ELEC-E8739 - AI in health technologies, Project 2022

Nuclei Atypia Scoring in Breast Cancer Histological Images Using Transfer Learning

Milla Kosonen

1 Introduction

Breast cancer is the most common cancer type diagnosed in women (WCRF 2020). Early detection of breast cancer increases the probability of successful treatment and recovery. Computer-aided diagnosis of breast cancer has huge potential to ensure early detection and suitable treatment for breast cancer patients. Nuclear atypia is one of the criteria for diagnosis and grading of the progress of cancer (Jafarbiglo et al. 2018). Nuclear atypia refers to variations in shape, size, and internal organization of nuclei as compared to normal epithelial nuclei (Mitos & Atypia 2014). The nuclei become more atypical in these aspects as cancer advances. Grading the histology images manually by pathologists is a time-consuming process and influenced by the pathologist's subjective view and experience level. For these reasons, automatization of atypia scoring is necessary, and it would make detecting cancerous cells more effective and provide valuable information about the progression of the disease. The objective of this project is to determine nuclei atypia scores in breast cancer hematoxylin-eosin-stained biopsy slides using Inception-v3 network with pretrained weights.

2 Related work

Many computer-aided image analysis techniques have been developed to analyze breast cancer histology images and other medical images such as x-rays. Most techniques developed to analyze breast cancer images incorporate convolutional neural networks for automated feature learning and classification. Jafarbiglo and Danyali (2021) have proposed two systems for nuclear atypia grading. Both systems use a patch-based approach where the full image is divided into patches. They experimented on different patch sizes and patches of size 64x64 were found to be the most suitable. The first system attempts to find patches that contain critical nuclei and use these patches as input to CNN. The second system uses CNN for feature extraction and LSTM-based approach for classification.

Vesal et al. (2018) have used transfer learning to classify breast cancer images into four classes, namely, normal, benign, in situ carcinoma and invasive carcinoma. They have compared the performance of Inception-v3 and ResNet50 with pretrained weights in the classification task. Like Jafarbiglo et al. (2021), Vesal et al. (2018) have used a patch-based approach but without detecting critical nuclei. Vesal et al. (2018) used patches of size 512x512 and ensured 50% overlap between patches. They used large patch size to ensure that the patches contain relevant information for the classification task. They

achieved 97.50% image-wise test accuracy using ResNet50 and 91.25% test accuracy using Inception-v3. Rakhlin et al. (2018) have also used transfer learning in the same four-class classification task. In addition to Inception-v3 and ResNet50 Rakhlin et al. (2018) have also incorporated VGG-16 for feature extraction. For classification, they have used LightGMB which is an efficient gradient boosting tree.

3 Methods

The dataset used in this project is the one provided for the Mitos & Atypia 14 contest (Mitos & Atypia 2014). The dataset provides H&E-stained breast cancer biopsy slides at x10, x20 and x40 magnifications scanned with two different scanners, namely, Aperio Scanscope XT and Hamamatsu Nanozoomer 2.0-HT. Pathologists have analyzed breast cancer biopsy slides at x20 magnification and assigned a nuclear atypia score 1, 2 or 3 according to the Elston and Ellis grading system (Elston & Ellis 1991). According to the Elston and Ellis grading system, cell receives score 1 when the nuclei are small, have regular outlines and uniformity of nuclear chromatin and vary little in size. Score 2 is assigned when the cells appear larger than normal, have open, vesicular nuclei with visible nucleoli, and there is moderate variability in both size and shape. Score 3 is given when a marked variation in size and shape is detected. The grading is initially done by two pathologists. In case of disagreement a third pathologist analyzes the images, and the final score is determined by majority.

In this project, the training and testing of the network were performed with the frames at x20 magnification scanned with Aperio-scanner. The total number of frames with assigned score is 297. The dataset is divided into train and test sets such that 70% of samples are used for training and the remaining 30% for testing. Then 20% of the training data is used for validation. The size of the frames is reduced from 1376x1539x3 to 458x513x3 and divided to patches of size 200x200x3 with 50% overlap between them to speed up the training process and increase the number of samples. The score of the full image is assigned to each patch. Each image generates 12 patches resulting in 3564 patch samples. Table 1 shows the number of samples and the class distribution in each set. The patches are used as input to the network and image-wise predictions are obtained by majority voting based on the patch-wise predictions.

