CNNs in lymph node metastasis

Deep learning models have potential to improve performance of automated computer-assisted diagnosis tools in digital histopathology and reduce subjectivity. The main objective of this study was to further improve diagnostic potential of convolutional neural networks (CNNs) in detection of lymph node metastasis in breast cancer patients by integrative augmentation of input images with multiple segmentation channels. For this retrospective study, we used the PatchCamelyon dataset, consisting of 327,680 histopathology images of lymph node sections from breast cancer. Images had labels for the presence or absence of metastatic tissue. In addition, we used four separate histopathology datasets with annotations for nucleus, mitosis, tubule, and epithelium to train four instances of U-net. Then our baseline model was trained with and without additional segmentation channels and their performances were compared. Integrated gradient was used to visualize model attribution. The model trained with concatenation/integration of original input plus four additional segmentation channels, which we refer to as ConcatNet, was superior (AUC 0.924) compared to baseline with or without augmentations (AUC 0.854; 0.884). Baseline model trained with one additional segmentation channel showed intermediate performance (AUC 0.870-0.895). ConcatNet had sensitivity of 82.0% and specificity of 87.8%, which was an improvement in performance over the baseline (sensitivity of 74.6%; specificity of 80.4%). Integrated gradients showed that models trained with additional segmentation channels had improved focus on particular areas of the image containing aberrant cells. Augmenting images with additional segmentation channels improved baseline model performance as well as its ability to focus on discrete areas of the image.

1. Introduction

Whether metastatic lesions are present in sentinel lymph nodes (SLN) is an important prognostic marker for early-stage breast cancer [1]. Large tumor size and perivascular invasion are associated with SLN involvement [2]. Therefore, the presence of metastatic tissue in SLN of breast cancer patients often represents a disseminated disease associated with poor prognosis and limited treatment options [3][4]. Since the status of SLN cannot be determined by clinical examination alone, SLN biopsies are routinely performed on early-stage breast cancer patients and are assessed by clinical pathologists for metastasis [1].

Accurate histopathological diagnosis empowers clinicians to recommend targeted treatment options specific for each patient [5]. Such histopathological diagnoses often occur in a time-limited setting during surgery, requiring a rapid classification of metastatic status, which greatly influences intraoperative decisions made whether to proceed with invasive treatment options or not [5][6]. For example, SLN-positive patients are recommended to receive axillary lymph node dissection, which is associated with significant permanent impairment [1]. However, detection procedures conducted by pathologists are often time consuming and subjective [7]. For example, metrics such as tumor cell percentage or quantification of fluorescent markers for estrogen receptor and/or HER-2 status are tasks that are often associated with inter-observer variability [8]. Furthermore, for the task of micro-metastases detection under simulated time constraints, pathologists have shown an underwhelming performance of 38% [3].

Whole-slide imaging systems have improved over the years, and are now capable of producing digitized, high-resolution, giga-pixel whole-slide images (WSI) of histopathology slides [9]. Using this technology, histopathological assessments can be done on a computer screen rather than using light microscopes. Digitization of workflow in pathology laboratories can reduce patient identification errors and save time for both pathologists and laboratory technicians [8]. Digitization of WSI has also enabled the development of automated computer-assisted diagnosis (CAD) platforms [9][10]. Automated computer-assisted diagnosis (CAD) has the potential to improve the speed and accuracy of histopathological diagnoses as well as reducing subjectivity [10].

Advancements in computer vision, most notably deep learning, has enabled researchers to extract more abstract features from large amounts of high-resolution medical images [11]. Therefore, high-resolution WSIs that contain complex features are suitable for application of deep learning strategies using convolutional neural networks (CNNs) [12]. The Cancer Metastases in Lymph Nodes Challenge 2016 (Camelyon16) found best algorithms to be performing significantly better than pathologists with time constraints and comparable to pathologists without time constraints [3]. Lymph Node Assistant (LYNA), an algorithm developed by Google AI Healthcare [5], managed to achieve 99.0% area under the curve in detection of micro- and macro-metastases from lymph node blocks [10]. Furthermore, pathologists with assistance from LYNA achieved 100% specificity and showed improved sensitivity over performance achieved by LYNA alone, which suggests the benefit of human intervention in CAD and room for improvement [10].

