

DEEP LEARNING FOR THE AUTOMATIC CLASSIFICATION OF TISSUE TYPES IN BREAST BIOPSIES

Javier Cordoba, Oscar Deniz and Gloria Bueno
 {javier.cordoba; oscar.deniz; gloria.bueno}@uclm.es
 VISILAB, E.T.S.I. Industriales, Ciudad Real, Spain

Abstract

Breast biopsies are crucial in the process of detecting a wide range of diseases such as breast cancer. The evaluation of these biopsies is performed by trained pathologists that are often overworked due to the increasing number of pathologies requested. Automatic tumour detection techniques have been developed, achieving very good results. In this work, we propose to classify breast biopsies in all the different types of tissue present in them. The tissue types were identified by hand-labeling them following the indications of an expert pathologist. Afterward, they were trained with different convolutional neural networks such as GoogleNet [12], AlexNet [4], SqueezeNet [5] and DenseNet [4]. Out of these four networks, GoogleNet outperformed all of them achieving 95.4% of accuracy. Finally, we tried to identify why the networks were underperforming while also suggesting how results could be improved.

Keywords: Deep learning · Convolutional neural networks · Breast biopsy · Tissue classification.

1 Introduction

In the last 10 years, several efforts have been made to apply well-known algorithms and techniques to try and detect the presence of tumours in whole slide images (WSI). Several approaches have been taken, ranging from the classic linear classifiers, self-organizing maps and decision trees to more novel approaches like artificial neural networks (ANN), template matching or support vector machines (SVM) using hand-crafted features [2]. The main disadvantages of these approaches are the limited number of features that can be hand-crafted, the fact that they cannot be modified while the model is being trained and the massive number of parameters to be optimized.

Lately, new architectures and techniques have been developed, which have made it possible to further increase the classification accuracy. Out of these architectures, the one that is currently more popular is the convolutional neural network

(CNN). These networks are similar to ANNs and are commonly used for image analysis as well as for other challenging applications such as speech recognition and natural language processing.

The main building block of CNNs is the convolutional layer, based on the concept of *convolution* or *kernel*. These kernels are small matrices that are applied to an image, the process of applying this kernel to an image is called a *convolution*. The concept of convolution is not new as it was widely used before CNNs as a way to apply filters to an image such as blur, sharpen, or edge detection.

The main advantages of CNNs against other methods are their capability of optimizing the kernels that make up the convolutional layers through network training, effectively removing the need of feature engineering and increasing the number of features explored in a single training session. Another advantage of these types of networks is that they are less prone to overfit, which is achieved by inserting a pooling layer (Figure 1) between consecutive convolutional layers. Other measures to control overfitting such as dropout are also used.

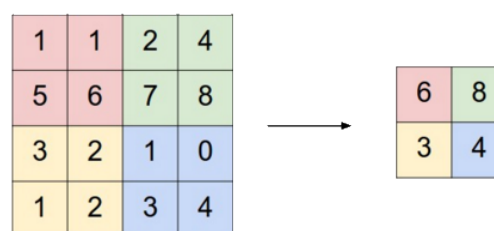


Figure 1: Example of max pooling operation with stride 2 (Original in [10])

But these advantages do not come free of drawbacks, one of the most relevant drawbacks is the increased number of parameters, which makes the training process more time consuming than typical ANNs. One way of solving this problem is to leverage the computing power of graphics processing units (GPU) in the form of parallel computing resulting in a dramatic reduction of the training time. Other approaches have been taken, such as using a Network in Network (NIN) architecture [7] that reduces the total training time by drastically reducing the total sum of parameters of the net-

work. Successful applications of this architecture are GoogleNet [12] and DenseNet [4].

Applications of CNNs [8, 14, 3] for tissue recognition are mostly focused classifying WSIs into tumoural and non-tumoural zones whereas in this work we use CNNs to try and classify all the different types of tissue present in a WSI as pathologists also use this information to guide their diagnostics.

This work is structured as follows. Section 2 describes the materials and datasets used. In Section 3 some state-of-the-art CNN architectures are discussed together with the chosen networks where the dataset will be trained. Finally, Section 5 draws some conclusions from the results obtained in Section 4.

