Using Deep Learning to Perform Multi-Class Classification on the Lung and Colon Cancer Histopathological Image Dataset (LC25000)

Deep Learning: 880663-M-6 Assignment

Report by: Erion Mediu, 2066763, E.mediu@tilburguniversity.edu

March 10th 2024

1 Problem Definition

The early detection and accurate classification of lung and colon cancers are critical challenges in medical diagnostics, directly impacting patient treatment outcomes. This project aims to harness deep learning neural networks, specifically, Convolutional Neural Networks (CNNs), to develop a precise model for classifying lung and colon cancer images from the LC25000 dataset. The primary objective is to fine-tune a deep learning model that surpasses random chance in classifying various histopathological images of lung and colon cancers, thus providing valuable assistance to professionals in the field. By leveraging computer vision techniques, the project aims to enhance diagnostic processes but also contributes to the ongoing research and frameworks in medical image analysis using deep learning technologies. The report progresses by dissecting the data through EDA, presenting a baseline model for benchmarking, and introducing an enhanced model alongside a transfer learning approach. It concludes with a thorough discussion of potential improvements and future research.

2 Exploratory Data Analysis

The main dataset Borkowski et al. (2023), contains 25,000 images with 5 classes of 5,000 color images each. This balanced nature of the data is consistent also after dropping 1280 duplicates as showcased in figure 1. The aforementioned duplicates were detected using a specialized script documented in the attached notebook of the report. Nonetheless, it remains important to mention that the data set consists of a data file and a label file respectively used as features and targets in the model training, and comes in a .npy format, containing image arrays or image tensors that can be used to regenerate the pictures. Thus after converting the 5 classes of target variables to One-Hot-Encoded format, 15 samples from the dataset were randomly selected. For each selected sample, the image and its corresponding label can be seen in figure 2. It is clear from this dissection that the dataset presents a diverse and high-quality collection of images, showcasing favorable conditions for model development.

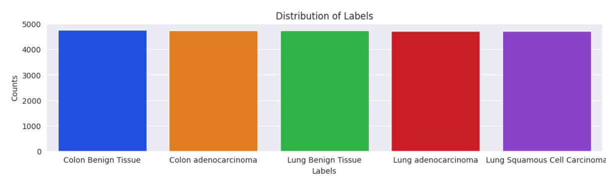


Figure 1: Bar plot to visualize the class label distribution of the dataset.

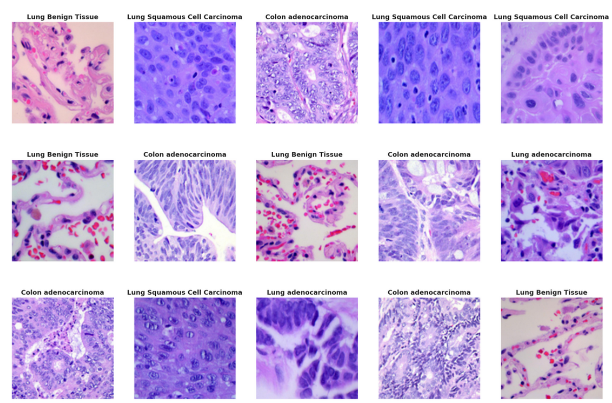


Figure 2: Diagnostic Diversity: A Visual Compilation of Lung and Colon Cancer Histopathologies Present in the Dataset

3 Results of the Baseline Model

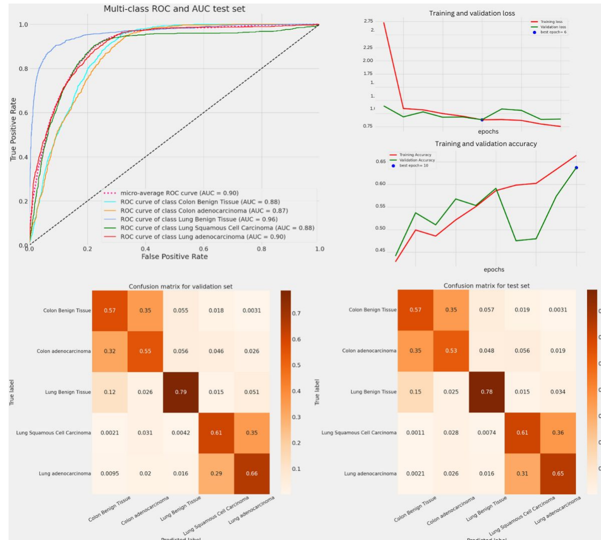


Figure 3: Baseline model evaluation metrics

To align with the predefined goals of the project, the baseline model serves as the foundational benchmark for comparison, with its configuration and parameters pre-defined, the results of such baseline model together with the framework showcased in figure 3 also dictate the methods that will be used for further evaluations. Using categorical cross-entropy as the loss function for the base model and Adam optimizer, the model is fit onto the data utilizing 10 epochs with a batch size of 32. The baseline model's optimal loss is seen at epoch 6 while peak accuracy is reached at epoch 10, this immediately hints at inconsistent learning dynamics.

