BREAST CANCER PN STAGING WITH DEEP LEARNING

Ilias Vasileiou[1*], Fotios Tagkalakis[1*], Argiris Diamandis[1], Evangelos Mitsianis[1], Evangelos Mavropoulos[1], Dimitrios Mallios[1], George Agrogiannis[2], Ioannis Pateras[3], Konstantinos Evangelou[3], Leena Joseph[4], Miles C Howe[4], Anshuman Chaturvedi[5], Jane Rogan[5], Paul A Townsend[6], Vassilis G Gorgoulis[1,3,6,7,8], Konstantinos Vougas[1,7,#]

1. DeepMed IO Ltd, 49 Peter St. Manchester, M2 3NG, UK.
2. 1st Department of Pathology, School of Medicine, National and Kapodistrian University of Athens, 75 Mikras Asias street, GR-11527, Athens, Greece.
3. Molecular Carcinogenesis Group, Department of Histology and Embryology, School of Medicine, National and Kapodistrian University of Athens, 75 Mikras Asias Str, Athens, GR-11527, Greece.
4. Department of Histopathology, Manchester University NHS Foundation Trust, Wythenshawe Hospital, Southmoor Road, Manchester, UK.
5. Department of Histopathology, The Christie NHS Foundation trust, Wilmslow Rd, Manchester, M20 4BX, UK.
6. Division of Cancer Sciences, Oglesby Cancer Research Building, Manchester Cancer Research Centre, Manchester Academic Health Science Centre, The University of Manchester, M20 4GJ, UK.
7. Biomedical Research Foundation of the Academy of Athens, 4 Soranou Ephessiou Str., Athens, GR-11527, Greece.
8. Center for New Biotechnologies and Precision Medicine, Medical School, National and Kapodistrian University of Athens, 75 Mikras Asias Str, Athens, GR-11527, Greece.

* These authors contributed equally to this work.
# To whom correspondence should be addressed:
Konstantinos Vougas, E-mail: kvougas@deepmed.io; Tel.: +44 731 044 143

ABSTRACT

Accurate and timely detection of metastases, as part of the globally recognized TNM standard, is pivotal for effective disease management and affects the therapeutic strategy along with the patient final outcome. Time constraints and high workload due to histopathologists under-staffing dramatically increases the probability of misdiagnosis in procedures such as the aforementioned one, which are laborious and time consuming. The current manuscript presents a method for the automatic determination of the pN-stage in lymph nodes from breast cancer patients that combines convolutional neural networks along with other machine learning algorithms. One of the contributions of this manuscript is a strategy for decreasing the training set size in a way that maintains its information content.

The proposed strategy utilized a subset (18%) of the patches available from Camelyon 16 train set in order to train a CNN, followed by a combination of machine learning classifiers trained on Camelyon 17 train set. The performance of the CNN was evaluated on Camelyon 17 train set and was found to generalize better than the baseline trained on the entire Camelyon 16 train set.

Index-terms: breast cancer, metastasis detection, deep learning, convolutional neural networks, machine learning

1. INTRODUCTION

The breast cancer TNM (Tumor, Node, Metastasis) staging system [1] plays a central role in the overall determination of the disease grading, which is the main factor guiding the therapeutic strategy. The TNM staging system aims to standardize the identification and quantification of the spread of cancerous cells to the regional lymph nodes (pN-stage). However, such a task is quite challenging, especially when the pathologist needs to identify small cancerous regions, known as micro-metastases, over large areas of normal tissue. The number of available pathologists, at a global scale, is inadequate to meet the diagnostic demand [2]. This translates to pathologists constantly working under pressure with limited amount of time to carry out their diagnostic tasks which causes delays in the delivery of the final diagnosis, dramatically increasing the chances of metastatic regions being missed leading to misdiagnosis with severe consequences to the final patient outcome.

Deep Neural Networks (DNN) consist of multiple layers of neurons, fully connected and interacting with one another, that are able to perform numerous abstractions and non-linear transformations [3]. These networks are very computationally expensive and require extremely high processing power for their training. This fact has been the main bottleneck for their utilisation in real-world problems. Recent developments, however in Graphic Proccesing Units (GPU) technology have made the use of DNNs feasible. As such, DNNs have been recently applied in a number of fields in order to perform prediction of drug toxicity in the liver [4], detection of the presence of

mutations in histopathology slides from lung cancer biopsies [5] and matching expert-pathologist performance in the detection of lymph node metastases in women with breast cancer [6] among others.