2 Materials

The dataset used consists of 7 normal (i.e., benign) and 12 malign hematoxylin-eosin stained slides (Figure 2) provided by the *University Hospital at Ciudad Real (Spain)*, all patients waived their consent. Although the dataset might look like it is imbalanced, this is compensated by the fact that the benign WSIs are an order of magnitude larger than their malign counterparts (8500 MB vs. 408 MB), this extra imbalance was finally solved by using less benign WSIs for the labeling step.

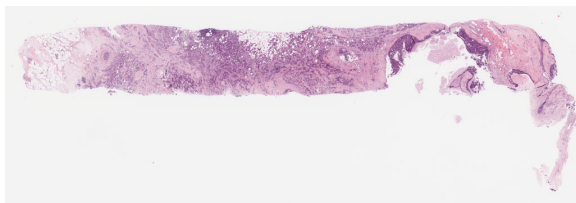


Figure 2: Example of a malign WSI

2.1 Labelling

WSIs were labeled with the help of a pathologist working at the hospital. The pathologist came to the laboratory and illustrated the visual and contextual differences between the different types of tissues present in a breast biopsy. With this help and a biopsy interpretation book [11] we labeled the WSIs by extracting patches of images that contained only one type of tissue. In cases where the type of tissue was not clear, then the image patch was discarded.

2.2 Dataset preparation

Extracted patches cannot be directly ingested by current CNNs, mainly due to their huge sizes. The

process of preparing the dataset for the ingestion by CNNs can be divided into two steps:

The first step is the random patch extraction, to match the image size with the CNN input size smaller patches must be extracted. The number of patches extracted was equivalent to the 70% of the total labeled image area with a size of 256x256 pixels. This size was chosen because most CNNs have an input size close to that size, also, since patch extraction is random images could overlap each other. A sample of the patches extracted in this step is shown in Figure 3.

The second step and the most tedious one is the non-related tissue patch cleaning because the shape of the extracted images is rectangular, and most of the tissues present in WSIs rarely have a rectangular shape. Figure 4 shows the importance of a good, clean dataset, mislabelling occurs due to the random patch extraction and the lack of heuristics applied to this process. Figure 5 shows the tissue distribution in the final dataset.

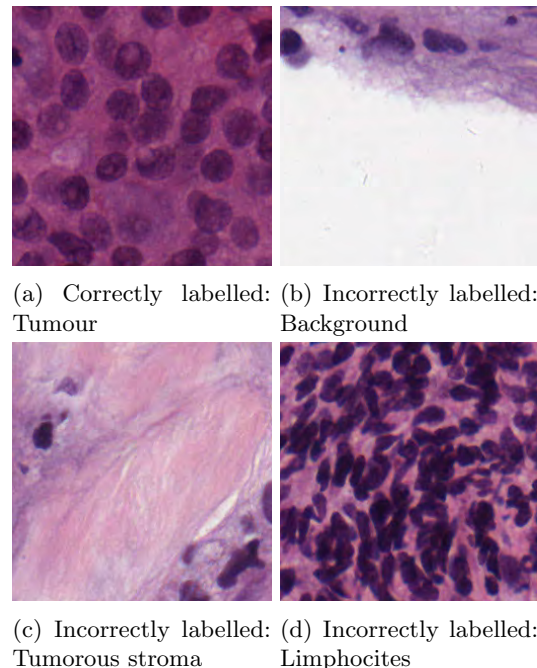


Figure 4: All of these images were labelled as *tumour* but only (a) truly belongs to that class. Dataset cleaning is crucial to obtain good results in the training phase.

3 Methods

As mentioned in section 1, NIN architectures allowed an improvement upon previous architectures where the network was limited by the number of parameters and therefore in its depth. This can be overcome by replacing a linear filter by a

