Breast Cancer Nuclear Atypia Scoring Using Convolutional Neural Networks

Shana Beniamin*
sbeniamin@.ryerson.ca
Ryerson University, Toronto, Canada.

***Abstract*—Nuclear atypia scoring for breast cancer is a tedious process for pathologists and prone to high inter-observer variation. While the number of pathologists is decreasing, the number of cancer cases is increasing. Thus, augmenting the pathologist workflow by automating the nuclear atypia scoring process is a research area of much interest. The literature review conducted in this paper indicates a shift from hand crafted feature based scoring algorithms to deep learning based scoring algorithms in recent years. The use of convolutional neural networks (CNNs) has been proven to increase the classification accuracy. Therefore, in this work, we will train a CNN to perform nuclear atypia scoring on the publically available MITOS-ATYPIA-14 Challenge dataset.**

***Index Terms*—Breast cancer, convolutional neural networks, nuclear atypia grading, digital pathology.**

I. Introduction

Breast cancer is the most common type of cancer in women; making up 25 % of cancer cases in women worldwide [2]. While non-invasive tests (ultrasounds, mammography, and MRI) are used for diagnosis, histopathological analysis of a breast biopsy specimen is the gold standard. Not only is the analysis of multiple tissue slides is a labor intensive process for pathologists, it is also prone to high diagnostic inter-observer variability between pathologists. Recent advancements in scanning technology enabled the development of computer aided diagnosis algorithms. These augment the pathologists' workflow and can improve the accuracy, efficiency and reliability of the breast cancer evaluation. The Nottingham Grading System (NGS), which is used for grading breast cancer, considers three characteristics (tubule formation, nuclear atypia, mitotic count) of a tissue and assigns each a score of 1, 2 or 3.

To determine the nuclear atypia score (NAS), a pathologist will examine the size and outline of nuclei as well as the distribution of chromatin. Smaller nuclei with a regular outline and a uniform chromatin distribution will obtain a score of 1. The more the nuclei stray from this description (irregularities in size, shape, number of nucleoli), the score increases (NAS2 or NAS3). Many existing nuclear atypia scoring techniques use handcrafted feature-based algorithms which are highly dependent on accurate nuclei segmentation. However, most regent algorithms utilize deep learning to extract high-level abstractions from the histopathological images, and have been shown to outperform the systems based on handcrafted feature descriptors [2].

In this work, we aim to automate the process of predicting the NAS by using convolutional neural networks (CNNs). We will train and test the CNN using the open source MITOS ATYPIA 14 dataset: https://mitos-atypia-14.grand-challenge.org/ [1].

II. Literature Review

Lu et al. [3] presents an automated technique for nuclear atypia scoring using a SVM classifier. A group of 142 textural and morphological features were extracted from segmented nuclei regions. The histogram is computed for each feature and they are combined to form a feature vector that is the input of the classifier. They evaluated their algorithm on the MITOS-ATYPIA-14 challenge dataset and were able to differentiate high power field images between NAS1 versus non-NAS1, NAS2 versus non-NAS2 and NAS3 versus non-NAS3; yielding area under receiver-operating characteristic curve (AUC) of 0.90, 0.86 and 0.87, respectively. For ternary classification, they obtained an average classification accuracy of 78.79%.

Doyle et al. [4] presents methodology for automatically performing binary classification of breast cancer grade (low vs. high) using a SVM. Over 3400 features are extracted from a dataset of 48 breast biopsy tissue slides; feature set dimensionality is reduced through spectral clustering. Different subsets of these features are used to train and test the classifier and their performance is compared. The highest classification accuracy of 93.3% was achieved using all the architectural features (Voronoi, Delaunay, MST, and nuclear features). While this accuracy is quite high, it is only performing binary classification. For clinical applications, the algorithm should distinguish between the three grades (low, intermediate, high) and this task is far more challenging.

Gandomkar et al. [5] employs a hybrid segmentation based and texture-based method to extract the necessary features for nuclear atypia scoring. An ensemble of trees (regression) is implemented. The first regression model is trained on six atypia related criteria (nuclei size, nucleoli size, anisonucleosis, chromatin density, regularity of nuclear contour, and membrane thickness) that the pathologists recorded. The second regression model is trained on textural

features. A third regression model was trained using the outputs from the other two regression models, and it was utilized to predict a final atypia grade. This algorithm's performance was evaluated on 300 images from the MITOS-ATYPIA-14 Challenge. When compared to 3 different pathologists, it achieved 93.8%, 92.9% and 93.1% agreement.

