DAN : Breast Cancer Classification from High Resolution Histology Images using *D*eep *A*ttention *N*etwork

Ritabrata Sanyal[1], Manan Jethanandani[2], Ram Sarkar[3]

[1] Kalyani Government Engineering College, West Bengal, India,
sritabrata@gmail.com
[2] LNM Institute of Information Technology, Rajasthan, India
mananjethanandani01@gmail.com
[3] Jadavpur University, West Bengal, India
ramjucse@gmail.com

Abstract. Millions of women succumb to breast cancer every year. Till date it is mainly diagnosed by core needle biopsy of the breast tissue, followed by analysis of the histopathological image to detect the presence of malignant tumor. In the past few years, deep learning pipelines have been proposed for carcinoma type classification from the breast histology images. They mostly entail in dividing the high resolution images into patches, followed by classifying the patches using Convolutional Neural Network, and finally integrating the patch wise results to predict the class of the image. But these methods give the same importance to all the patches, and do not focus on the most salient regions of the image. In this paper, we present a novel attention mechanism, which aids the network to specifically focus on the most relevant parts of the image, that is the design of the network allows for learning a weighted representation of all the constituent patches of an image. Experimental results reveal that our model achieved a 85.50% and 96.25% for patch wise and image wise classification accuracies respectively on the ICIAR 2018 breast histopathological images dataset. Our proposed method outperforms some state-of-the-art methods to the best of our knowledge.

Keywords: Breast cancer, Histology imaging, Deep learning, Attention network, CNN, BLSTM

1 Introduction

Breast cancer accounts for one of the major causes behind death of women worldwide. One of the most commonly used diagnostic technique is the microscopic analysis of Hematoxylin and Eosin(H & E) stained breast histopathological images. However, manual analysis suffers from various drawbacks due to the complexity of the histopathological images, and thus the process often becomes time consuming and the results are contingent on the pathologist's subjectivity. Hence it is necessary to develop highly accurate, automated breast

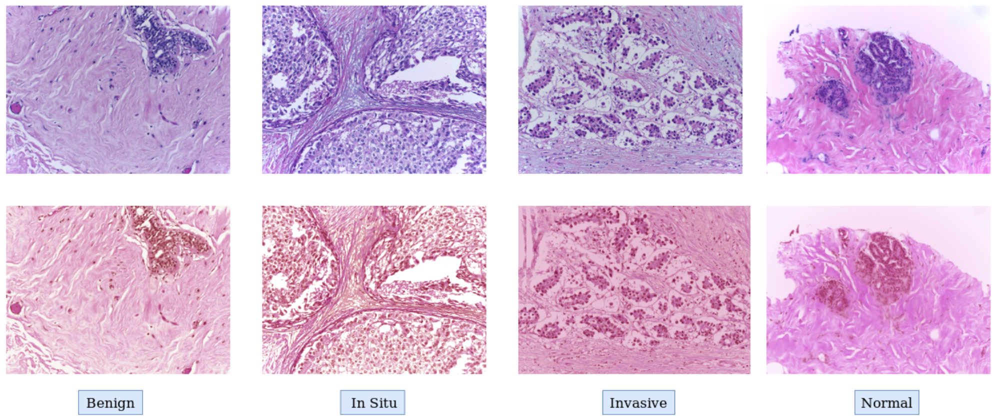

Fig. 1: Sample high resolution images taken from the ICIAR-2018 dataset are shown in the first row, and their corresponding stain normalized versions are shown in the second row.

histopathology image analysis methods. The recent success of Deep Learning based models in various computer vision problems has inspired researches to apply those techniques to detect carcinoma from high resolution breast histology images [1,3,7,11]. Most of the methods proposed in the literature, follow a basic outline—at first, the high resolution images are divided into patches which is followed by classifying the patches by extracting features from them with using CNN, finally the patch wise results or features are integrated using any conventional classifier like Support Vector Machine (SVM) to predict the class of the image. Since these methods do not consider the contextual relationship between neighbouring patches, Yan et al. [12] have proposed a Bi-directional Long Short Term Memory (BLSTM) model to integrate the patch wise features for image wise prediction. This retains the contextual information of the patches, but suffers from one limitation—their model does not focus on the most relevant regions of the high resolution image. This is important because not all regions of an image may contribute equally to predict the final outcome. In order to aid the network to *attend* some specific relevant spatial regions of an image, we propose a novel Deep Attention Network. Our proposed network learns a function which weighs the contribution of the constituent patches for predicting the class of the image. To the best of our knowledge, this is the first work which uses attention mechanism in the context of breast histology image classification.

2 Dataset

We use the Breast Cancer Histology Challenge (BACH) 2018 dataset [2], hosted as part of the ICIAR 2018 conference, to train and test our model. The dataset consists of high resolution 2048×1536 breast histology images, with each pixel covering 0.42 µm x 0.42 µm of breast tissue area. The images are labelled into

4 different categories namely, Benign, Normal, In Situ carcinoma and Invasive carcinoma, with 100 images per category.