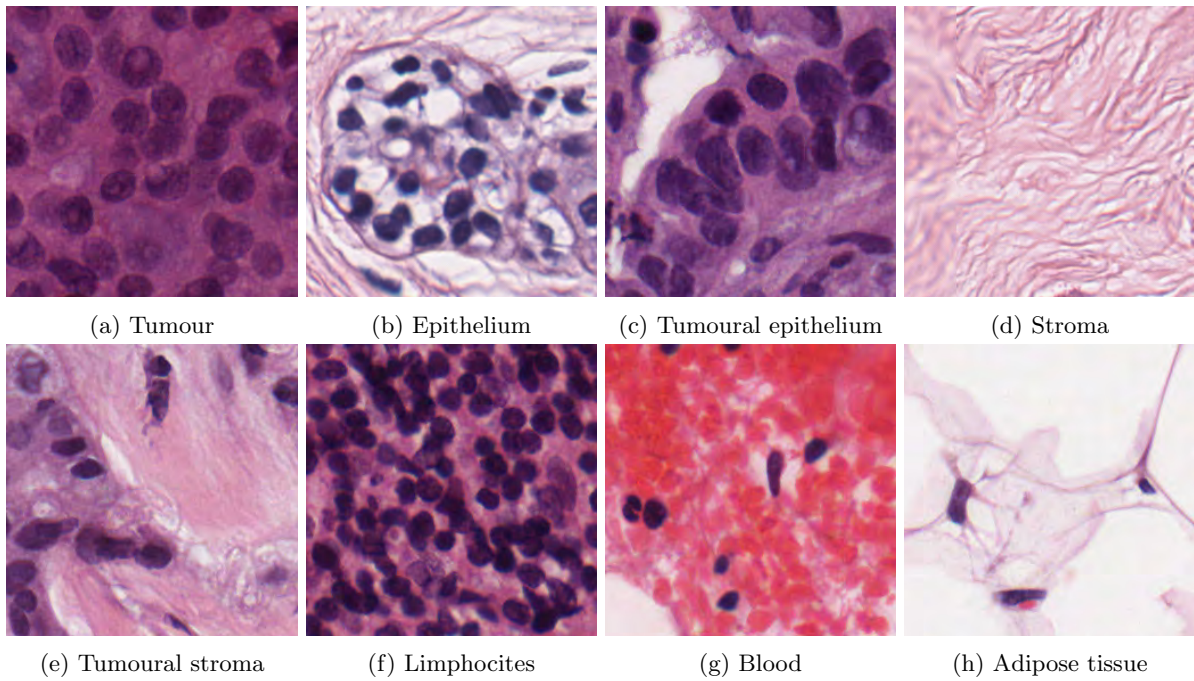


Figure 3: Patch images samples for every tissue present in a WSI

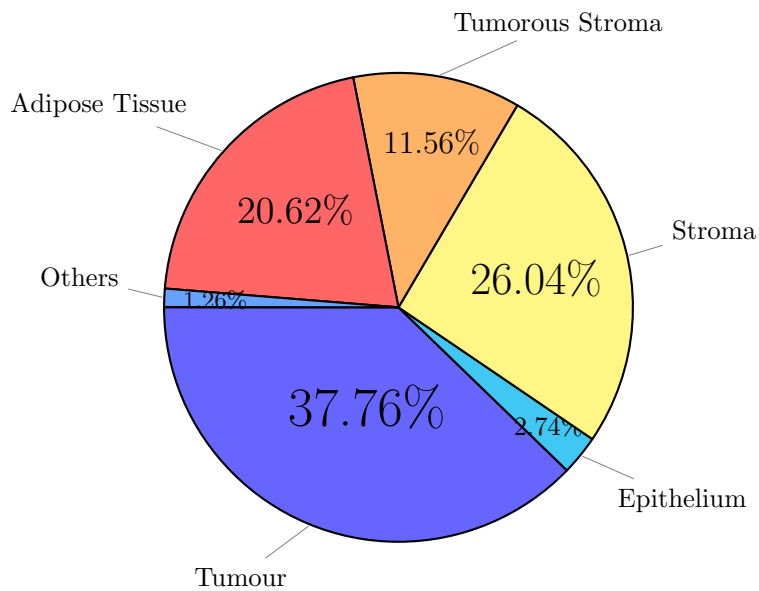
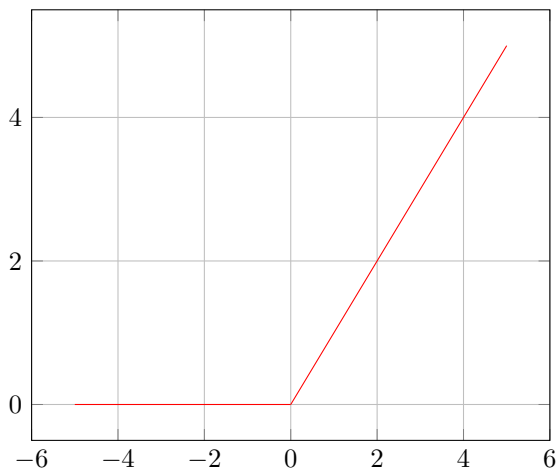