Wan et al. [6] used pixel-based, object-based, and CNN-derived semantic-level descriptors to distinguish between low, intermediate, and high breast cancer grades. The dataset had 106 hematoxylin and eosin stained breast cancer tissue images (24 low, 62 intermediate, and 20 high grade). The histopathological images are first color normalized, then the nuclei are segmented, and the features extracted from the segmented nuclei are used to train cascaded support vector machine (SVM) classifiers. The pixel-based descriptors use the image pixels in the segmented nuclei to quantify image sharpness, contrast, changes in intensity, and discontinuities. Pixel based descriptors are composed of texture features (Kirsch filters, Gabor filters, first-order and Haralick features), histogram of oriented gradients (HoG) features, and local binary pattern (LBP) features. Object-level features capture the spatial distribution of the nuclei. The topological features are extracted using the Voronoi diagram (VD), Delaunay triangulation (DT), and minimum spanning tree (MST). Semantic-level features (obtained using a CNN) capture interpretable high-level concepts (i.e. presence/absence of nucleoli, necrosis, and lymphocytes). The 3-layer CNN model is trained on segmented nuclei labeled as low, intermediate, or high grade. It is comprised of two successive convolutional (9 x 9 kernel) and max pooling (2 x 2 kernel) layers, followed by a fully connected classification layer (38 neurons form the final layer are connected to output 3 neurons; each corresponding to a grade). This quantifies the proportions of nuclei belonging to the different cancer grades. The scores from each individual SVM classifier were combined to obtain the final cancer grade prediction. This approach is computationally efficient and achieves an accuracy of 0.92, 0.77, and 0.76 for binary classification of low versus high, low verses intermediate, and intermediate versus high grades respectively. Ternary classification (low, intermediate, high) yields accuracy of 0.69.

Xu et al. [7] presented a Multi-Resolution Convolution Network (MR-CN) with Plurality Voting (MR-CN-PV) to perform automated nuclear atypia scoring. The MR-CN consists of three single resolution convolutional networks (based on AlexNet) which consider three different image resolutions (x10, x20, x40 magnifications) to predict the nuclear atypia score. These three scores are combined using a plurality voting strategy to obtain a final atypia score prediction. This approach was evaluated on the publically available MITOS-ATYPIA-14 challenge dataset. Each single resolution convolutional network consists of 5 convolutional (ranging from 11 x 11 to 3 x 3 kernels) and 3 max pooling (3 x 3 kernel) layers. A Rectified Linear Unit (ReLU) function was used for activation. This approach achieved 2nd place results in the MITOS-ATYPIA-14 Challenge. It integrates image features from multiple field of views but has the downside of having to train multiple models (one for each field of view) to achieve this.

Maqlin et al. [8] proposed a deep belief network (DBN-DNN) for nuclear atypia scoring. Restricted Boltzmann Machines (RBMs) are generative neural networks which learn the probability distribution of the input data and were incorporated in each layer of the DBN-DNN. Each one has 13 hidden layers, and they are stacked together to form a beep belief network and then a final layer is added to form a deep neural network. The DBN-DNN is fine-tuned using a classical back-propagation algorithm. From the MITOS-ATYPIA-14 Challenge, 80 breast cancer images are segmented and 20 features are extracted to train the DBN-DNN classifier to predict the NAS. It obtained in accuracy of 90% for ternary classification. While this method is very fast and accurate, it relies on features extracted from segmented nuclei, and segmenting nuclei accurately is a challenging task.

Jafarbiglo et al. [9] propose a CNN based NAS classifier which contains three convolutional layers; each one followed by local response normalization, ReLU and a 2x2 max pooling layer. The first, second, and third convolutional layers utilize 16, 32, and 52 5x5 pixel filters, respectively. Two fully connected layers are incorporated after the convolutional layers; the first contains 64 neurons and uses ReLU activation, while the second contains 3 neurons and uses a softmax function to compute the probability of each of the three classes. They employ data augmentation (180 degree rotation, and flipping) and pre-processing (convert RGB to LAB color space) to better differential colors. This approach is evaluated on the MITOS-ATYPIA-14 Challenge dataset and was found to achieve an accuracy of 84.23% for ternary classification.

Golatkar et al. [10] proposed a deep learning-based approach to classify HE stained breast tissue in into four classes (normal tissue, benign lesion, in situ carcinoma, and invasive carcinoma). While this is not the same as nuclear atypia scoring, there may be useful techniques that can translate to our desired task. They utilize images from the BACH challenge and fine-tune an Inception-v2 CNN (pre-trained on the ImageNet dataset) for classification. By only extracting patches with high nuclear density, they are able to discard uninformative regions. Majority voting is used to determine the final image class. They replaced the fully connected layer from the top of the network with a global average pooling layer, fully connected layer with 1024 neurons and a softmax classifier (4 neurons). The two stage training process involved (1) freezing the convolutional layers to only train newly added layers using an RMSProp optimizer for 25 epochs and (2) fine tuning the last two inception blocks and newly added layers using an SGD optimizer with a learning rate of 0.0001 for 50 epochs. Their approach had an average accuracy of 85% over the four classes and 93% when simply distinguishing between caner and non-cancer tissues.

