

Neural Networks 2024–SI  
**Histopathology of cancer detection**

Christian Chávez

May 7, 2024

**Abstract**

The project focuses on creating a binary image classification model for cancer diagnosis, specifically targeting metastases in histopathology images. It uses convolutional neural networks (CNNs) and transfer learning to train the model, optimizing hyperparameters. The model's performance is assessed, providing insights for future advancements in cancer detection using deep learning techniques. The study evaluates validation metrics, visualization, and interpretation of model predictions, highlighting relevant regions in cancer histopathology.

**Contents**

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Problem statement</b>	<b>2</b>
2.1	Data description . . . . .	2
<b>3</b>	<b>Exploratory data analysis</b>	<b>3</b>
3.1	Preprocesing . . . . .	3
3.2	Augmentation . . . . .	3
<b>4</b>	<b>Methodology</b>	<b>4</b>
4.1	Model architecture . . . . .	5
4.2	Hyperparameters . . . . .	6
<b>5</b>	<b>Results and discussion</b>	<b>6</b>
5.1	Loss plot . . . . .	7
5.2	Confusion matrix . . . . .	7
5.3	Enhancement of the model . . . . .	8
5.3.1	ROC and AUC . . . . .	8
5.4	Discussion . . . . .	9
<b>6</b>	<b>Conclusion</b>	<b>9</b>
<b>A</b>	<b>Code</b>	<b>10</b>
<b>References</b>		<b>10</b>

# 1 Introduction

Histopathology image analysis is crucial in modern healthcare, especially in cancer diagnosis and prognosis. With the advent of digital pathology and vast amounts of histopathological image data, there is growing interest in leveraging machine learning techniques, particularly neural networks, for automated cancer detection. Detecting metastases in histopathology images presents a significant challenge due to the subtle nature of metastatic cells and the need for precise localization within tissue samples. Convolutional neural networks (CNNs) have emerged as powerful tools for image analysis tasks, with transfer learning being particularly effective for medical image analysis tasks. Researchers can leverage the knowledge encoded in the pretrained networks to achieve superior performance on targeted tasks. This study presents a baseline model for binary image classification, focusing on the detection of metastases within histopathology images. The primary objective is to establish a foundational framework for automated metastasis detection in histopathology images using state-of-the-art deep learning techniques. Through experimentation and analysis, the study aims to evaluate the efficacy of the proposed methodology and provide valuable insights for further advancements in computer-aided diagnosis systems for cancer detection.

## 2 Problem statement

This project aims to classify digital histopathology images to detect metastases, secondary cancerous growths from primary tumors. Metastases can appear as single cells in large tissue areas, making it challenging to accurately identify even the smallest clusters. The goal is to build a model capable of identifying even the smallest metastatic clusters, aiding in early cancer detection and diagnosis, by differentiate between metastase-containing and non-metastase images.

### 2.1 Data description

See [Table 2.1](#)

Table 2.1: Data description

Description	Value
Data Size	220025
Data Dimension	2
Data Structure	[‘id’, ‘label’]
Number of training images	220k
Number of evaluation images	57k
Image dimensions	96 x 96
Channels	3
Bits per channel	8
Data type	Unsigned char
Compression	Jpeg

### 3 Exploratory data analysis

See [Figure 3.1](#) for exploration of random samples for comparison.