In the current work, we present a pipeline for processing WSIs containing LNs from breast cancer patients and diagnosing the pN stage through the combination of a Convolutional Neural Network (CNN-sub-class of DNNs specifically designed to process images) [3], along with other machine learning algorithms. The datasets used for pipeline development, machine learning training as well as predictive performance evaluation was the Camelyon 16 [7] training set for the CNN and the Camelyon 17 [7] training set having only WSI-level labels and not pixel-level annotation for the remaining pipeline.

2. METHOD

The pipeline presented in the current manuscript is presented in Figure 1 and consists of the following steps:

a. Tissue detection & patch extraction
b. CNN training for slim dataset compilation
c. CNN training on slim dataset
d. Inference of the Camelyon 17 training set; heatmap & segments generation
e. Training of machine learning algorithms at the WSI-level using segment-level derived features segment-level to predict WSI-level labels

2.1. Tissue detection & patch extraction

A WSI is a giga-pixel deep zoom image where the highest resolution level is approximately 200000 x 100000 pixels on the highest resolution level (zero level) and approximately 4GBs in size. Efficiently processing such an image has proven to be challenging. The initial step in the processing pipeline is to identify the tissue regions. It must be noted that accurate tissue identification is extremely important because in the vast majority of cases tumors are located at the edges of the LN tissue; therefore insufficient detection methods can potentially have a direct effect on the final diagnosis. We extensively experimented on Camelyon 16 & 17 WSIs and we found that the use of Otsu thresholding on grayscale transformed images followed by a series of morphological image operations allows the extraction of highly detailed tissue masks from the original WSI. Patches of 299 X 299 pixels size are then extracted from the WSI to be used as input to the CNN.

2.2. Training of patch-level CNN

An ImageNet pre-trained Inception-V1 [8] was used as the patch-level CNN. Normally, Inception-V1 uses patches of 224x224 pixel size, however patches of 299x299 pixels were used for the proposed model and an extra step of averaging was performed before the fully connected layer. The aforementioned scheme was chosen after initial experimention with Inception-V3 architecture. More specifically, it was found that a 0.5% increase in validation set ROC-AUC was achieved using the proposed strategy.

The model was trained only with patches from Camelyon16 train set [7] using data augmentation such as rotations, horizontal/vertical flips & color augmentation. 80% of WSIs were used for train set and the rest for the validation set. The ROC-AUC metric on the validation set was the criterion for model selection. All extracted patches from the respective WSIs were used for an initial round of training. Consequently, inference was performed on train and validation sets to create a list of normal patches being correctly inferred as normal with high confidence (probability of a patch being negative > 90%). We randomly removed 90% of these patches from both sets which resulted in much smaller training and validation sets (18% of the initial size), namely the slim training and validation sets. Finally, a second round of model training was performed using the aforementioned slim sets on an ImageNet pre-trained Inception-V1. All training was performed on an in-house NVIDIA-DGX1 deep-learning training server.

2.3. Training of machine learning classifiers for WSI-level label inference

Using the model described in 2.2, we inferred the Camelyon 17 training set and for each WSI we generated probabilistic heat-maps for detecting the presence of tumor hot-spots. We consequently performed segmentation on these heat-maps. Having in our possession a model that predicts only 103 out of 318 WSIs as negative, it is essential to use a high tumor probability segmentation threshold to compensate for the fact that several negative areas are incorrectly classified as metastases. The segmentation threshold we selected was 0.9. Segment-level features including but not limited to area, solidity, perimeter, inertia, compactness, circularity were extracted from the resulting segments. The values of each feature were grouped for all segments of each WSI. Their distribution characteristics were fed as second-level features to two machine learning algorithms in order to predict the final label per WSI (normal, ITC, micro and macro). More specifically, a Random Forest classifier and a combination of Random Forest and XGBoost (RF was

used to perform feature selection and XGBoost was used to perform the classification task) were initially considered. The classifiers' performance was evaluated and optimised based on a stratified 5-fold cross-validation scheme on Camelyon 17 train set. A voting scheme between the above mentioned classifiers was chosen as the final model. For the cases were the 2 classifiers disagreed the median label was chosen.

3. RESULTS

The original dataset which was 22 million patches in size delivered quadratic weighted k values of 0.6430 and 0.7142 at the WSI and patient levels respectively while the slim dataset, which was 4 million patches (18% the size of the original one), delivered 0.7316 and 0.8502 respectively (Tables 1-4). It was evident that the slim dataset generalised better while requiring much less computational resources to be trained, therefore we used only this one for all the processes described in the current manuscript.