Motlagh et al. [11] classified breast cancer sub-types using 7909 images from the BreakHis database using deep learning framework. Fine-tuning CNNs and applying transfer learning causes faster convergence and obtains better results than training from scratch. Thus, they initialized the weights of different layers of their network using the Inception and ResNet models pre-trained on the very large ImageNet dataset. The malignant subtype (ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma) classification of the ResNet V1 152 and ResNet V1 50 had accuracies of 96.4% and 94.6%, respectively, when all layers were fine tuned. ResNet V1 152 could achieve an accuracy of 90% when only the last-layer was fine tuned. ResNet models were also found to outperform the Inception models. The models were trained on 85% of the dataset and tested using the remaining 15%.

Han et al. [12] also classified breast cancer subtypes using a newly proposed deep learning model; a Class Structure based Deep CNN (CSDCNN). It learns discriminative and sematic hierarchical features and utilizes feature space distance constraints to specify feature similarities of different classes. The model is very deep and can learn high-level features which help to discriminate between classes. The CSDCNN model was based on the GoogLeNet, is 22 layers deep (convolutional and maxpooling layers, and the last layer uses ReLU activation) and was trained for 5000 iterations using a learning rate of 0.01. They evaluated their model on the BreakHis dataset achieved 93.2% accuracy for their multi-classification method.

III. Problem Statement

Current trends suggest a shortage of pathologists in the US and the diagnostic workload for each pathologist has risen by 41.7%. Shortages result in suboptimal patient care; delayed cancer diagnoses and diagnostic errors [13]. In addition to diagnostic errors due to the shortages, breast cancer diagnosis is subjective in nature and has poor reproducibility [13]. Automating the nuclear atypia scoring process will help to reduce the pathologist workload and improve diagnostic reliability [2].

Most of the existing automation techniques require a preliminary step of accurately segmenting the nuclei in breast cancer tissue digital pathology images. Then, handcrafted features (morphology and texture based) are extracted from the nuclei and used in conjunction with machine learning classifiers to classify the nuclear atypia score. However, segmenting nuclei accurately is very challenging and any errors in this step can negatively affect the scoring process. Furthermore, there may be other useful information in the images that is not represented by the hand-crafted features. This may explain why recent deep learning based methods are achieving better classification accuracy than the hand crafted feature based approaches [2].

IV. References

[1] "MITOS-ATYPIA-14 - Grand Challenge," *Grand Challenge*. [Online]. Available: https://mitos-atypia-14.grand-challenge.org/. [Accessed: 21-Jun-2021].

[2] A. Das, M. S. Nair, and S. D. Peter, "Computer-Aided Histopathological Image Analysis Techniques for Automated Nuclear Atypia Scoring of Breast Cancer: a Review," *Journal of Digital Imaging*, vol. 33, no. 5, pp. 1091–1121, 2020.

[3] C. Lu, M. Ji, Z. Ma, and M. Mandal, "Automated image analysis of nuclear atypia in high-power field histopathological image," *Journal of Microscopy*, vol. 258, no. 3, pp. 233–240, 2015.

[4] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clusteringwith textural and architectural image features," *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2008.

[5] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, "Computer-Assisted Nuclear Atypia Scoring of Breast Cancer: a Preliminary Study," *Journal of Digital Imaging*, vol. 32, no. 5, pp. 702–712, 2019.

[6] T. Wan, J. Cao, J. Chen, and Z. Qin, "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," *Neurocomputing*, vol. 229, pp. 34–44, 2017.

[7] J. Xu, C. Zhou, B. Lang, and Q. Liu, "Deep Learning for Histopathological Image Analysis: Towards Computerized Diagnosis on Cancers," *Deep Learning and Convolutional Neural Networks for Medical Image Computing*, pp. 73–95, 2017.

[8] P. Maqlin, R. Thamburaj, J. J. Mammen, and M. T. Manipadam, "Automated Nuclear Pleomorphism Scoring in Breast Cancer Histopathology Images Using Deep Neural Networks," *Mining Intelligence and Knowledge Exploration*, pp. 269–276, 2015.

[9] S. K. Jafarbiglo, H. Danyali, and M. S. Helfroush, "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.

[10] A. Golatkar, D. Anand, and A. Sethi, "Classification of Breast Cancer Histology Using Deep Learning," *Lecture Notes in Computer Science*, pp. 837–844, 2018.

[11] M. H. Motlagh, M. Jannesari, H. R. Aboulkheyr, P. Khosravi, O. Elemento, M. Totonchi, and I. Hajirasouliha, "Breast Cancer Histopathological Image Classification:

A Deep Learning Approach," *018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 2405-2412, 2018.

[12] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Scientific Reports*, vol. 7, no. 1, 2017.

[13] D. M. Metter, T. J. Colgan, S. T. Leung, C. F. Timmons, and J. Y. Park, "Trends in the US and Canadian Pathologist Workforces From 2007 to 2017," *JAMA Network Open*, vol. 2, no. 5, 2019.