	Score 1	**Score 2**	**Score 3**	**Total**
Train	120	1536	324	1980
Validation	36	396	72	504
Test	120	732	228	1080

Table 1 The class distribution and total number of samples in each set.

3.1 Preprocessing

The H&E-stained images must be normalized before training because the stain can be unevenly distributed within the tissue (Das et al. 2020). A stain normalization method implemented in StainTools-package (Byfield 2018) is used to perform stain normalization. The normalization is implemented using stain matrix estimation via method proposed by Vahadane et al. (2016). Figure 1 shows the target image and examples of frames before and after normalization.

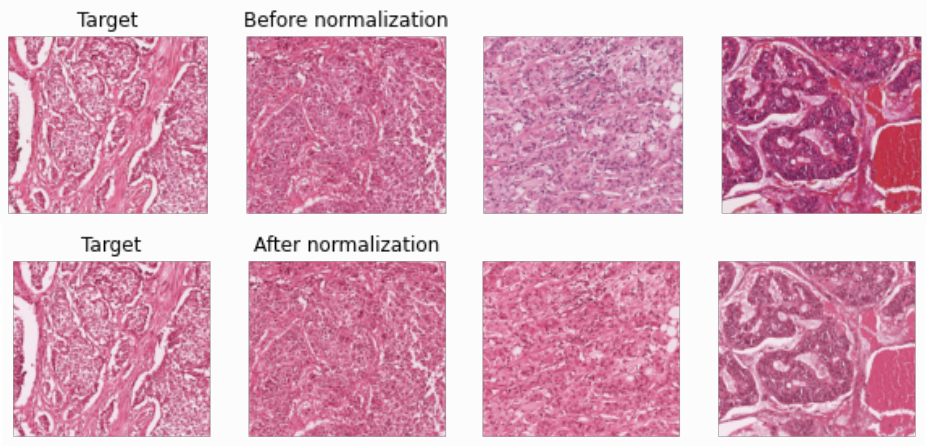

Figure 1 The target image and examples of images before and after stain normalization.

3.2 Model

The model used is Inception-v3 with pretrained weights from training on the ImageNet dataset. Transfer learning has been used in medical domain applications to obtain better results. It has been shown to increase speed of convergence and performance especially in small data regimes (Raghu et al. 2019). Inception-v3 model has been used in medical image classification tasks for example by Guan et al. (2019) to differentiate cervical lymphadenopathy in cytological images, by Mujahid et al. (2022) to detect pneumonia in x-ray images and by Vesal et al. (2018) and Rakhlin et al. (2018) for breast cancer histology image classification.

Due to the small number of training images, transfer learning is used to improve accuracy. The patches of size 200x200x3 are used as the input to the network. Average pooling layer is used to flatten the output from the Inception-v3 model to a vector of length 2048. Fully connected layer is then used to perform the classification. Since the problem is multiclass classification, softmax activation function is used in the final layer to obtain the probability distribution of the classes. The final prediction is obtained by choosing the class with highest probability. The full network architecture is shown in figure 2. The network was trained using stochastic gradient descent optimizer with 0.001 learning rate and 0.9 Nesterov momentum. The class distribution, seen in table 1, is imbalanced. There are more samples with score 2 than score 1 or 3. For this reason, class weights are incorporated to the training process.