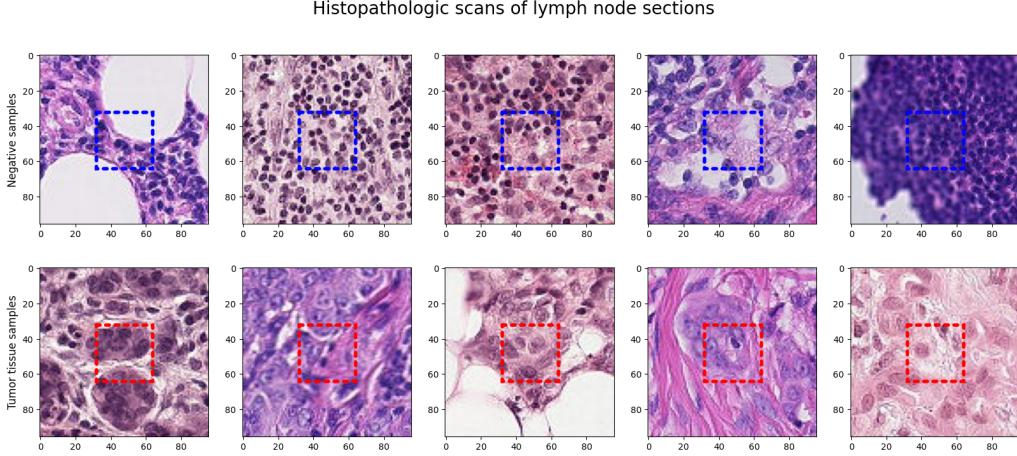


Figure 3.1

After an exploration, we found there is presence of outliers ([Figure 3.2](#)), as totally black or white images. There was 1 extremely dark image and 6 extremely bright images. These images will be removed.

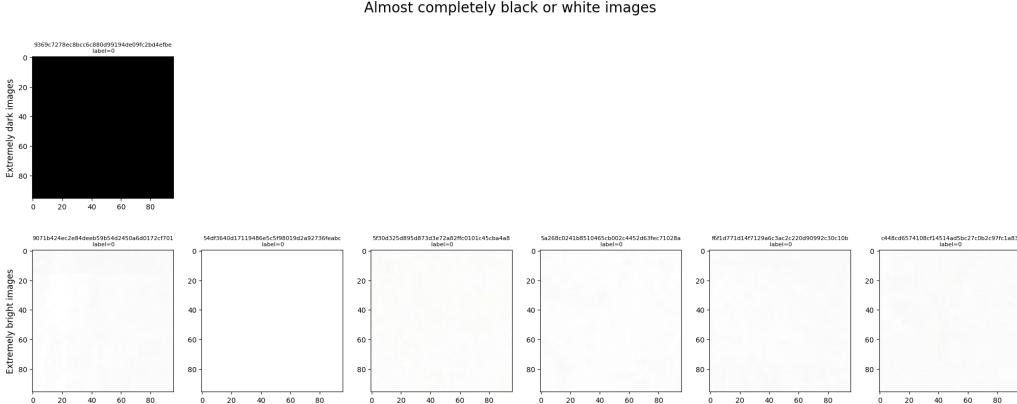


Figure 3.2

#### 3.1 Preprocessing

We know that the label of the image is influenced only by the center region (32 x 32px) so it would make sense to crop our data to that region only. However, some useful information about the surroundings could be lost if we crop too close. This hypothesis could be confirmed by training models with varying crop sizes. [Figure 3.3](#) shows the cropped images of [Figure 3.1](#).

#### 3.2 Augmentation

Addressing overfitting in machine learning models requires a multifaceted approach, and one effective strategy involves employing image augmentation techniques to introduce variability

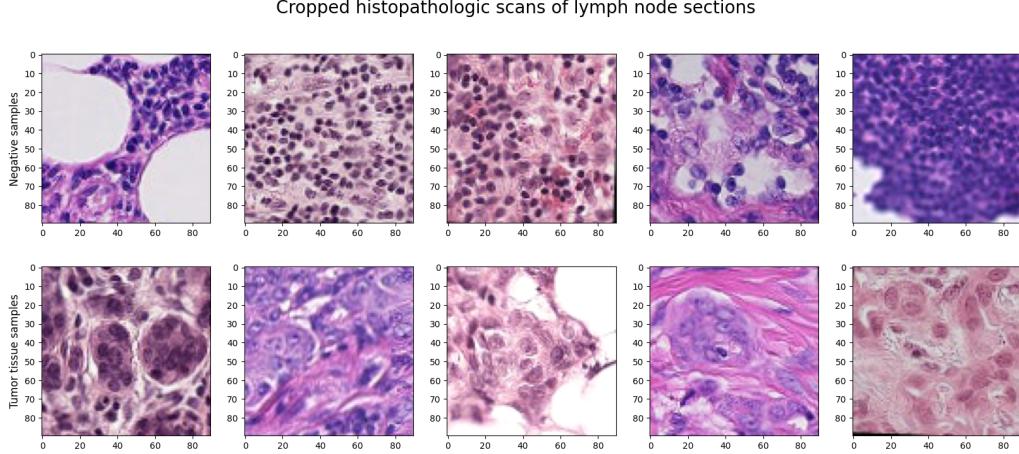


Figure 3.3

into the training data. By incorporating transformations such as random rotation, cropping, flipping, and adjustments to lighting conditions, augmented images enrich the dataset and help prevent the model from memorizing specific patterns. Integrating these augmentations into the image loading process facilitates seamless integration into the training pipeline, while test time augmentation (TTA) further enhances model performance by applying the same transformations during inference, thereby leveraging ensemble learning principles. Leveraging the OpenCV library for efficient image processing operations enables effective implementation of augmentation, contributing to improved model generalization and mitigating the risk of overfitting. See Figure 3.4.