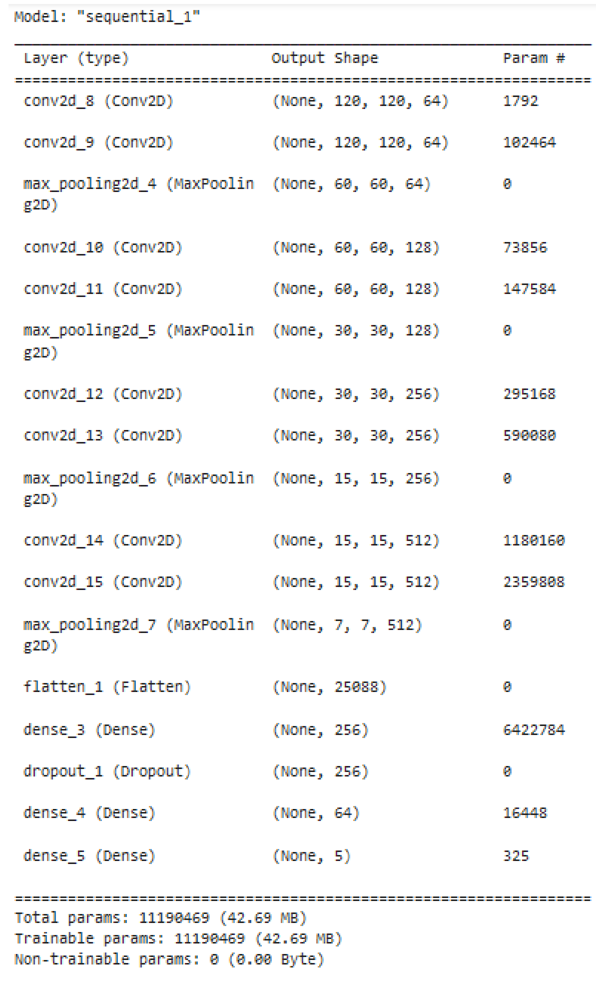
Thus potential for model refinements in relation to model depth and learning rate which we explore in the following sections. Meanwhile, the ROC curve and AUC scores, which are critical for assessing model performance across classes exceed the baseline value of 0.5. Indicating that the model's predictions are derived from learned patterns rather than random guessing. In such use cases high level of discrimination is crucial, yet the confusion matrix in conjunction with the aforementioned ROC-AUC curve, highlighted the inherited inability of the baseline model to distinguish classes apart from the Lung Benign Tissue class with high efficiency. This hints at the implications that the chosen architecture may have on the model. To conclude, the test and validation set performance metrics comparison is presented in table 1. While, the trends observed in the confusion matrix and class-specific performance highlight opportunities for improving the model's ability to generalize across classes more evenly.

4 Enhanced Model and Results

To enhance the model's performance in terms of accuracy and loss reduction, we focused on architectural improvements without modifying the data preprocessing techniques. Our approach to hyperparameter tuning aimed to utilize the distinct characteristics of the dataset across six key parameters. Firstly, we considered the network's depth. According to Chen and Tsou (2022), while increasing the network's depth can be advantageous, overly deep networks might suffer from reduced accuracy. Secondly, we adjusted the kernel size in the second convolutional layer, increasing it from the standard (3,3) to (5,5). This change allows the model to capture more high-level features early on at the expense of increasing the number of parameters, which, theoretically, could heighten the complexity of the network.