Figure 5: Percentages of tissues present in the final dataset



$$ReLU(x) = \max(0, x)$$

Figure 6: ReLU function

multilayer perceptron (MLP) [9]. An MLP is a feedforward ANN that consists of three or more layers. Except for the input layer all neurons use a nonlinear activation function, such as the sigmoid, hyperbolic tangent and more recently the rectified linear units (ReLU) [1] defined as in Figure 6. ReLU is preferred over the sigmoid or hyperbolic tangent because of its non-saturating properties, both sigmoid and hyperbolic tangent functions have a maximum value of 1 and a minimum value of -1 hence making large and dissimilar values have almost the same value, this is called the *vanishing gradient problem*.

A successful implementation of a NIN is GoogleNet’s *Inception* which is depicted in Figure 7.

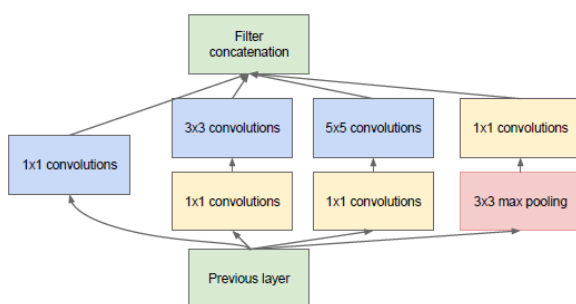


Figure 7: GoogleNet inception module. 1x1 convolution layers are used for dimensionality reduction, reducing the total parameters of the network hence reducing the total training time and increasing the network total depth. (Original in [13])

Other well-known networks, such as DenseNet overcome the vanishing gradient problem by using *dense blocks*. Dense blocks are comprised of a set of convolutional layers where every layer

has access to the feature map of every preceding layer. This allows the network to backpropagate the gradient much faster than traditional dense networks. The layers after every dense block are called *transition layers*; they consist of a 1x1 convolution operation; this follows the same reasoning in GoogleNet, 1x1 convolutions drastically reduce the number of parameters. A diagram of the dense blocks and transition layers can be seen in Figure 8.

Lastly, two more networks were chosen, AlexNet [6] and SqueezeNet [5]. AlexNet is a feedforward CNN, being the first to win the ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2012 challenge. It does not incorporate modern additions such as the previously discussed NIN or the 1x1 convolutional layers. Additionally, SqueezeNet is a network focused on achieving the same level of accuracy of AlexNet with 50 times fewer parameters and less than 0.5 MB model size, thus paving the way for CNN usage in embedded devices.

In conclusion, the networks chosen were:

- AlexNet
- SqueezeNet
- GoogleNet
- DenseNet-201

3.1 Training

The original dataset was split into two smaller datasets, training-validation and test. 93% of the original dataset was assigned to training-validation while 7% was set aside for testing the final network. Training-validation was in turn split into two subsets, training and validation, 85% of the training-validation dataset was assigned to training and 15% to validation so as to tune the parameters of the networks. Figure 9 shows the final percentages against the original dataset.

After obtaining three datasets the following data augmentations were applied only to the training images:

- Translation in both axis.
- Scaling in both axis.
- Reflection in the X axis.