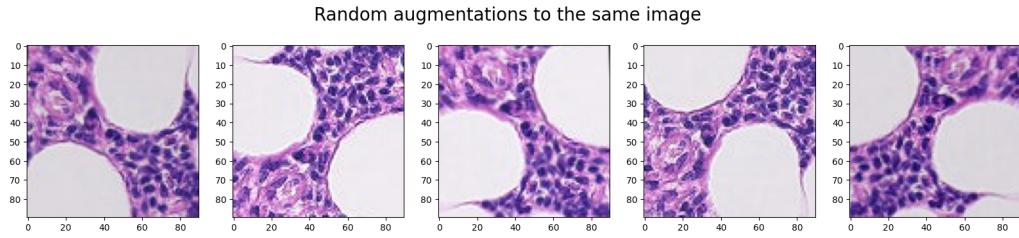


Figure 3.4

Based on the preliminary analysis, we'll utilize a pre-trained convolutional neural network (convnet) to conduct image classification.

## 4 Methodology

The study uses a dataset of histopathology images from the PCam dataset, containing 220k training and 57k evaluation images. The images are preprocessed to focus on relevant features and remove any corrupted data. Transfer learning is used for model architecture, with pretrained convolutional neural networks (CNNs) like ResNet and DenseNet. The training adopts a 1cycle policy, optimizing learning rates and weight decay [7]. The model is then trained with lower learning rates for further fine-tuning. Performance is evaluated using validation metrics like accuracy, precision, recall, and F1-score. Random samples, predictions, and examples of

incorrectly labeled images are analyzed to understand model behavior and identify improvement areas. Validation metrics like ROC curve and AUC are calculated to quantify model performance.

## 4.1 Model architecture

In the realm of image recognition, various deep learning structures have demonstrated remarkable performance. Convolutional Neural Networks (CNNs) form the basis, with trailblazers such as AlexNet, VGGNet, and GoogLeNet leading the charge. ResNet and DenseNet stand out for their exceptional accuracy, whereas MobileNet and EfficientNet prioritize computational efficiency. The existent literature points out to the fact that densenet is one of the most effective architectures when it comes to image recognition and image classification, as noted in [8, 9, 5, 2].

We will use a pretrained convnet model and transfer learning to adjust the weights to our data. Going for a deeper model architecture will start overfitting faster. The code of the implementation can be found at [https://colab.research.google.com/drive/1L43tKjEDZE6ZBf6\\_v\\_eW-w0btKpKAh8m?usp=sharing](https://colab.research.google.com/drive/1L43tKjEDZE6ZBf6_v_eW-w0btKpKAh8m?usp=sharing).

The paper by [4] proposes an architecture (densenet169) with a simple connectivity pattern, connecting all layers directly to each other to maximize information flow. Each layer receives inputs from all preceding layers and passes its feature maps to all subsequent layers, preserving the feed-forward nature of the network. In [Table 4.1](#) we can see the detailed architecture of densenet169.

Table 4.1: Architecture of DenseNet-169. Taken from [1].

Layers	Output Size	DenseNet 169	
Convolution	$112 \times 112$	$7 \times 7$ conv, stride 2	
Pooling	$56 \times 56$	$3 \times 3$ max pool, stride 2	
Dense Block (1)	$56 \times 56$	$1 \times 1$ conv	$\times 6$
		$3 \times 3$ conv	
Transition Layer (1)	$56 \times 56$	$1 \times 1$ conv	
Dense Block (2)	$28 \times 28$	$2 \times 2$ average pool, stride 2	
Transition Layer (2)	$28 \times 28$	$1 \times 1$ conv	$\times 12$
Dense Block (3)	$14 \times 14$	$3 \times 3$ conv	
Transition Layer (3)	$14 \times 14$	$1 \times 1$ conv	
Dense Block (4)	$7 \times 7$ conv	$1 \times 1$ conv	$\times 32$
		$3 \times 3$ conv	
Classification Layer	$1 \times 1$	$7 \times 7$ global average pool	
	1000	1000D fully-connected, softmax	

Medical diagnostics relies heavily on the ability to accurately detect metastases, secondary cancerous growths from primary tumors. DenseNet169, a convolutional neural network architecture, offers a promising solution for this task. Its unique connectivity pattern maximizes information flow between network layers, allowing for robust inter-layer communication and feature reuse. This allows the model to capture intricate patterns of metastatic growths, even in the smallest clusters. *This project uses DenseNet169 to classify histopathology images for metastasis detection, fine-tuning its weights to the specific dataset.* Its inherent balance between

model complexity and overfitting makes it suitable for tasks with limited training data.

