this research. This model achieved a test accuracy of 85% and it was trained for 10 epochs with a learning rate of 0.001. Because it is a complicated deep learning framework its performance can be improved if the number of iterations is increased and increasing the neurons in the dense layers could also potentially improve the model's performance. It can be observed that from the accuracy and loss plots that the

training and validation results were consistent with less cases of overfitting.

Figure IX: ResNet-50 Accuracy and Loss

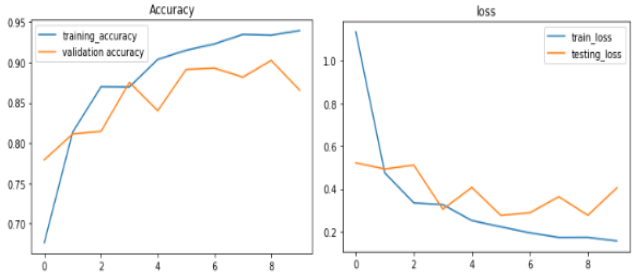
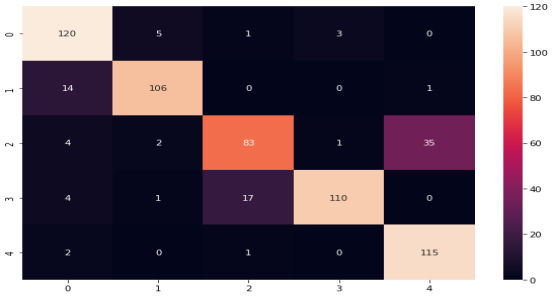


Figure X: ResNet-50 Performance on Test Data

	precision	recall	f1-score	support
0	0.83	0.93	0.88	129
1	0.93	0.88	0.90	121
2	0.81	0.66	0.73	125
3	0.96	0.83	0.89	132
4	0.76	0.97	0.86	118
accuracy			0.85	625
macro avg	0.86	0.86	0.85	625
weighted avg	0.86	0.85	0.85	625

Figure XI: ResNet-50 Confusion Matrix



If the confusion matrix is compared with the one obtained with VGG-16 it can be seen that more instances are misclassified by the model. For instance, 35 samples belonging to lung squamous cell carcinoma category were misclassified by the model as Colon benign tissues. Lung squamous cell carcinomas are not identified as benign or malignant. It is only found out after several tests are performed. If this model was being used to solve a real-world problem the pathologist would label the image as a lung benign tissue and close the case. It would put the life of the patient at risk. Therefore, the VGG-16 model outperforms ResNet50 at histopathological image classification problem.

C. Inception-V3

The test accuracy of this model is only 87% and from the plots it can be observed that there is overfitting between the training and test data. The model can be improved with a smaller learning rate as well as with a greater number of iterations. In the first epoch the training accuracy achieved is 85.69% but the validation accuracy is only 34.40%. This explains the plots shown below, and this is normal because the model is still learning but in the tenth epoch the model is still overfitting on the training data because the training accuracy achieved is 97.81% while the validation accuracy is only 84.48%. The ninth epoch achieves the best validation accuracy of 94.88%.

Figure XII: Inception V3 Loss and Accuracy

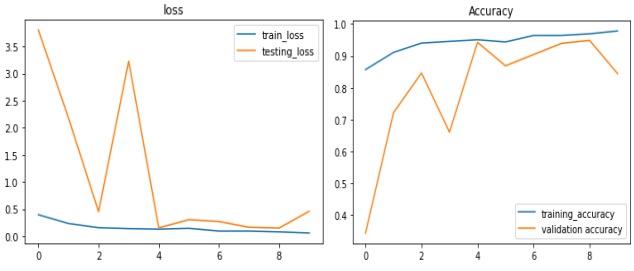
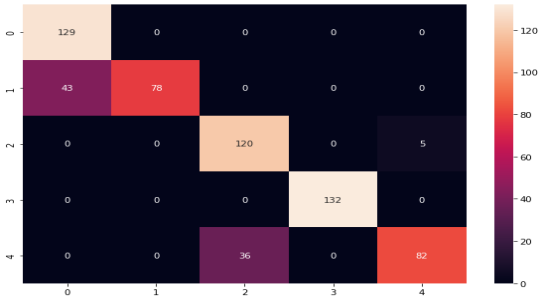


