Bunk, S.A.O.

[Email address]

Abstract

[Draw your reader in with an engaging abstract. It is typically a short summary of the document.   
When you’re ready to add your content, just click here and start typing.]

BEP Report mockup

[Document subtitle]

# Abstract

Table of Contents

[Abstract 1](#_Toc7953096)

# Introduction

In 2017, in the US alone, an estimated 252,710 new cases of invasive breast cancer and 63,410 new cases of in situ breast cancer were diagnosed amongst women [1]. Every woman undergoing treatment has some form of analysis performed on them to determine the appropriate treatment for their situation. These treatment options include techniques such as radiation therapy and chemotherapy [2][3]. Treatment may lead to a reduction in size or disappearance of the tumour. However, it is also possible that the tumour remains the same size, but the cellularity of the tumour reduces, meaning the tumour now consists of a lower number of larger cells.

Assessment of the cancer cellularity of a tumour is done to determine the effectiveness of previously applied treatment. The pathological examination of tissue removed during surgery allows for the determination of tumour cellularity. In current clinical practice pathologists manually determine the cellularity of tissue slides, which is a highly subjective and labour-intensive task. The variability in observers reduces the reliability and quality of the cellularity assessment [4].

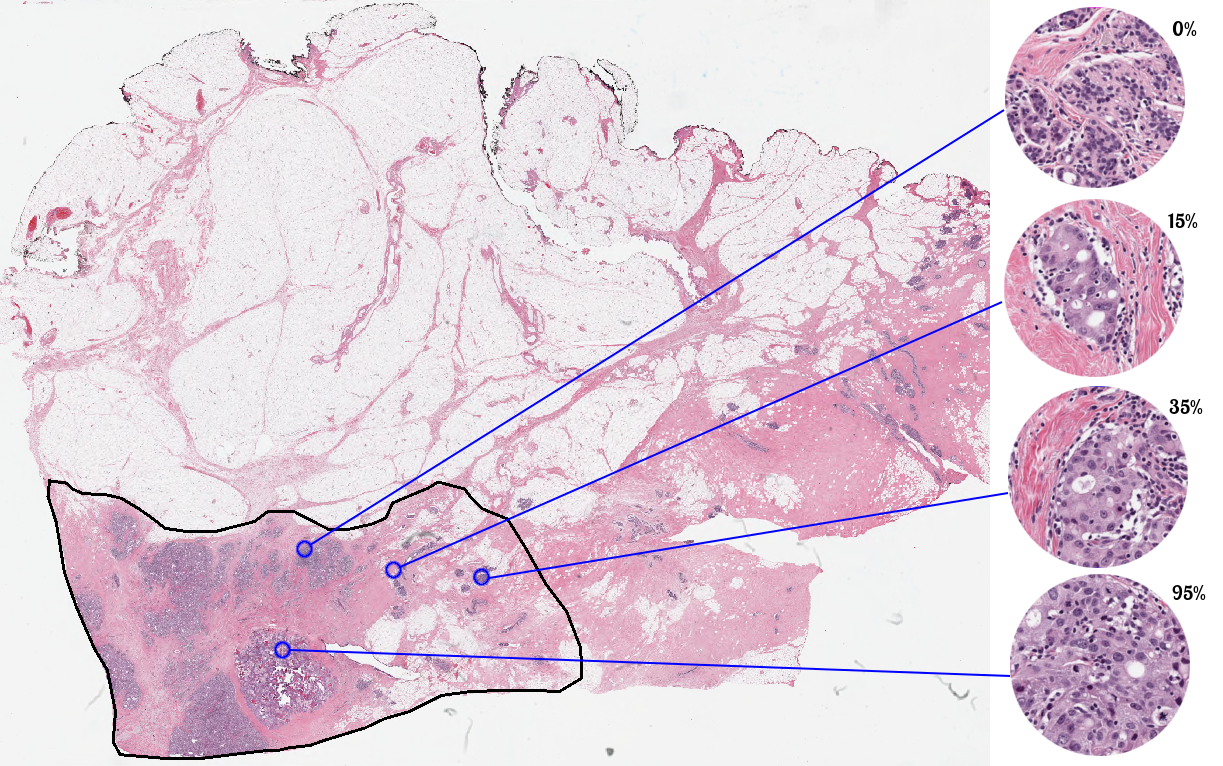


Figure 1: Haematoxylin and eosin stained slide used in the assessment of breast cancer cellularity. On the right the cellularity of specific points of the slide is estimated [4].

Automated assessment by image analysis can remove the intra- and inter-observer variability of cellularity determination and help increase the reproducibility of cellularity scores. In this study, we aim to assess cancer cellularity using deep learning techniques. Neural networks have achieved good results in the histopathological image analysis task and are highly useful tools for the computer-aided analysis of medical images [5][6][7].

There exist networks that are pre-trained and can easily be adapted to a different purpose, whilst retaining high performance in the new task. A selection of these networks we will look at in this paper is InceptionV3 [8], VGG19 [9], Xception [10] and ResNet50 [11], all of which were pre-trained on the ImageNet dataset and are included with the Keras python library. The networks achieved top-5 accuracies on the ImageNet dataset of 0.937, 0.900, 0.945 and 0.921 respectively [12]. In this paper we will retrain these networks to be used in the cancer cellularity determination task.