All networks were trained for 6 epochs, with the exception of DenseNet-201 that was trained for 3 epochs due to the extremely long training times. No early stopping conditions were set, validation

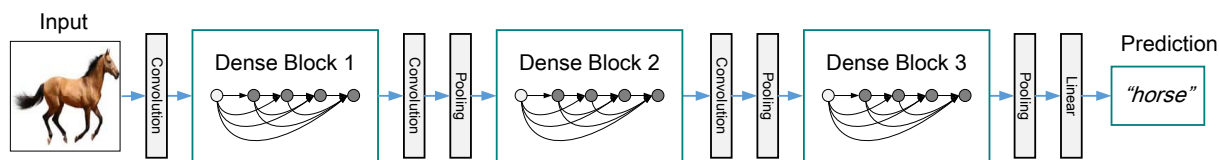


Figure 8: DenseNet dense blocks can be observed in conjunction with transition layers. (Original in [4])

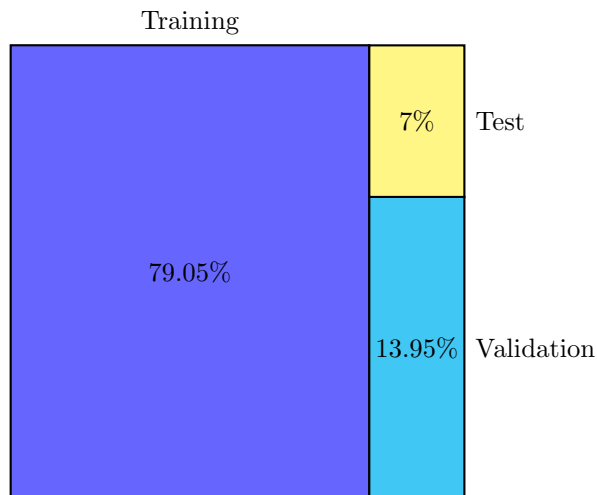


Figure 9: Number of images present in each of the three datasets.

frequency was set to 3 iterations, learning rate was set to 0.0003 and the solver was *SGDM*. The learning rate and validation frequency parameters were chosen ad-hoc.

The technique used to train the networks was transfer learning. Transfer learning takes advantage of pre-trained networks to train only the last two layers, those who take as input the features extracted by the previous layers and classify them into the new classes. This significantly reduces both the necessary data to train the network and the time required to do so.

4 Results

Training was performed in a laptop computer with an NVIDIA Geforce GTX 970M , training times in minutes *versus* the accuracy of each network are shown in Figure 10. As the figure shows, the best performing network with a 95.4% accuracy is GoogleNet, with SqueezeNet following it closely with a 93.6%. On top of this, it is surprising for DenseNet to be the worst network, taking into account the massive number of parameters it has when compared with GoogleNet (20 million vs. 7 million).

By looking at GoogleNet’s confusion matrix (Figure 11) an observation can be made, the network

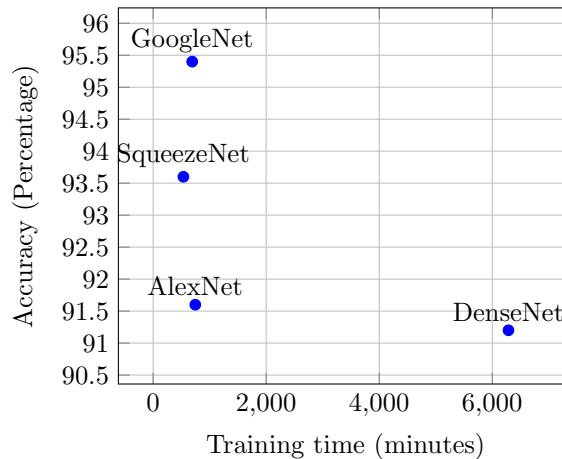


Figure 10: Training time vs Accuracy plot

struggles to differentiate between tumoural and tumoural stroma tissues. This observation was confirmed by looking at the classification errors on the test dataset, the conclusion drawn was that a particular WSI had very similar tumoural and tumoural stroma tissue, as shown in Figure 12. Lastly, all networks underperform in both lymphocytes and tumoural epithelium tissues due to the small number of images present in the dataset.

5 Conclusions

This work focuses on classifying all the tissues present in a breast biopsy instead of trying to classify each patch as positive or negative (tumour or non-tumour), four convolutional networks were trained of which GoogleNet scored the highest, with 95.4% accuracy, it outperformed DenseNet, a much deeper network in a tenth of the training time.

Future work could consider the use of an expert pathologist to label the biopsies, increasing the number of samples for underrepresented tissues, applying other data augmentation transformations or normalizing color in the patches due to the differences in staining between WSIs.