To mitigate this, the improved model combines this with regularization techniques such as dropout, and L2 regularization on later layers. Thus, betting on the idea that larger kernels can enhance the model's ability to learn complex patterns without necessarily memorizing the noise in the training data, Tomen and Gemert (2021). As for the optimizer as a hyperparameter, the model employed the Adamax optimizer, in line with Kandel and Castelli (2020) which also looks at optimizer performance using CNNs on Histopathology Images. Kandel and Castelli (2020) argues that "Globally, the best result on the considered dataset was achieved by Adamax optimizer and DenseNet network". Within the same study, the optimal learning rate for the model was 10^5. To conclude the list of changes, an EarlyStopping callback was utilized to monitor validation loss, ceasing training when no improvements were observed for two consecutive epochs, a precautionary measure used to prevent the network from overfitting Kandel and Castelli (2020). The performance of the improved model compared to the baseline model can be assessed through several angles. From the ROC curve and AUC score, it's evident that the fine-tuned model achieves near-perfect scores, which indicate an excellent classification performance across all classes. Compared to the baseline, in terms of the confusion matrices[a], the enhanced model resolves the previously discussed issue with far fewer misclassifications. In the enhanced model, we see a steady decline in loss and an increase in accuracy, which indicates better learning dynamics. The use of the Adamax optimizer and the chosen learning rate likely contributed to this.

[a]The values in the confusion matrix are normalized to lie between 0 and 1, representing the proportion of each true class per predicted class.



Model: "sequential_1"

Layer (type)	Output Shape	Param #
conv2d_8 (Conv2D)	(None, 120, 120, 64)	1792
conv2d_9 (Conv2D)	(None, 120, 120, 64)	102464
max_pooling2d_4 (MaxPooling2D)	(None, 60, 60, 64)	0
conv2d_10 (Conv2D)	(None, 60, 60, 128)	73856
conv2d_11 (Conv2D)	(None, 60, 60, 128)	147584
max_pooling2d_5 (MaxPooling2D)	(None, 30, 30, 128)	0
conv2d_12 (Conv2D)	(None, 30, 30, 256)	295168
conv2d_13 (Conv2D)	(None, 30, 30, 256)	590080
max_pooling2d_6 (MaxPooling2D)	(None, 15, 15, 256)	0
conv2d_14 (Conv2D)	(None, 15, 15, 512)	1180160
conv2d_15 (Conv2D)	(None, 15, 15, 512)	2359808
max_pooling2d_7 (MaxPooling2D)	(None, 7, 7, 512)	0
flatten_1 (Flatten)	(None, 25088)	0
dense_3 (Dense)	(None, 256)	6422784
dropout_1 (Dropout)	(None, 256)	0
dense_4 (Dense)	(None, 64)	16448
dense_5 (Dense)	(None, 5)	325

Total params: 11190469 (42.69 MB)
Trainable params: 11190469 (42.69 MB)
Non-trainable params: 0 (0.00 Byte)

Figure 4: Enhanced model summary

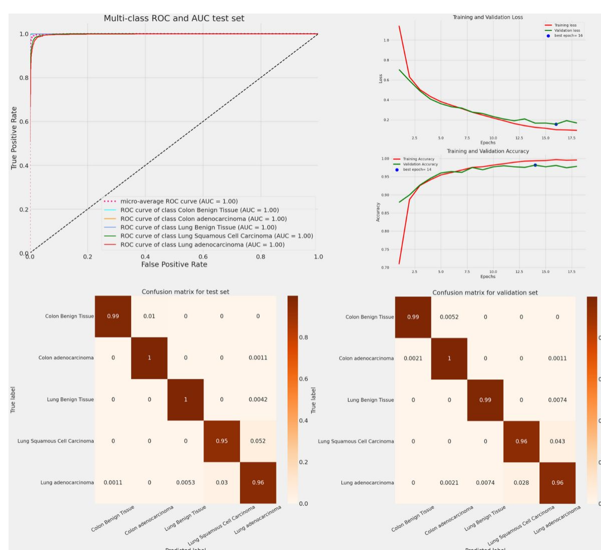


Figure 5: Enhanced model evaluation metrics

The increase in model depth and parameters has enhanced the model's capability, leading to a leap in performance, as documented in table 1. Simultaneously, the integration of regularization techniques and dropout has effectively lowered the risk of overfitting, while maintaining a balance between model complexity and generalization ability. Finally, the implementation of an EarlyStopping mechanism, in conjunction with the Adamax optimizer and the chosen learning rate, showcases the approach toward optimizing loss minimization while maintaining computational efficiency. This selection was concluded after many intermediate model iterations and proved to elevate the model's accuracy while also making it practical for deployment.

