

# Deep Learning with Histology Images

## Cancerous Cell and Cell Type identification

For this project, we utilised convolutional neural networks (CNNs) to perform deep learning on the set of histology patch images. Various CNN architectures were used as a basis to tune and optimise the performance of models capable of classifying cancerous cells and cell types.

## Cancer Cell Classification

### Selecting a base model

Two CNN architectures were identified as suitable for this classification task and were tuned using different hyperparameters and training parameters.

The spatially-constrained model, which was found to be the most effective model for classification of whether cells were cancerous or not, was proposed originally by the authors of the paper from which the dataset was sourced (Sirinukunwattana, Raza, Tsang, Snead, Cree & Rajpoot, 2016) for the purposes of identifying cell types.

The Lenet CNN architecture is well-suited to classification of small images and so was also used to test our model. It is a popular model (Géron, 2019) having been used and tested widely, meaning that it is well-suited to serve as a baseline to compare the other model to while also being suitable for images of a similar size and type, originally having been designed to effectively classify images from the MNIST dataset of handwritten digits (Géron, 2019).

### Model tuning

A number of different hyperparameters were adjusted and tested in an attempt to develop a model with maximum f1-score and area under the ROC curve (AUC). These included adding and adjusting the strength of a dropout layer for regularisation purposes, adjusting the learning rate throughout model training and using data augmentation.

Zero-padding was used to ensure images were of sufficient size to be processed as inputs by the LeNet architecture.

### Evaluation

As stated earlier, F1-score was emphasised for evaluating model performance because this takes into account the number of false-negatives and false-positives (via precision and recall). Minimising the number of false negatives was regarded as important because the consequences of failing to diagnose cancer are serious, with severe illness and death likely outcomes. Meanwhile, false-positives can cause significant distress to patients, lead to unnecessary further invasive procedures such as surgery or consume valuable resources for treating actual cancer sufferers. The F1-score takes both into account but also balances these concerns. Area under curve was also used as a guide when developing the model as it is a good general indicator of the usefulness of a model in classification problems.

Accuracy, on the other hand, was not considered as important because it tends to severely exaggerate the performance of the model when there is a class imbalance. Furthermore, it does not suit this problem well due to the real-world implications of this problem; the measure of accuracy weights the consequences of true and false positives as well as true and false negatives equally.

Compared to comparable deep learning cancer cell classification models in research papers, the models that we tested and tuned performed relatively well, suggesting that our model development was on the right track.

	<b>LeNet base</b>	<b>Spatially-Constrained base</b>	<b>LeNet Tuned</b> (Adjusted learning rate)	<b>Spatially-Constrained tuned</b> (Dropout layer and adjusted learning rate)	<b>Classification of breast cancer histology images using Convolutional Neural Networks</b>  (Araújo et al., 2017)	<b>Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases -Alexnet CNN</b>  (Janowczyk & Madabhushi, 2016)
Maximum Accuracy	89.35%	89.84%	88.76%	89.50%	77.6%	84.68%
Maximum F1-score	0.843750	0.849635	0.832075	0.841871		0.7648

Note that the value of these comparisons should not be over-emphasised as datasets vary in nature and complexity, but it can help to indicate whether a model is within a reasonable and expected range of performance, as well as the relative usefulness of said model.

It may be the case that, as the patch size is capped at 27x27 pixels, little more information can be gleaned from the image data than is already being extracted. This must be considered in balance with the possibility that creating even smaller patch sizes could lead to stronger results as less pixels without significance are used as input variables (Janowczyk & Madabhushi, 2016). A far more exhaustive analysis of potential models would be required to determine this however.

## Ultimate Judgement on Cancer Cell Classification

Because it demonstrated the highest maximum F1-score (unlike its tuned counterpart), it is our judgement that the spatially-constrained model with early stopping during training should be used for this classification task. Early stopping will prevent further training when the loss function begins to diverge. Reducing the learning rate on the other hand had the effect of decreasing the variability of the loss function and lead to more reliably high performance for classification across the validation and test datasets. When a more reliable level of classification performance is desired, it may be well advised to use the tuned model which prevents the loss function from diverging during training.

# Cell-Type Identification

## Transfer Learning and Model Selection

The best tuned model from the cancerous cell identification model was used and tuned further for the purpose of classifying cell types. This was used as a base model. Different optimizers were experimented with, as well as activation functions, and the best one (in terms of the selected performance metric) was selected.

Two other models (Janowczyk & Madabhushi, 2016) (Hamad, Ersoy & Bunyak, 2018) were then used from further research, tuned similarly and all were evaluated to choose the best model.

## Evaluation

From the EDA, it was found that epithelial cells are the only ones that are cancerous - thus making them the most important to classify because we want to know if patients have cancer. Therefore, for primarily classifying epithelial cells, the F1-Score for that class suffices.

When looking at the classification of other cell types, the macro-average F1-Score was used as a performance metric, as it equally weights the F1-Score of all cell types, regardless of the class imbalance.

	Custom-Tuned Model	AlexNet	FCRN+CNN
Macro-Avg F1-Score	0.76	0.72	0.66
Weighted F1-Score	0.79	0.77	0.72

From the results, we can see that the custom-tuned model outperforms the other 2 models taken into consideration

## Ultimate Judgement

### Enhancing the cell-type identification by combining the cancer cell identification model

As noted by Janowczyk and Madabhushi "regions of cancer are typically manifested in the epithelium" (Pg. 38, 2016). This is reflected strongly in the main dataset provided, insofar as all of the patch samples which have been manually identified as cancerous are also of epithelial type. Therefore, as a part of our ultimate judgement, we used the spatially-constrained model in conjunction with the custom-tuned model to make a decision on the classification of cell type. This was done by using both models to make a prediction on data, and where they both disagree as to whether a cell is cancerous (epithelial), we use the model that classifies an image as cancerous (or epithelial). This is because we would rather have more false positives, and people getting tested thoroughly and diagnosed than false negatives.

## References

Géron, A., 2019. Hands-on machine learning with Scikit-Learn and TensorFlow. 2nd ed. Sebastopol: O'Reilly, p.463.

Sirinukunwattana, K., Raza, S., Tsang, Y., Snead, D., Cree, I. and Rajpoot, N., 2016. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. *IEEE Transactions on Medical Imaging*, 35(5), pp.1196-1206.

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A. and Campilho, A., 2017. Classification of breast cancer histology images using Convolutional Neural Networks. *PLOS ONE*, 12(6), p.e0177544.

Janowczyk, A. and Madabhushi, A., 2016. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7(1), pp.29-47.

Hamad, A., Ersoy, I., and Bunyak, F., 2018 Improving Nuclei Classification Performance in H&E Stained Tissue Images Using Fully Convolutional Regression Network and Convolutional Neural Network. *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1-6, doi: 10.1109/AIPR.2018.8707397.