The voting scheme scored quadratic weighted k values of 0.9012 and 0.9090 at the WSI and patient levels respectively of the Camelyon 17 training set (Tables 5,6), as determined through the 5-fold cross validation described above. This strategy was used to infer the labels of the Camelyon 17 test set that constitutes our initial submission to the contest.

4. DISCUSSION

The current work demonstrates a machine-learning training strategy that combines a CNN along with RF and XGBoost models to automatically diagnose the WSI-level for each lymph node and finally with the use of the Camelyon 17 rule based system, the pN-stage of breast cancer patients is computed. The current work does not include training with mainstream hard-negative mining strategies as seen in a number of submissions in the Camelyon 17 contest. We have selected to explore the basic hypothesis that in order to achieve better results, the initial training-set must be optimised for the specific CNN to represent all the different entities to be learned in a balanced and non-redundant manner. As such we have used the ImageNet pre-trained Inception-V1 to extract the information of which patches are easily learned by the network. A percentage of these patches was then removed and the initial network was re-trained. This process resulted in a CNN that generalises better to Camelyon 17 train set (Tables 1-4) delivering a ~10% and ~15% increase of the quadratic weighted k values at the WSI and patient levels respectively. It must be noted that this impressive increase in performance was achieved using a "slim" dataset, 18% of the initial size. This fact leads us to conclude that although we dramatically reduced the size of the dataset we did not decrease its information content and we relatively enriched it with images that were either harder for the network to learn or were under-represented in the original dataset. It must be also noted that the slim-dataset was 5 times faster to train.

5. CONCLUSION

A machine learning strategy was developed for the automatic diagnosis of the pN-stage from breast cancer patients.

A data-reduction strategy was used to reduce the size of the initial training-set that resulted in a slim-training-set 18% of the original one. The performance of the method has been proven to be competitive with other state of the art algorithms as determined by our robust stratified k-fold cross-validation on the Camelyon dataset.

Future work that will constitute our next submissions to the Camelyon 17 contest include: (a) incorporation of novel strategies to build multiple slim datasets each representing a unique view of the original data and (b) incorporation of robust hard-negative mining strategies.

6. REFERENCES

[1] Meier FA. The landscape of error in surgical pathology. InError Reduction and Prevention in Surgical Pathology 2015 (pp. 3-26). Springer, New York, NY.

[2] Meeting pathology demand - Histopathology workforce census 2017/2018, The Royal College of Pathologists (report)

[3] LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015 May;521(7553):436.

[4] Xu Y, Dai Z, Chen F, Gao S, Pei J, Lai L. Deep learning for drug-induced liver injury. Journal of chemical information and modeling. 2015 Oct 13;55(10):2085-93.

[5] Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, Moreira AL, Razavian N, Tsirigos A. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature medicine. 2018 Oct;24(10):1559.

[6] Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Laak JA, Hermsen M, Manson QF, Balkenhol M, Geessink O. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama. 2017 Dec 12;318(22):2199-210.

[7] Litjens G, Bandi P, Ehteshami Bejnordi B, Geessink O, Balkenhol M, Bult P, Halilovic A, Hermsen M, van de Loo R, Vogels R, Manson QF. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience. 2018 May 31;7(6):giy065.

[8] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. InProceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).

[9] Bandi P, Geessink O, Manson Q, Van Dijk M, Balkenhol M, Hermsen M, Bejnordi BE, Lee B, Paeng K, Zhong A, Li Q. From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. IEEE transactions on medical imaging. 2018 Aug 27;38(2):550-60.