Comparar arquitecturas y ajustar hiperparámetros

## 4.2 Hyperparameters

The model's performance is optimized using Leslie Smith's one cycle policy [7], which offers a disciplined approach to hyperparameter selection. This policy help us determine the best possible values for the hyperparameters, thereby saving training time by avoiding suboptimal configurations. The Fastai library's implementation allows for seamless integration into the workflow. The optimal learning rate is determined by identifying the point before loss plateaus and divergence. Weight decay is chosen at the largest value that permits high learning rate. A small grid search is conducted across weight decay values for enhanced performance. We aim to determine the weight decay parameter that strikes a balance between achieving a low loss and allowing for the highest learning rate before encountering a sharp increase.

See [Figure 4.1](#). We want to select the largest weight decay that gets to a low loss and has the highest learning rate before shooting up. Out of the tested WD's, 1e-4 seems like the largest WD that allow us to train with maximal learning rate.

Next, we train only the heads while keeping the rest of the model frozen. Otherwise, the random initialization of the head weights could harm the relatively well-performing pre-trained weights of the model. After the heads have adjusted and the model somewhat works, we can continue to train all the weights.

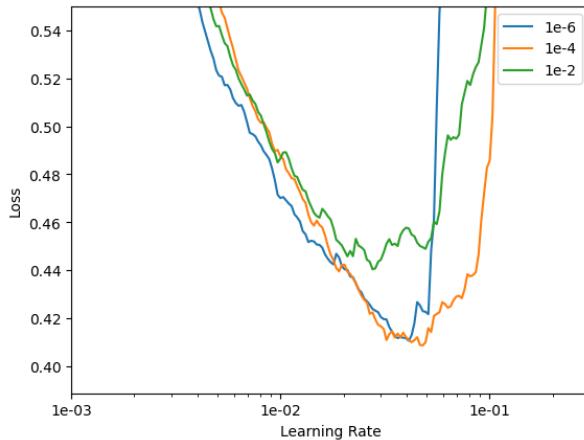


Figure 4.1

We next proceed to train the proposed model. The results are presented in the following section.

## 5 Results and discussion

The accuracy of the model is 95 %. We will finetune the model and compare later the results. For now we present the loss plot and confusion matrix.

## 5.1 Loss plot

As we can see in [Figure 5.1](#), there is a small rise after the initial drop which is caused by the increasing learning rate of the first half cycle. The losses are temporarily rising when the change in the learning rate drives the model out of local minima. However, this is fixed in the end when the learning rates are decreased.

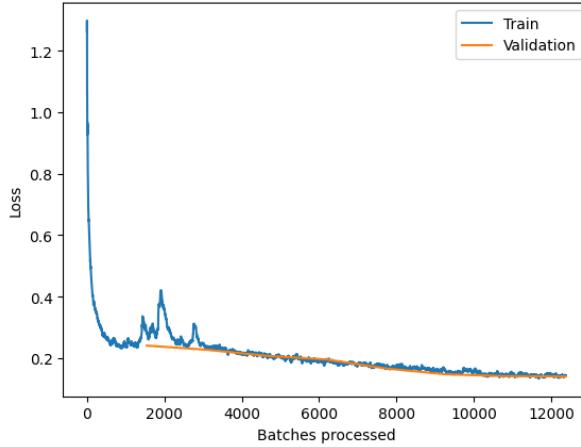


Figure 5.1

## 5.2 Confusion matrix

The confusion matrix provides a summary of the model's performance by presenting the counts of true negatives, false positives, false negatives, and true positives in a tabular format. It assesses the model's ability to distinguish between different classes, such as tumors and negative samples. See [Figure 5.2](#). The model shows promising accuracy, but further training is needed to improve its performance.