Assessing the effectiveness of any method on a certain task requires an evaluation metric. To compare automatically assessed cancer cellularity scores to manually obtained ones, a fair metric needs to be chosen. This study ventures to compare the three different networks with a multitude of metrics. We utilize existing metrics like Spearman correlation, Kendall’s tau, mean-squared error and prediction probability [13]. It is expected that networks which performed better on the ImageNet dataset will show the highest scores [14]. The results showed that Xception had the best scores for Kendall’s Tau, prediction probability, and Spearman correlation, but the worst score for mean squared error. VGG19 performed the worst according to the three metrics Xception scored the best in, and the best in the metric Xception scored the worst in. InceptionV3 scored between both other networks in every metric.

The dataset the networks will be applied on is the BreastPathQ dataset. This dataset is available in the SPIE-AAPM-NCI BreastPathQ: Cancer Cellularity Challenge. The training set contains 2579 patches extracted from 96 whole slide images. The whole slide images were made from breast tissue obtained from 64 patients. The training set has one tumour cellularity score assigned per patch [4].

# Methods

The training set was split up into three parts based on the patient ID to create our own training, validation and test datasets. 45 patients had their associated patches turned into 1489 train patches, 8 patients and their 429 patches became the validation dataset, and 10 patients and their 476 patches became the test dataset.

The choice was made to not include the top of the pre-trained networks InceptionV3, VGG19, Xception and ResNet50, as the top’s main function is to classify the input into the 1000 categories of the ImageNet dataset [8]. Instead, three custom layers were added to replace the top layers. These layers were constructed such that they would yield a single output, which would be the input image’s cellularity.

Figure 2: Layers added on top of the pre-trained networks.

A variety of data augmentation techniques were applied to the image patches. During training, these techniques were applied randomly to the image patches within a certain range as shown in Table 1.

Table 1: The augmentation techniques and their range values.

|  |  |
| --- | --- |
| Augmentation Technique | Value range (+/-) |
| Rotation | π |
| X translation | \*image width |
| Y Translation | \*image height |
| Rescaling | \*image size |
| X flipping | Yes |
| Y flipping | Yes |
| Per-pixel rescaling |  |
| Channel Shifting | 15 |
| Shearing | \*image size |
| Zooming | \*image size |

InceptionV3, VGG19 and Xception were trained for 100 epochs using the Adam optimizer with a learning rate of 0.001, with a mean squared logarithmic error loss. All but the three added custom layers had their weights frozen for this first batch of training. The epochs were the validation mean squared error was lower than in the previous epoch had their weights saved.  
  
After completing the first batch of training, the InceptionV3 and Xception networks also had some of their convolutional layers re-trained. For the InceptionV3 network the first 41 layers and for the Xception the first 66 layers remained frozen, the rest of the layers was re-trained using the Adam optimizer with a learning rate of 0.0001 and the same loss type as the first batch of training. Again, the epochs with the lowest validation mean squared error had their weights saved.

Each network had four different instances of itself trained, whose best saved weights were used on the dataset to produce predictions. These predictions were saved such that statistical analysis could be performed on the predictions of the test set. The predictions made on the test set were compared to the ground truth of the test set using four different statistical metrics, namely Kendall’s Tau, Prediction Probability, Mean Square Error and Spearman Correlation.

# Results

Table 2: Mean metric values for each network and their four instances.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Kendall's Tau | Prediction Probability | Mean Square Error | Spearman Correlation |
| InceptionV3 | 0,691392196 | 0,86938006 | 0,029771092 | 0,853678169 |
| VGG19 | 0,666924443 | 0,85630965 | 0,02569758 | 0,833274084 |
| Xception | 0,738082899 | 0,894326617 | 0,030283267 | 0,882558549 |

# Sources

[1] American Cancer Society. *Breast Cancer Facts & Figures 2017-2018*. Retrieved from <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf>

[2] Early Breast Cancer Trialists' Collaborative Group. (2011). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials. *The Lancet*, *378*(9804), 1707-1716.

[3] Von Minckwitz, G., Untch, M., Blohmer, J. U., Costa, S. D., Eidtmann, H., Fasching, P. A., ... & Jackisch, C. (2012). Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol, 30(15), 1796-1804.

[4] Grand-Challenges. *SPIE-AAPM-NCI BreastPathQ: Cancer Cellularity Challenge.* Retrieved from <https://breastpathq.grand-challenge.org/>

[5] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

[6] Veta, M., Heng, Y. J., Stathonikos, N., Bejnordi, B. E., Beca, F., Wollmann, T., ... & Hedlund, M. (2019). Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. *Medical image analysis*, *54*, 111-121.

[7] Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*.

[8] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[9] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[10] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[12] Keras, *Applications*, retrieved from <https://keras.io/applications/>

[13] <http://spiechallenges.cloudapp.net/competitions/14>

[14] Kornblith, S., Shlens, J., & Le, Q. V. (2018). Do better imagenet models transfer better?. *arXiv preprint arXiv:1805.08974*.