Figure XIII: Inception-V3 Performance on Test Data

	precision	recall	f1-score	support
0	0.75	1.00	0.86	129
1	1.00	0.64	0.78	121
2	0.77	0.96	0.85	125
3	1.00	1.00	1.00	132
4	0.94	0.69	0.80	118
accuracy			0.87	625
macro avg	0.89	0.86	0.86	625
weighted avg	0.89	0.87	0.86	625

Figure XIV: Confusion Matrix



The overall test accuracy achieved by this model is 87% and from 14th figure shown above it can be seen that the model is able to perfectly classify colon adenocarcinoma samples it was tested on and achieved an 100% precision, recall, and f1-score.

D. Convolutional Neural Network

The architecture of this model is described in the methodology section above and it is trained for 25 epochs with the default settings of the adam optimizer. This model took approximately 2 to 3 hours to train and learn features from the images belonging to the train set. This model outperformed the Inception-V3 and the ResNet-50 model, but it took 25 epochs and more time to train. The other models took approximately 15 to 20 minutes to train for 10 epochs with the ResNet-50 taking up the most time to train. The error of the model is 0.03200.

Figure XV: CNN Loss and Accuracy

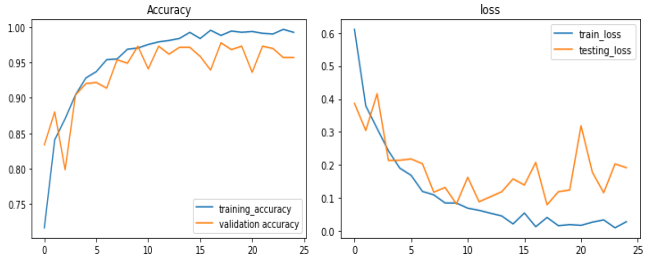
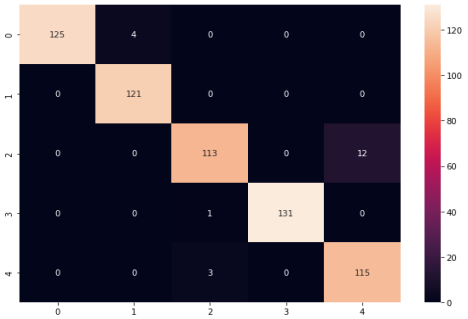


Figure XVI: CNN Performance on Test Data

	precision	recall	f1-score	support
0	1.00	0.97	0.98	129
1	0.97	1.00	0.98	121
2	0.97	0.90	0.93	125
3	1.00	0.99	1.00	132
4	0.91	0.97	0.94	118
accuracy			0.97	625
macro avg	0.97	0.97	0.97	625
weighted avg	0.97	0.97	0.97	625

Figure XVII: Confusion Matrix



VII. Future Work

Medical images are known for being noisy and distorted in most cases. The dataset used in this research has been pre-processed very well and it has been cleaned. All the images are consistent, and this is never the case with real medical images. This means that if this model were to be converted into an application to provide assistance to the pathologists it will definitely produce inaccurate results and perform poorly on distorted and noisy images. To solve this issue DCGAN can be used as proposed by Srivastava et al. [6]. It can be utilized to preprocess the histopathological images and add inconsistency and noise to the images to generate more data that is fake. Another preprocessing technique that can be used is to rotate the images in different angles because for real scenarios these input tissue slides are not consistent and don't usually have the same orientation. When it comes to feature extraction and classification the VGG-16 model, based on the experimental results produced in this paper, can be used and improved with more compute resources. With more compute resources such as GPUs, memory, and run time this model can be improved by adding more dense layers before the output layer and more neurons to learn more intricate features from the input medical images.