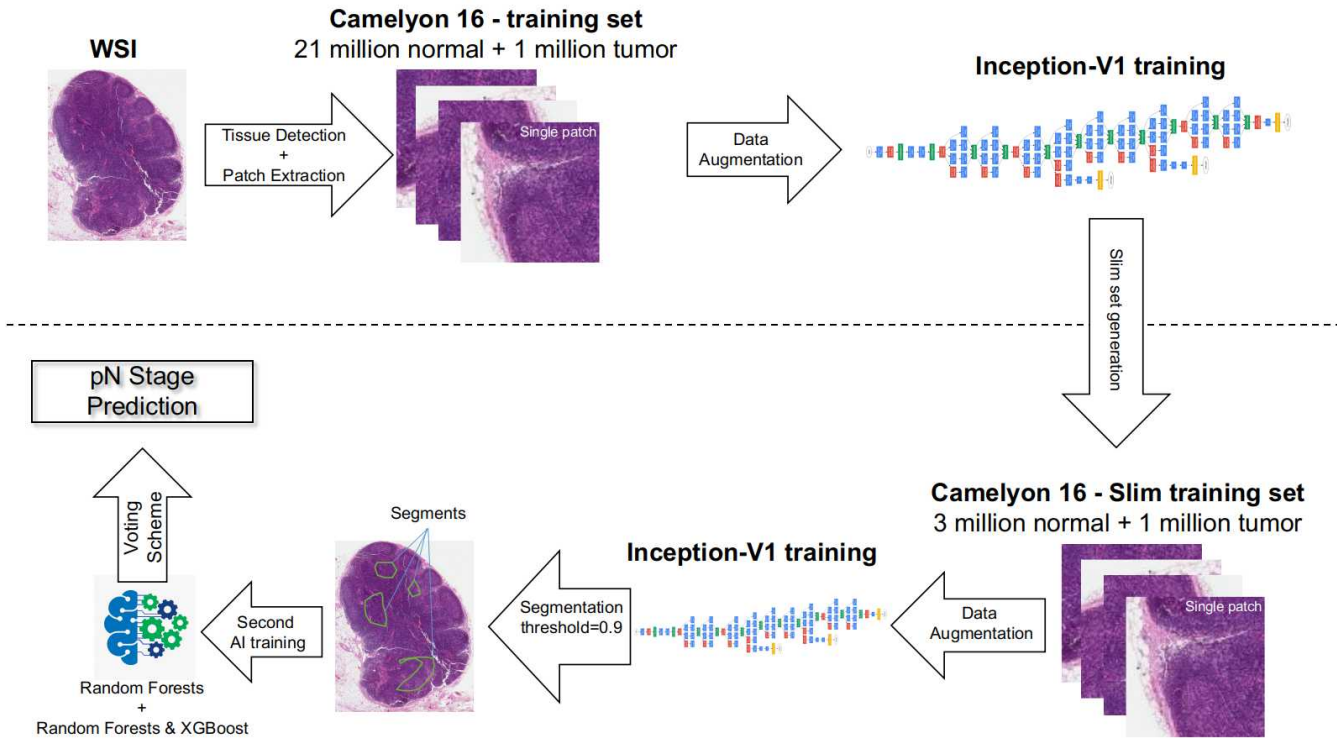

Figure 1. Schematic representation of the pipeline utilised in the current manuscript

Table 1. Confusion Matrix: Inception V1 - original dataset WSI-Level

k	0.6430	Predicted			
		Negative	ITC	Micro	Macro
Ground Truth	Negative	105	141	72	0
	ITC	2	21	13	0
	Micro	3	3	52	1
	Macro	0	0	17	70

Table 2. Confusion Matrix: Inception V1 - original dataset Patient-Level

k	0.7142	Predicted				
		pN0	pN0(i+)	pN1(mi)	pN1	pN2
Ground Truth	pN0	0	9	15	0	0
	pN0(i+)	0	5	6	0	0
	pN1(mi)	0	4	17	0	0
	pN1	0	0	4	19	7
	pN2	0	0	0	0	14

Table 3. Confusion Matrix: Inception V1 - slim dataset WSI-Level

k	0.7316	Predicted			
		Negative	ITC	Micro	Macro
Ground Truth	Negative	99	197	22	0
	ITC	4	29	3	0
	Micro	0	7	52	0
	Macro	0	1	24	62

Table 4. Confusion Matrix: Inception V1 - slim dataset Patient-Level

k	0.8502	Predicted				
		pN0	pN0(i+)	pN1(mi)	pN1	pN2
Ground Truth	pN0	0	20	4	0	0
	pN0(i+)	0	11	0	0	0
	pN1(mi)	0	4	17	0	0
	pN1	0	0	3	23	4
	pN2	0	0	0	1	13

Table 5. Confusion Matrix: Final Ensemble - slim dataset WSI-Level

k	0.9012	Predicted			
		Negative	ITC	Micro	Macro
Ground Truth	Negative	310	4	4	0
	ITC	34	1	1	0
	Micro	10	4	40	5
	Macro	2	1	12	72

Table 6. Confusion Matrix: Final Ensemble - slim dataset Patient-Level

k	0.9090	Predicted				
		pN0	pN0(i+)	pN1(mi)	pN1	pN2
Ground Truth	pN0	22	1	1	0	0
	pN0(i+)	11	0	0	0	0
	pN1(mi)	5	0	16	0	0
	pN1	0	0	0	29	1
	pN2	0	0	0	2	12