5 Transfer learning model

The key advantage of transfer learning lies in the ability to utilize the knowledge encoded in the weights of models trained on large and diverse datasets. In our approach, we employ the ResNet50 architecture, a deep convolutional neural network that has been pre-trained on the vast ImageNet dataset. The transfer of weights to a new task enables the model to benefit from the pre-learned features, such as edges, textures, and patterns, which are broadly applicable to a wide array of image recognition tasks. As depicted in Figure 6, we append three dense layers to ResNet50, introduced a 20 percent dropout rate, and an EarlyStopping callback with a patience of three to prevent overtraining while considering computational efficiency. This strategy differs from the enhanced model. The decision to increase patience was made because the transfer learning model showed promising performance early on while also maintaining a quick runtime. Therefore, it seemed beneficial to allow the model more time to train, potentially capturing additional accuracy gains. This contrasts with the approach taken in the enhanced model, where computational efficiency was prioritized over the pursuit of marginal performance improvements.

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 resnet50 (Functional)       (None, 4, 4, 2048)        23587712

 global_average_pooling2d (  (None, 2048)              0
 GlobalAveragePooling2D)

 dense_6 (Dense)             (None, 256)               524544

 dropout_2 (Dropout)         (None, 256)               0

 dense_7 (Dense)             (None, 64)                16448

 dense_8 (Dense)             (None, 5)                 325

=================================================================
Total params: 24129029 (92.04 MB)
Trainable params: 541317 (2.06 MB)
Non-trainable params: 23587712 (89.98 MB)
_________________________________________________________________
```

Figure 6: Transfer learning model summary

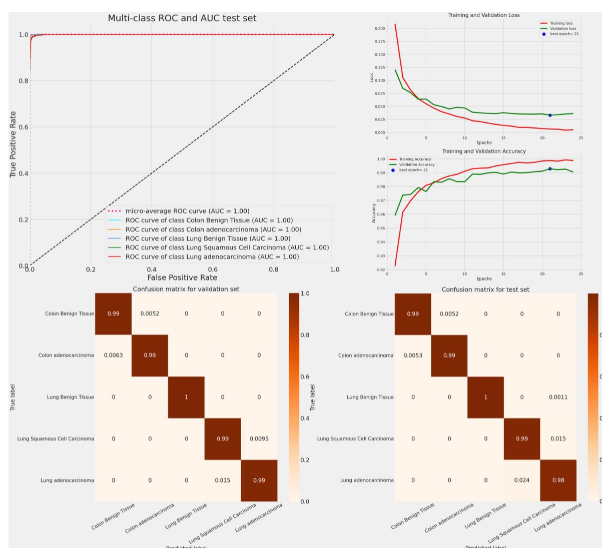


Figure 7: Transfer model evaluation metrics

The enhanced model, which incorporated changes like increased depth, kernel size, and added regularization, showed notable improvements over the baseline. The adjustments aimed to enhance the model's capacity to capture more complex patterns without overfitting. While the performance is impressive, especially in the context of learning dynamics (observing a steady decline in loss and increasing accuracy), it still falls short of the transfer learning model's achievements. As can be seen in figure 6, the transfer learning model has achieved near-perfect ROC and AUC scores across all classes in the multi-class classification task. It outperforms both the baseline and the enhanced models in terms of accuracy and the ability to generalize, as seen in the confusion matrices, which display fewer misclassifications. The enhanced capabilities compared to the baseline and enhanced models, as we argue in the following discussion, can be attributed to the rich feature set that ResNet50 brings to the task. These pre-learned features provide a starting point that is already very well-suited for image recognition tasks. Furthermore, the added layers on top of the ResNet50 base enable the model to adapt these features to the specific dataset and task, resulting in improved discrimination also between difficult-to-distinguish classes.

	Validation Set Performance			Test Set Performance		
	(1)	(2)	(3)	(1)	(2)	(3)
Accuracy	0.6372	0.9808	0.9928	0.6262	0.9793	0.9898
Precision:	0.6415	0.9808	0.9928	0.6314	0.9794	0.9898
Recall	0.6373	0.9807	0.9928	0.6238	0.9792	0.9898
F1-Score	0.6388	0.9807	0.9928	0.6282	0.9792	0.9898

Table 1: Comparative performance metrics of the three evaluated models, where models (1), (2), and (3) correspond to the Baseline, Enhanced, and Transfer Learning models, respectively. Values are presented as decimals, rounded to four digits for clarity.

6 Discussion

Building upon the exploration and results from the baseline, enhanced, and transfer learning models, this project came across a wide spectrum of network architectures and factors that impact classification. While existing literature highlights several pathways for enhancing algorithmic performance, for this project the main papers upon which hyperparameter tuning of the enhanced model was based are a few. Namely Chen and Tsou (2022) who argues that increased depth is crucial yet needs to be followed by regularisation so that it can be beneficial. This accuracy decreases with the increase of layers coming from vanishing gradients and degradation problems. He, Zhang, Ren, and Sun (2015) illustrates that a 20-layer network achieves lower training errors compared to a 56-layer network, suggesting that the latter struggles with generalization on new data, rendering it less efficient. While Kandel and Castelli (2020) guided the selection of the learning rate and optimizers. Early stopping, dropout rate, and kernel size modification were consequences of intermediate models and an ongoing attempt to increase accuracy and minimize loss while keeping on with computational constraints.