Weights previously trained on large-scale datasets such as ImageNet [13] can be used to initiate training of the model on a different task. Such strategy known as transfer learning have reportedly shown to facilitate faster convergence and better prediction performance for CNNs in digital pathology [14]. For example, Nishio et al. [15] have shown that VGG16 [16] with transfer learning performed better overall than same models trained without transfer learning. However, transfer learning does not guarantee better performance, because performance of models trained with the same architecture and pre-trained weights have been observed to differ greatly [16].

Data augmentation strategies, such as stain color normalization and morphological transformations of the input images, are often employed for digital histopathology image analyses, to improve model generalizability and robustness [17][18]. Algorithms such as WSI color standardizer (WSICS) [19] and Stain Normalization using Sparse AutoEncoders (StaNoSA) [17] demonstrated that data augmentation can improve performance of existing CAD systems for tasks such as necrosis quantification and nuclei detection, respectively. Therefore, we sought other data augmentation approaches to further improve performance of existing CAD models in histopathology.

Pathologists look for histological features such as nuclei, mitotic figures, tissue types, and multicellular structures such as tubules to make and justify their diagnoses. For example, pixel-wise detection of cytological features such as epithelial cell nuclei, epithelial cell cytoplasm, and the lumen were used for the higher-level tasks of gland segmentation and prediction of tumor grade on the Gleason grading scheme in prostate cancer [20][21]. Another study showed that local descriptors such as the distribution of cell nuclei was one of the most significant features used by a random forest model to detect metastasis from digital pathology images [22].Therefore, we investigated if we could further improve the performance of baseline CNN models by providing multiple segmentation channels of the input images with pixel-wise histological annotations of such features. Each of these segmentation channels can be extracted by U-net, a CNN model designed for semantic segmentation of biomedical images [23], which can then be integrated onto the original images depth-wise prior to input into the baseline model. We hypothesized that training CNN models with additional multiple segmentation channels will boost its performance over the baseline model. The specific aims of this project were: 1) train and evaluate a baseline CNN model for detecting breast cancer metastasis from digital histopathology images of lymph node sections using the PatchCamelyon (PCam) dataset ; 2) train four instances of a U-net model for semantic segmentation of histological features including the nucleus, mitotic figures, epithelium, and tubule using four independent datasets curated previously [24]; 3) train and evaluate a second instance of the baseline model with additional segmentation channels of images from the same test set to compare to the baseline model.

2. Discussion

Deep neural networks were inspired by the organization of the human visual cortex [25]. By designing a model which mimics the human brain, researchers were able to gain significant advances in various fields, notably in computer vision and CAD [26][27]. Likewise, the central motivation of this study was to modify a model to mimic how a pathologist sees a histology image and assess the model's performance. In the eyes of a pathologist, histological features like cell nuclei, cell type, cell state, and multicellular structures are recognized naturally, which all contribute to the pathologist's ability to recognize malignancy from a given histology image [21]. Objective and quantitative segmentation of histologic primitives such as the nuclei and glandular structures is one of the major interests of digital pathology . Accordingly, we extracted multiple segmentation channels that captured such histological features, which were used to augment input images during the training phase. As previously demonstrated by the whole-slide image color standardizer (WSICS) algorithm, which reduced the effects of stain variations and further improve performance of a CAD system by incorporating spatial information, we incorporated the spatial information of histological structures to improve our model's classification performance.

For our problem of detecting metastatic cells from digital histopathology images of sentinel lymph node sections extracted from breast cancer patients, we observed improvements in both sensitivity and specificity when the models were provided with one or more additional segmentation channels. Deep neural networks and features generated by these models have

been criticized for their lack of interpretability. However, we also showed that through the IG algorithm [28] that the models trained with additional segmentation channels were able to establish regions of interest containing malignant-looking cells or structures when the baseline model could not.