VIII. Conclusion

The pre-trained VGG-16 model after moderate fine-tuning produced the best test accuracy of 99%. These results suggest that RGB medical images such as histopathological images can be trained with the proposed fine-tuned model for only a few epochs and it will produce a good accuracy. Because the dataset used in this research is preprocessed to convert the input images to a more consistent set of image data, it is difficult to turn the proposed model into a usable application for real medical image classification that is highly distorted. It is suggested that additional preprocessing measures are taken to introduce noise and inconsistency to the dataset and this should dramatically improve the performance of the model for a real-world medical application. Early detection of cancer is the only preventative measure that can be taken to treat and cure patients. Deep learning models such as VGG-16 can be used to assist pathologists make faster and more accurate clinical decisions to provide better treatment plans.

References

[1] S. Lee, "Lung cancer statistics," *Canadian Cancer Society*. [Online]. Available:https://cancer.ca/en/cancer-information/cancer-types/lung/statistics.

[2] P. H. A. of Canada, "Government of Canada," *Canada.ca*, 17-Mar-2021. [Online]. Available: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/colorectal-cancer.html.

[3] A. Hekler, J. S. Utikal, A. H. Enk, W. Solass, M. Schmitt, J. Klode, D. Schadendorf, W. Sondermann, C. Franklin, F. Bestvater, M. J. Flaig, D. Krahl, C. von Kalle, S. Fröhling, and T. J. Brinker, "Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images," *European Journal of Cancer*, 18-Jul-2019. [Online].Available:https://www.sciencedirect.com/science/article/pii/S0959804919303806.

[4] J. Chang, J. Yu, T. Han, H. -j. Chang and E. Park, "A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer," 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 2017, pp. 1-4, doi: 10.1109/HealthCom.2017.8210843.

[5] T. Kaur and T. K. Gandhi, "Automated Brain Image Classification Based on VGG-16 and Transfer Learning," 2019 International Conference on Information Technology (ICIT), 2019, pp. 94-98, doi: 10.1109/ICIT48102.2019.00023.

[6] D. Srivastav, A. Bajpai and P. Srivastava, "Improved Classification for Pneumonia Detection using Transfer Learning with GAN based Synthetic Image Augmentation," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 433-437, doi: 10.1109/Confluence51648.2021.9377062.

[7] M. F. I. Soumik, A. Z. B. Aziz and M. A. Hossain, "Improved Transfer Learning Based Deep Learning Model For Breast Cancer Histopathological Image Classification," 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 2021, pp. 1-4, doi: 10.1109/ACMI53878.2021.9528263.

[8] A. S. B. Reddy and D. S. Juliet, "Transfer Learning with ResNet-50 for Malaria Cell-Image Classification," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0945-0949, doi: 10.1109/ICCSP.2019.8697909.

[9] M. Fradi, M. Afif, E. -H. Zahzeh, K. Bouallegue and M. Machhout, "Transfer-Deep Learning Application for Ultrasonic Computed Tomographic Image Classification," 2020 International Conference on Control, Automation and Diagnosis (ICCAD), 2020, pp. 1-6, doi: 10.1109/ICCAD49821.2020.9260569.

[10] K. Team, "Keras Documentation: Keras applications," *Keras*. [Online]. Available: https://keras.io/api/applications/.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv.org*, 10-Apr-2015. [Online]. Available: https://arxiv.org/abs/1409.1556.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv.org*, 10-Dec-2015. [Online]. Available: https://arxiv.org/abs/1512.03385.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *arXiv.org*, 11-Dec-2015. [Online]. Available: https://arxiv.org/abs/1512.00567.

[14] “Loss functions: Build an AI model: Peltarion platform,” *Peltarion*. [Online]. Available: https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions.

[15] Larxel, “Lung and colon cancer histopathological images,” *Kaggle*, 14-Apr-2020. [Online]. Available: https://www.kaggle.com/andrewmvd/lung-and-colon-cancer-histopathological-images.