Acknowledgement

This work was partially funded by AIDPATH EC project Ref. No 612471 <http://aidpath.eu/>

Confusion Matrix

Output Class	Tumour	517	0	2	0	44	6	0	0	80.9%	9.1%	
	Epithelium	0	37	0	0	0	0	0	0	100%	0.0%	
	Tumoural Epithelium	0	0	0	0	0	0	0	0	NaN%	NaN%	
	Stroma	0	1	0	365	0	0	0	0	99.7%	0.3%	
	Tumoural stroma	11	0	0	25.1%	119	0	0	0	91.5%	8.5%	
	Lymphocytes	1	0	0	0	119	0	0	0	83.3%	16.7%	
	Blood	0	0	0	0	0	5	0	0	100%	0.0%	
	Adipose Tissue	0	0	0	0	0	0	289	0	100%	0.0%	
			97.7%	97.4%	100%	100%	72.8%	45.5%	100%	100%	95.4%	4.6%
			2.3%	2.6%	100%	0.0%	27.2%	54.5%	0.0%	0.0%	4.6%	
	Target Class	Tumour	Epithelium	Tumoural Epithelium	Stroma	Tumoural stroma	Lymphocytes	Blood	Adipose Tissue			

(b) CoResNet confusion matrix

Confusion Matrix

Output Class	Tumour	520	0	2	0	42	11	0	0	93.4%	6.6%	
	Epithelium	0	38	0	0	0	0	0	0	100%	0.0%	
	Tumoural epithelium	0	0	0	0	0	0	0	0	NaN%	NaN%	
	Stroma	0	0	0	339	0	0	0	0	100%	0.0%	
	Tumoural stroma	9	0	0	26	120	0	0	0	77.4%	22.6%	
	Lymphocytes	0	0	0	0	0	0	0	0	NaN%	NaN%	
	Blood	0	0	0	0	0	0	5	0	100%	0.0%	
	Adipose tissue	0	0	0	0	0	0	0	289	100%	0.0%	
			88.3%	100%	100%	92.9%	74.1%	0.0%	100%	100%	83.6%	16.4%
			1.7%	0.0%	100%	7.1%	25.9%	100%	0.0%	0.0%	16.4%	
	Target Class	Tumour	Epithelium	Tumoural epithelium	Stroma	Tumoural stroma	Lymphocytes	Blood	Adipose tissue			

(c) AlexNet confusion matrix

Confusion Matrix

Output Class	Tumour	529	0	2	0	104	10	0	0	82.0%	18.0%	
	Epithelium	0	38	0	0	1	0	0	0	97.4%	2.6%	
	Tumoural epithelium	0	0	0	0	0	0	0	0	NaN%	NaN%	
	Stroma	0	0	0	365	0	0	0	0	100%	0.0%	
	Tumoural stroma	0	0	0	0	57	0	0	0	100%	0.0%	
	Lymphocytes	0	0	0	0	41	1	0	0	100%	0.0%	
	Blood	0	0	0	0	0	0	5	0	100%	0.0%	
	Adipose tissue	0	0	0	0	0	0	0	289	100%	0.0%	
			100%	100%	100%	100%	35.2%	5.1%	100%	100%	91.6%	8.4%
			0.0%	0.0%	100%	0.0%	64.8%	90.9%	0.0%	0.0%	8.4%	
	Target Class	Tumour	Epithelium	Tumoural epithelium	Stroma	Tumoural stroma	Lymphocytes	Blood	Adipose tissue			