Our findings suggest that even for models of CAD with considerably high predictive performance, their performance can be further improved by augmenting input images with multiple additional segmentation channels. Diagnostic errors are expensive both for the patient and the healthcare system because false positive results can lead to unnecessary calls for additional diagnostic tests or treatments on a healthy individual, and false negative results can lead to a lack of care for patients who need early medical intervention [29][30]. Furthermore, both types of errors can lead to potential litigations. Therefore, it is important to consider our method of data augmentation to further improve the performance of existing CAD tools and those in development. However, it should be noted that although the IG algorithm was able to visualize the differences in feature attribution between models, we still do not have a clear understanding as to why some models have focused appropriately on regions containing malignancy and yet made incorrect decisions on some of the images. Nonetheless, proper focus and extraction of regions of interest can potentially relieve the burden of pathologists, who serve majority of their time scanning benign areas without malignancy . Moreover, the ability of automated CAD tools to speedily and objectively quantify histopathological features such as tumor cell percentage and disease grade is much needed [8].

Many of our predecessors in digital histopathologic image analysis have used transfer learning techniques, mostly by using weights from CNN architectures pre-trained on large generalized image datasets such as ImageNet , to reduce training time and to benefit from potential performance benefits [31]. Although there was a significant reduction in training time, the performance results were highly variable, even with the same pre-trained CNN architectures . In our study, we observed that VGG16 with transfer learning performed better than the baseline, albeit with substantially higher number of parameters. Our approach to augment the training phase of CNN models can also be seen as a method of transfer learning, albeit different from our predecessors in that 1) we transferred knowledge gained from the same type of images, specifically from histopathology; and 2) rather than transferring only the weights, we used entire pre-trained networks in parallel to extract new segmentation channels from the same input image [31]. These two key differences potentially contributed to the improvements in performance benefits that were observed in this study, including convergence at lower loss value and increased generalizability to unseen data, with little additional computational cost to the classifier models.

However, a major limitation of this study was that the annotated histology images used to train the U-nets were not from the same tissues. For example, the nuclei and tubule segmentation datasets were images from colorectal cancer patients [32] whereas the epithelium and mitosis segmentation datasets were images from breast cancer patients . Furthermore, our main benchmark dataset, PCam, consisted of images from sentinel lymph node sections . Training the U-nets and the subsequent baseline model with a single dataset with multiple annotations for nuclei, mitotic figures, multicellular structures, and other histological features has potential to improve model performance even further.

3. Conclusions

In summary, we demonstrated that improvements were made in both sensitivity and specificity when deep learning models were trained with additional segmentation channels of input images. IG analysis suggested that these additional segmentation channels help the models to orient their attention to specific regions of the image containing malignancies, although we found examples where better focus did not necessarily lead to correct classification. However, further analyses should be repeated using larger datasets with better resolutions and deeper models in the future to investigate if our results can be replicated under those circumstances. Interpretation of deep learning models still remains a challenge and presents room for improvements.

Furthermore, the feature segmentation pipeline using U-net can be extended to segment other, more complex histological features such as different tumor tissues, inflammation, and necrosis, among many others. We demonstrate that data augmentation with prior extracted features have potential to further improve the performance of CAD tools in digital histopathology and other tasks in medical image analyses, in which even small improvements in performances has significant implications for the patient's clinical outcomes.

References

1. The Expert Panel on SLNB in Breast Cancer. Sentinel Lymph Node Biopsy in Breast Cancer in Early-Stage Breast Cancer, George, R.; Quan, M.L.; McCready, D.; McLeod, R.; Rumble, R.B., Reviewer; Cancer Care Ontario: Toronto, ON, Canada, 2009.