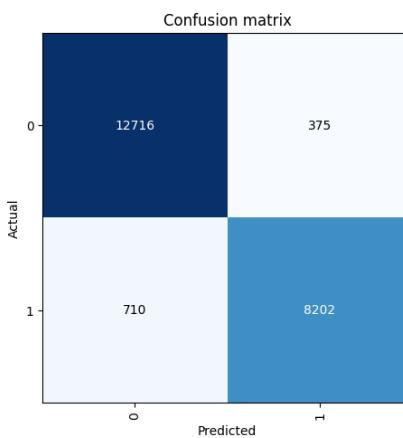


Figure 5.2

### 5.3 Enhancement of the model

The baseline model is refined by unfreezing its parameters and continuing its training. This process refines the model's understanding of intricate features and patterns specific to the dataset. The initial layers serve as a foundation, honed through exposure to vast amounts of visual data. As the model's weights adjust, it adapts to the nuances of the dataset. To mitigate the risk of destabilizing the learned representations, training should proceed with lower learning rates. This strategic fine-tuning process enables the model to delve deeper into the data, enhancing its ability to detect subtle distinctions and nuances.

The accuracy of the new model is 0.97 %. We obtained the following loss plot (Figure 5.3a) and confusion matrix (Figure 5.3b).

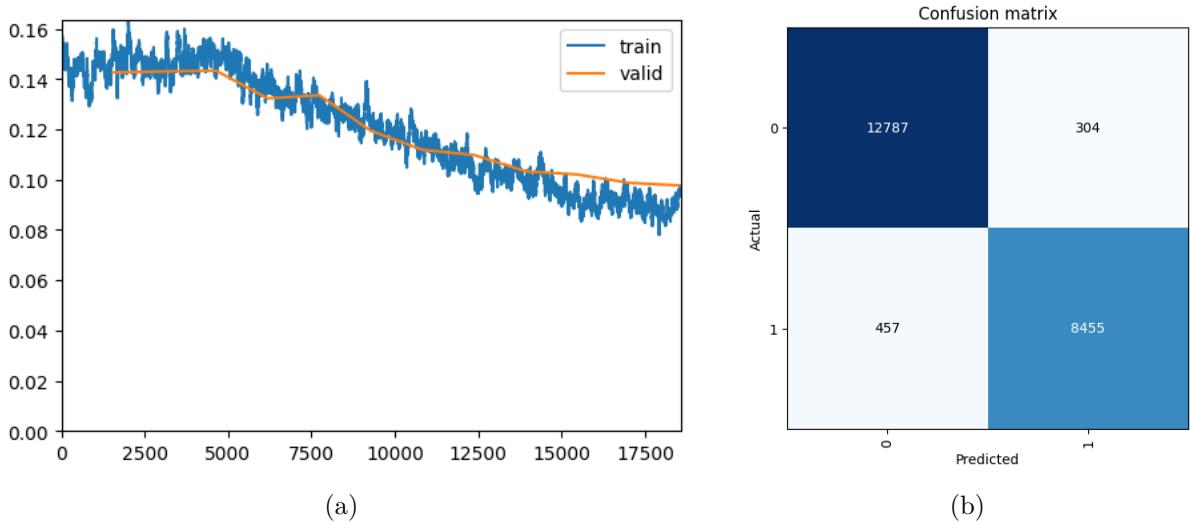


Figure 5.3

#### 5.3.1 ROC and AUC

The Area Under the ROC Curve (AUC) is a crucial metric for evaluating Convolutional Neural Networks (ConvNets) due to its ability to assess discrimination across all classification thresholds. This intuitive measure is particularly useful for ConvNets, which often use output probabilities as confidence indicators. AUC's resilience to class imbalance ensures reliable performance evaluation across diverse datasets, making it popular in real-world applications. Loss plots provide insights into training dynamics but may not directly correlate with classification prowess in class imbalance scenarios. AUC's holistic evaluation, adaptability to binary and multiclass tasks, and ability to evaluate unseen data make it the preferred metric for gauging ConvNet performance in image classification.

We present the roc curve in Figure 5.4. The area under the curve is 0.99, which indicates an exceptionally high level of model performance in discriminating between classes. This highlights the reliability of the model for the task of image classification regarding the identification of metastases.