Nonetheless, it became clear in this project that the exploration and implementation of CNNs for the classification of lung and colon cancer images from the LC25000 dataset reflects a microcosm of the broader challenges and innovations shaping the field of computer vision. As highlighted in recent literature M. Tan and Le (2019), CNNs have been pivotal in advancing vision-based applications, yet they face significant hurdles, such as the need for extensive training data and high computational demands. Transfer learning has been proposed as a strategic response to the limitations imposed by data scarcity and computational resources. This technique, rooted in the principle of leveraging pre-acquired knowledge from one domain to enhance performance in another, has increasingly become the focus of contemporary research. The literature suggests that transfer learning not only addresses the issue of data scarcity by utilizing models pre-trained on large, diverse datasets but also offers a path to more efficient computation, C. Tan et al. (2018). Such transfer learning models provide a rich set of features that can be repurposed for tasks with significantly smaller datasets, thereby reducing the computational burden of training from scratch. Moreover, regularization techniques employed in conjunction with transfer learning, such as L2 regularization and dropout, play a crucial role in the mitigation of overfitting. This ensures that the model remains generalizable to new, unseen data, a critical aspect of deploying models in real-world scenarios C. Tan et al. (2018).

Even though in this project within the enhanced model the aforementioned problem of vanishing gradients doesn't seem to be present since a model suffering from vanishing gradients would typically show very slow or no improvement, especially in the early layers. The choice of the transfer learning model was shaped by this potential occurrence. There is a large selection of pre-built models in the literature but a deep residual network, often referred to as ResNet, is a type of convolutional neural network (CNN) that was introduced to solve the problem of vanishing gradients which would make it difficult for the model to learn and update its parameters effectively, He et al. (2015). Additionally, deeper networks as briefly touched upon before can also suffer from a degradation problem, where adding more layers leads to higher training error, contrary to what one might logically expect. In summary, transfer learning literature emphasizes its role in overcoming data limitations and computational challenges, suggesting a future where deep learning models are more accessible and efficient across various fields. The use of pre-trained models, alongside fine-tuning and regularization, is a robust strategy for exploiting CNNs' capabilities, particularly where data scarcity and computational resources are concerns. Retrospectively the insights of the literature display objectively the reasons behind the increased performance of the enhanced model from the baseline and at the same time the limitations. This logically leads to the belief that future projects and endeavors may benefit from allocating the resources to optimizing such pre-trained models as opposed to building models from scratch.

References

Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A., & Mastorides, S. M. (2023). Lc25000 lung and colon histopathological image dataset. Retrieved from https://github.com/tampapath/lung_colon_image_set/

Chen, F., & Tsou, J. Y. (2022). Assessing the effects of convolutional neural network architectural factors on model performance for remote sensing image classification: An in-depth investigation. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102865. Retrieved from https://www.sciencedirect.com/science/article/pii/S156984322200067X doi: https://doi.org/10.1016/j.jag.2022.102865

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep residual learning for image recognition.*

Kandel, I., & Castelli, M. (2020, 09). Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images. *Journal of Imaging*, *6*. doi: 10.3390/jimaging6090092

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. *CoRR*, *abs/1808.01974*. Retrieved from http://arxiv.org/abs/1808.01974

Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, *abs/1905.11946*. Retrieved from http://arxiv.org/abs/1905.11946

Tomen, N., & Gemert, J. (2021, 01). Spectral leakage and rethinking the kernel size in cnns. , 5138-5147.