2. Veronesi, U.; Viale, G.; Paganelli, G.; Zurrida, S.; Luini, A.; Galimberti, V.; Veronesi, P.; Intra, M.; Maisonneuve, P.; Zucca, F.; et al. Sentinel lymph node biopsy in breast cancer: Ten-year results of a randomized controlled study. Ann. Surg. 2010, 251, 595–600, doi:10.1097/SLA.0b013e3181c0e92a.

3. Bejnordi, B.E.; Litjens, G.; Timofeeva, N.; Otte-Höller, I.; Homeyer, A.; Karssemeijer, N.; van der Laak, J.A.W.M. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017, 318, 2199–2210, doi:10.1001/jama.2017.14585.

4. Bándi, P.; Geessink, O.; Manson, Q.; van Dijk, M.; Balkenhol, M.; Hermsen, M.; Bejnordi, B.E.; Lee, B.; Paeng, K.; Zhong, A.; et al. From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. IEEE Trans. Med. Imaging 2019, 38, 550–560, doi:10.1109/TMI.2018.2867350.

5. Liu, Y.; Kohlberger, T.; Norouzi, M.; Dahl, G.E.; Smith, J.L.; Mohtashamian, A.; Olson, N.; Peng, L.H.; Hipp, J.D.; Stumpe, M.C. Artificial intelligence–based breast cancer nodal metastasis detection insights into the black box for pathologists. Arch. Pathol. Lab. Med. 2019, 143, 859–868, doi:10.5858/arpa.2018-0147-OA.

6. Komura, D.; Ishikawa, S. Machine Learning Methods for Histopathological Image Analysis. Comput. Struct. Biotechnol. J. 2018, 16, 34–42, doi:10.1016/j.csbj.2018.01.001.

7. Vesal, S.; Ravikumar, N.; Davari, A.; Ellmann, S.; Maier, A. Classification of breast cancer histology images using transfer learning. In ICIAR 2018, LNCS 10882; Campilho, A., Ed.; Springer: Cham, Switzerland, 2018; pp. 812–819.

8. Griffin, J.; Treanor, D. Digital pathology in clinical use: Where are we now and what is holding us back? Histopathology 2017, 70, 134–145, doi:10.1111/his.12993.

9. Ghaznavi, F.; Evans, A.; Madabhushi, A.; Feldman, M. Digital imaging in pathology: Whole-slide imaging and beyond. Annu. Rev. Pathol. 2013, 8, 331–359, doi:10.1146/annurev-pathol-011811-120902.

10. Steiner, D.F.; MacDonald, R.; Liu, Y.; Truszkowski, P.; Hipp, J.D.; Gammage, C.; Thng, F.; Peng, L.; Stumpe, M.C. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. Am. J. Surg. Pathol. 2018, 42, 1636–1646, doi:10.1097/PAS.0000000000001151.

11. Madabhushi, A.; Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med. Image Anal. 2016, 33, 170–175, doi:10.1016/j.media.2016.06.037.

12. Veeling, B.S.; Linmans, J.; Winkens, J.; Cohen, T.; Welling, M. Rotation Equivariant CNNs for Digital Pathology. In MICCAI 2018 LNCS 11071; Frangi, A.F., Ed.; Springer: Cham, Switzerland, 2018; pp. 210–218.

13. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255, doi:10.1109/CVPR.2009.5206848.

14. Ahmad, H.M.; Ghuffar, S.; Khurshid, K. Classification of breast cancer histology images using transfer learning. In Proceedings of the 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST 2019), Islamabad, Pakistan, 8–12 January 2019; pp. 328–332, doi:10.1109/IBCAST.2019.8667221.

15. Nishio, M.; Sugiyama, O.; Yakami, M.; Ueno, S.; Kubo, T.; Kuroda, T.; Togashi, K. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. PLoS ONE 2018, 13, e0200721, doi:10.1371/journal.pone.0200721.

16. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of ICLR 2015, San Diego, CA, USA, 7–9 May 2015.