(c) AlexNet confusion matrix

Confusion Matrix

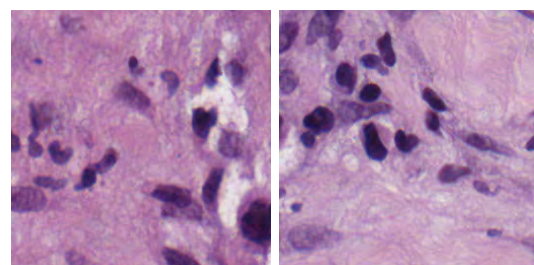
Output Class	Tumour	445	0	2	0	28	9	0	0	91.9%	8.1%	
	Epithelium	0	38	0	0	0	0	0	0	100%	0.0%	
	Tumoural Epithelium	0	0	0	0	0	0	0	0	NaN%	NaN%	
	Stroma	0	0	0	365	0	0	0	0	100%	0.0%	
	Tumoural stroma	84	0	0	0	134	0	0	0	81.5%	18.5%	
	Lymphocytes	0	0	0	0	0	2	0	0	100%	0.0%	
	Blood	0	0	0	0	0	0	5	0	100%	0.0%	
	Adipose tissue	0	0	0	0	0	0	0	289	100%	0.0%	
			84.1%	100%	100%	100%	32.7%	18.2%	100%	100%	91.2%	8.8%
			15.9%	0.0%	100%	0.0%	17.3%	81.8%	0.0%	0.0%	8.8%	
	Target Class	Tumour	Epithelium	Tumoural epithelium	Stroma	Tumoural stroma	Lymphocytes	Blood	Adipose tissue			

(d) DenseNet confusion matrix

Figure 11: Confusion matrices for each of the four networks trained. Target class corresponds to the true classes while output class corresponds to the classed predicted by the network.

References

- [1] Agarap, Abien Fred. “Deep Learning using Rectified Linear Units (ReLU)”. In: *CoRR* abs/1803.08375 (2018). arXiv: 1803.08375. URL: <http://arxiv.org/abs/1803.08375>.
- [2] Cheng, H.D. et al. “Automated breast cancer detection and classification using ultrasound images: A survey”. In: *Pattern Recognition* 43.1 (2010), pp. 299–317. ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2009.05.012>. URL: <http://www.sciencedirect.com/science/article/pii/S0031320309002027>.
- [3] Cruz-Roa, Angel et al. “Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks”. In: ().
- [4] Huang, Gao, Liu, Zhuang, and Weinberger, Kilian Q. “Densely Connected Convolutional Networks”. In: *CoRR* abs/1608.06993 (2016). arXiv: 1608.06993. URL: <http://arxiv.org/abs/1608.06993>.
- [5] Iandola, Forrest N. et al. “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 1MB model size”. In: *CoRR* abs/1602.07360 (2016). arXiv: 1602.07360. URL: <http://arxiv.org/abs/1602.07360>.
- [6] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems 25*. Ed. by Pereira, F. et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- [7] Lin, Min, Chen, Qiang, and Yan, Shuicheng. “Network In Network”. In: *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada*,



(a) Patch labelled as tumour (b) Patch labelled as tumoural stroma

Figure 12: A particular WSI had very similar tumour and tumoural stroma likely causing the networks to underperform.

- April 14-16, 2014, Conference Track Proceedings*. 2014. URL: <http://arxiv.org/abs/1312.4400>.
- [8] Liu, Yun et al. “Artificial Intelligence–Based Breast Cancer Nodal Metastasis Detection”. In: *Archives of Pathology & Laboratory Medicine* 0.0 (0). PMID: 30295070, null. DOI: 10.5858/arpa.2018-0147-0A. eprint: <https://doi.org/10.5858/arpa.2018-0147-0A>. URL: <https://doi.org/10.5858/arpa.2018-0147-0A>.
- [9] Rosenblatt, Frank. “Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.” In: *Psychological review*. 1958.
- [10] Santos, Leonardo Araujo. *Pooling Layer*. URL: https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/pooling_layer.html.
- [11] Schnitt, Stuart and Collins, Laura. *Biopsy interpretation of the breast*. Wolters Kluwer Health. ISBN: 9780781791465.
- [12] Szegedy, Christian et al. “Going Deeper with Convolutions”. In: *CoRR* abs/1409.4842 (2014). arXiv: 1409.4842. URL: <http://arxiv.org/abs/1409.4842>.
- [13] Tsang, Sok-Ho. *Review: GoogLeNet (Inception v1)- Winner of ILSVRC 2014 (Image Classification)*. URL: <https://medium.com/coinmonks/paper-review-of-googlenet-inception-v1-winner-of-ilsvlc-2014-image-classification-c2b3565a64e7>.
- [14] Wang, Dayong et al. “Deep Learning for Identifying Metastatic Breast Cancer”. In: *arXiv e-prints*, arXiv:1606.05718 (June 2016), arXiv:1606.05718. arXiv: 1606.05718 [q-bio.QM].



© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>).