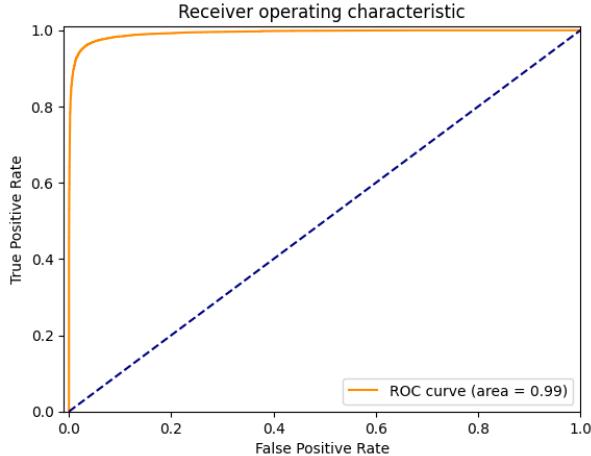


Figure 5.4

## 5.4 Discussion

This study developed a binary image classification model for cancer diagnosis, specifically targeting metastases in histopathology images. The model was trained using convolutional neural networks (CNNs) and transfer learning, and optimized hyperparameters to achieve optimal performance. The evaluation of the model's performance provides insights for future advancements in cancer detection using deep learning techniques. The Area Under the ROC Curve (AUC) was used as a crucial metric for evaluating ConvNets, considering the entire range of classification thresholds. AUC is robust to class imbalance and provides an intuitive measure of model performance across different decision thresholds. The baseline model achieved a high accuracy of 95%, with further enhancement resulting in an accuracy of 97%. Various visualization techniques were employed to analyze model performance and identify areas for improvement. The ROC curve and AUC analysis demonstrated the model's reliability for image classification, particularly in identifying metastases. The study contributes to ongoing efforts in leveraging deep learning techniques for automated cancer detection and emphasizes the importance of comprehensive evaluation metrics like AUC in assessing model performance.

## 6 Conclusion

In summary, this study introduces a robust methodology for developing a baseline model aimed at metastasis detection in histopathology images through the application of sophisticated deep learning techniques. Leveraging transfer learning and pretrained convolutional neural networks, the model demonstrates encouraging performance in automating cancer diagnosis processes. These findings represent a significant stride towards enhancing diagnostic precision and ultimately improving patient outcomes in the field of oncology. Moving forward, future research endeavors could focus on refining the model architecture, exploring additional data augmentation strategies, and validating the model's efficacy on external datasets. The continued advancement of accurate computer-aided diagnosis systems holds immense potential for transforming cancer detection and treatment paradigms, promising a brighter outlook for both patients and healthcare providers.

## A Code

The code for this project can be found at <https://github.com/christian-chavez/cancer-detection-project>

## References

- [1] Khalid A. AlAfandy et al. *Investment of Classic Deep CNNs and SVM for Classifying Remote Sensing Images*. 2020. DOI: [10.25046/aj050580](https://doi.org/10.25046/aj050580). URL: <http://dx.doi.org/10.25046/aj050580>.
- [2] Tavishee Chauhan, Hemant Palivela, and Sarveshmani Tiwari. “Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging”. In: *International Journal of Information Management Data Insights* 1.2 (2021), p. 100020.
- [3] *Histopathologic Cancer Detection*. URL: <https://www.kaggle.com/competitions/histopathologic-cancer-detection>.
- [4] Gao Huang et al. *Densely Connected Convolutional Networks*. July 2017. DOI: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243). URL: <http://dx.doi.org/10.1109/CVPR.2017.243>.
- [5] Zhiwen Huang et al. “Medical image classification using a light-weighted hybrid neural network based on PCANet and DenseNet”. In: *Ieee Access* 8 (2020), pp. 24697–24712.
- [6] Geert Litjens et al. *1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset*. en. May 2018. DOI: [10.1093/gigascience/giy065](https://doi.org/10.1093/gigascience/giy065). URL: <http://dx.doi.org/10.1093/gigascience/giy065>.
- [7] Leslie N. Smith. “A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay”. In: (2018). DOI: [10.48550/ARXIV.1803.09820](https://arxiv.org/abs/1803.09820). URL: <https://arxiv.org/abs/1803.09820>.
- [8] Ke Zhang et al. “Multiple feature reweight densenet for image classification”. In: *IEEE access* 7 (2019), pp. 9872–9880.
- [9] Ziliang Zhong et al. “Cancer image classification based on DenseNet model”. In: *Journal of physics: conference series*. Vol. 1651. 1. IOP Publishing. 2020, p. 012143.