Layer (type)	Output Shape	Param #
input_5 (InputLayer)	[(None, 128, 128, 3)]	0
inception_v3 (Functional)	(None, 2, 2, 2048)	21802784
global_average_pooling2d_3 (GlobalAveragePooling2D)	(None, 2048)	0
dense_3 (Dense)	(None, 3)	6147

Total params: 21,808,931
Trainable params: 6,147
Non-trainable params: 21,802,784

Figure 2 Model architecture

4 Results

After training, predictions were made on the test set. Predictions on the set resulted in 0.6102 patch-wise accuracy and 0.7222 image-wise accuracy. Confusion matrix of the predictions is shown in figure 3. ROC-curves were plotted for each class such that each class in turn is considered as the positive class. ROC-curves are presented in figure 4. The confusion matrix shows that the model confuses scores 1 and 2 often. The model detects score 3 well which is good since it is the most severe case of nuclear atypia and the most important case to detect. ROC-curves also show that grade 3 is predicted most accurately. AUC-value for grade 2 is 0.69. AUC-values for grades 1 and 3 are 0.73 and 0.83 respectively.

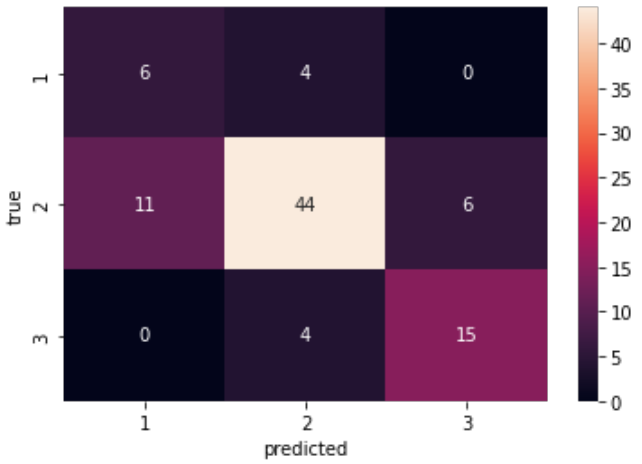

Figure 3 Confusion matrix of test set predictions.

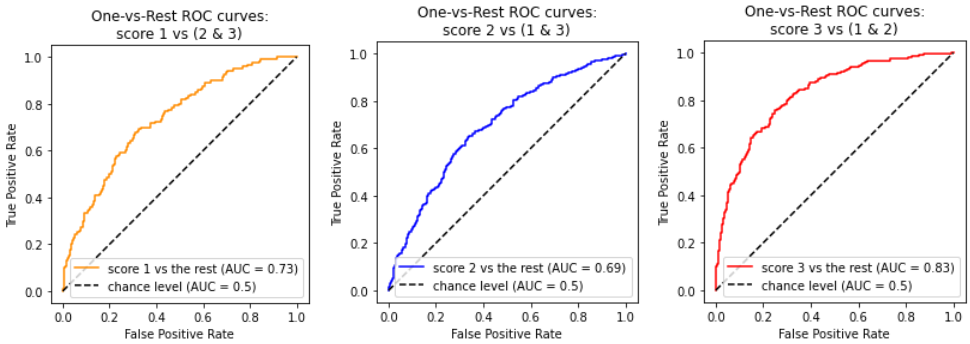

Figure 4 ROC-curves and AUC-values for each class.

Table 2 shows precision, specificity, F-score, and accuracy of each class separately and the overall value of these metrics. The results of Jafarbiglo et al. (2021) from the CNN+LSTM-based system are shown in table 3 for comparison. Precision tells the rate of true positive predictions out of all positive predictions. Specificity, also called true negative rate, tells the rate of true negative predictions out of all cases that should have been negative. F-score is used to measure accuracy.

$$Precision = \frac{TP}{TP + FP}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F - score = 2 * \frac{precision * recall}{precision + recall}$$

Score	Precision	Specificity	F-score	Accuracy
1	0.35	0.84	0.44	0.60
2	0.85	0.72	0.78	0.72
3	0.71	0.89	0.75	0.79
Overall	**0.64**	**0.82**	**0.66**	**0.72**

Table 2 Performance evaluation of classifying the images with the Inception-v3 model.

Score	Precision	Specificity	F-score	Accuracy
1	0.88	0.93	0.89	0.91
2	0.90	0.95	0.86	0.82
3	0.83	0.90	0.85	0.87
Overall	**0.87**	**0.93**	**0.87**	**0.87**

Table 3 Performance evaluation of classifying the images with system II by Jafarbiglo et al. (2021).