17. Janowczyk, A.; Basavanhally, A.; Madabhushi, A. Stain Normalization using Sparse AutoEncoders (StaNoSA): Application to digital pathology. Comput. Med. Imaging Graph. 2017, 57, 50–61, doi:10.1016/j.compmedimag.2016.05.003.

18. Tellez, D.; Litjens, G.; Bándi, P.; Bulten, W.; Bokhorst, J.; Ciompi, F.; van der Laak, J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med. Image Anal. 2019, 58, 101544, doi:10.1016/j.media.2019.101544.

19. Bejnordi, B.E.; Litjens, G.; Timofeeva, N.; Otte-Höller, I.; Homeyer, A.; Karssemeijer, N.; van der Laak, J.A.W.M. Stain specific standardization of whole-slide histopathological images. IEEE Trans. Med. Imaging 2016, 35, 404–415, doi:10.1109/TMI.2015.2476509.

20. Naik, S.; Doyle, S.; Feldman, M.; Tomaszewski, J.; Madabhushi, A. Gland segmentation in prostate histopathological images. J. Med. Imaging 2017, 4, 027501, doi:10.1117/1.jmi.4.2.027501.

21. Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological Image Analysis: A Review. IEEE Rev. Biomed. Eng. 2009, 2, 147–171, doi:10.1109/RBME.2009.2034865.

22. Valkonen, M.; Kartasalo, K.; Liimatainen, K.; Nykter, M.; Latonen, L.; Ruusuvuori, P. Metastasis detection from whole slide images using local features and random forests. Cytometry A 2017, 91, 555–565, doi:10.1002/cyto.a.23089.

23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In MICCAI 2015, Part III, LNCS 9351; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241, doi:10.1007/978-3-319-24574-4_28.

24. Janowczyk, A.; Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J. Pathol. Inform. 2016, 7, 29, doi:10.4103/2153-3539.186902.

25. Shrestha, A.; Mahmood, A. Review of deep learning algorithms and architectures. IEEE Access 2019, 7, 53040–53065, doi:10.1109/ACCESS.2019.2912200.

26. Albarqouni, S.; Baur, C.; Achilles, F.; Belagiannis, V.; Demirci, S.; Navab, N. AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans. Med. Imaging 2016, 35, 1313–1321, doi:10.1109/TMI.2016.2528120.

27. Menden, M.P.; Wang, D.; Mason, M.J.; Szalai, B.; Bulusu, K.C.; Guan, Y.; Yu, T.; Kang, J.; Jeon, M.; Wolfinger, R.; et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat. Commun. 2019, 10, 2674, doi:10.1038/s41467-019-09799-2.

28. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic Attribution for Deep Networks. In Proceedings of Machine Learning Research, ICML 2017, Sydney, Australia, 6–11 August 2017; Precup, D., Teh, Y.W., Eds.; 2017; Volume 70, pp. 3319–3328.

29. Peck, M.; Moffat, D.; Latham, B.; Badrick, T. Review of diagnostic error in anatomical pathology and the role and value of second opinions in error prevention. J. Clin. Pathol. 2018, 71, 995–1000, doi:10.1136/jclinpath-2018-205226.

30. Renshaw, A.A.; Gould, E.W. Reducing false-negative and false-positive diagnoses in anatomic pathology consultation material. Arch. Pathol. Lab. Med. 2013, 137, 1770–1773, doi:10.5858/arpa.2013-0012-OA.

31. Mormont, R.; Geurts, P.; Maree, R. Comparison of deep transfer learning strategies for digital pathology. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; Volume 2018, pp. 2343–2352, doi:10.1109/CVPRW.2018.00303.

32. Sirinukunwattana, K.; Raza, S.E.A.; Tsang, Y.; Snead, D.R.J.; Cree, I.A.; Rajpoot, N.M. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. IEEE Trans. Med. Imaging 2016, 35, 1196–1206, doi:10.1109/TMI.2016.2525803.