5 Conclusions

In this project, a transfer learning-based method was used for classification of breast cancer histology images. The images were divided into patches of size 200x200 to solve the problem of large image size and small dataset. This kind of patch-based approach is used for breast cancer image analysis for example by Jafarbiglo et al. (2021) and Vesal et al. (2018). When using patches for nuclear atypia scoring, however, it is important to detect the patches that contain the nuclei that are the most important for the classification since all regions of the original image might not contain atypical nuclei. In this project this kind of detection was not used but the patch size was chosen to be relatively large to ensure that critical nuclei would be included in the patches. To improve performance further, a nuclei segmentation algorithm could be used before classification.

Vesal et al. (2018) achieved 91.25% test accuracy using Inception-v3 network for a four-class classification of breast cancer images. Jafarbiglo et al. (2021) used the same dataset as used in this project. They achieved 86.67% accuracy using their proposed CNN feature extraction + LSTM based system. In this project, 72.2 % accuracy was achieved on the testing set which is close to the state-of-the-art results.

References

Byfield, P. 2018. StainTools-package. Available at: https://staintools.readthedocs.io/en/latest/

Das, A., Nair, M.S. & Peter, S.D. 2020. Computer-Aided Histopathological Image Analysis Techniques for Automated Nuclear Atypia Scoring of Breast Cancer: a Review. Journal of Digital Imaging, vol. 33, p. 1091–1121. doi: https://doi.org/10.1007/s10278-019-00295-z

Elston C.W. & Ellis I.O. 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. Histopathology, vol. 19:5. p. 403–410. doi: https://doi.org/10.1111/j.1365-2559.1991.tb00229.x

Guan, Q., Wan, X., Lu, H. et al. 2019. Deep convolutional neural network Inception-v3 model for differential diagnosing of lymph node in cytological images: a pilot study. Annals Of Translational Medicine, vol. 7:14. doi: https://doi.org/10.21037/atm.2019.06.29

Jafarbiglo, S. K., Danyali H. & Helfroush, M. S. 2018. Nuclear Atypia Grading in Histopathological Images of Breast Cancer Using Convolutional Neural Networks. 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), 2018, p. 89-93. doi: https://doi.org/10.1109/ICSPIS.2018.8700540.

Jafarbiglo, S.K. & Danyali H. 2021. Nuclear atypia grading in breast cancer histopathological images based on CNN feature extraction and LSTM classification. CAAI Transactions on Intelligence Technology. vol. 6:4. p. 426-439. doi: https://doi.org/10.1049/cit2.12061

Mitos & Atypia. 2014. Detection of Mitosis and Evaluation of Nuclear Atypia Score in Breast Cancer Histological Images. 22nd International Conference on Pattern Recognition. Available at: https://mitos-atypia-14.grand-challenge.org/Description/

Mujahid, M., Rustam, F., Álvarez, R. et al. 2022. Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network. Diagnostics (Basel). vol. 12:5:1280. doi: https://doi.org/10.3390/diagnostics12051280

Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. 2019. Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems. doi: https://doi.org/10.48550/arXiv.1902.07208

Rakhlin, A., Shvets, A., Iglovikov, V. & Kalinin, A.A. 2018. Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis. In: Campilho, A., Karray, F., ter Haar Romeny, B. (eds) Image Analysis and Recognition. ICIAR 2018. Lecture Notes in Computer Science, vol 10882. doi: https://doi.org/10.1007/978-3-319-93000-8_83

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S. et al. 2016. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. IEEE Transactions in Medical Imaging, vol. 35:8. p. 1962–1971. doi: https://doi.org/10.1109/tmi.2016.2529665

Vesal, S., Ravikumar, N., Davari, A., Ellmann, S., & Maier, A. 2018. Classification of breast cancer histology images using transfer learning. In International conference image analysis and recognition. p. 812-819. doi: https://doi.org/10.48550/arXiv.1802.09424

World Cancer Research Fund (WCRF). 2020. Breast cancer statistics. Available at: https://www.wcrf.org/cancer-trends/breast-cancer-statistics/ (Accessed 11.12.2022)