3 Methodologies

Our proposed system constitutes of four components namely, i) Preprocessing, ii) Patch extraction, iii) Patch wise feature extraction, and iv) Image level with attention mechanism . The details of each component are elaborated in the subsequent sections.

3.1 Preprocessing

Preprocessing the histology images is of paramount importance for development of an accurate prediction model. Hematoxylin and Eosin (H & E) stained images are susceptible to unwanted color shifts, mainly due to the variations in slide scanners and level of stain absorption. These color shifts may introduce a bias in our model preventing it from generalizing well. Thus, to reduce these variations, the images are standardized before feeding them to the model using a process called—- stain normalization. In our work, Vahadane normalization [10], a stain normalization method based of sparse non-negative matrix factorization (SNMF) is used.

3.2 Patch extraction

CNN models trained on high resolution (2048×1536) images suffer from overfitting, and the paucity of images in our dataset exacerbates the issue. Hence the preprocessed images are divided into patches before feeding them to a CNN. The patches are extracted in a sliding window fashion——a 512×512 window, with a stride of 2 that is with 50% overlap, slides over the image, resulting in 35 patches per image. We empirically observed that models trained with these settings to be giving the best results.

3.3 Patch wise feature extraction

We fine-tune a pre-trained CNN network for patch wise classification and feature extraction. We do not train the network from scratch as it would require more data, compared to a fine-tuning process through which only the higher level features are learned. A state-of-the-art CNN architecture namely VGG19 [9], pre-trained on the ImageNet dataset, is fine-tuned on the breast histology patch dataset. The top fully connected layers of the VGG19 are replaced with a Global Average Pooling (GAP) layer [5] to allow the network to consume images of any arbitrary dimensions. A GAP layer takes a $3 - D$ feature map as input, and outputs a $1 - D$ vector of length equal to the number of channels (depth) of the map. For feature extraction, we first apply the GAP operation on 7 internal convolutional layers, namely the 4^{th}, 6^{th}, 8^{th}, 10^{th}, 12^{th}, 14^{th}, 16^{th} conv. layers with 128, 256, 256 and the rest with 512 number of channels respectively. The outputs of all the GAP operations are then concatenated to form a vector of length 2688, which is thus used as the feature representation of a patch. This type of hierarchical feature extraction is done to retain richer, fine multilevel features like local textures, edges, corners, intensity variations which are lost in the higher convolutional layers.

3.4 Image level classification

For classifying a high resolution image, we first extract the feature vectors of the constituent patches, as described earlier. The 35 feature vectors are then fed to a BLSTM network, one at each timestep, to retain long term and short term correlations between the patches and to incorporate contextual information between neighbouring patches. Essentially, BLSTM functions as an encoder network and the hidden state vector at every BLSTM timestep contains the encoded information of the corresponding patch feature. At every timestep, an Attention Unit (AU) takes in the hidden state of the BLSTM and the corresponding input patch feature vector, for computing a weight that is to be assigned to that particular patch. The weight signifies the importance of that patch for prediction of the image class. Finally, the AU generates a weighted sum of all the hidden BLSTM state vectors, which is then fed to a Multi Layer Perceptron (MLP) decoder and finally a Softmax layer to output the posterior probabilities of the 4 image classes.

More formally, we extract a set of feature vectors, $\psi = \{x_i; i = 1, 2, ..., 35\}$ from the 2048×1536 image, where x_i is the feature vector of the i^{th} patch. The hidden state at t^{th} timestep of the BLSTM encoder, h_t is formally expressed as,

$$\overrightarrow{h_t}, \overleftarrow{h_t} = BLSTM(\overrightarrow{h_{t-1}}, \overleftarrow{h_{t-1}}, x_t) \tag{1}$$

$$h_t = Concat(\overrightarrow{h_t}, \overleftarrow{h_t}) \tag{2}$$

where $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are the hidden states of the left to right and right to left LSTMs respectively, and $t = 1, 2, ...35$. The AU assigns a positive weight ϕ_t to the hidden state at every timestep, h_t and generates its weighted sum r where,

$$r = \sum_{t=1}^{35} \phi_t h_t \tag{3}$$

where $\phi \in R^{35}$ is a vector of attention weights such that $\sum_t \phi_t = 1$. Thus r represents the relative significance of every patch of the image, used by the decoder to predict the class. The attention weights are calculated as follows,

$$\beta_t = v^T tanh(W_h h_t + W_x x_t + b) \tag{4}$$

$$\phi_t = \frac{exp(\beta_t)}{\sum_{t=1}^{35} exp(\beta_t)} \tag{5}$$

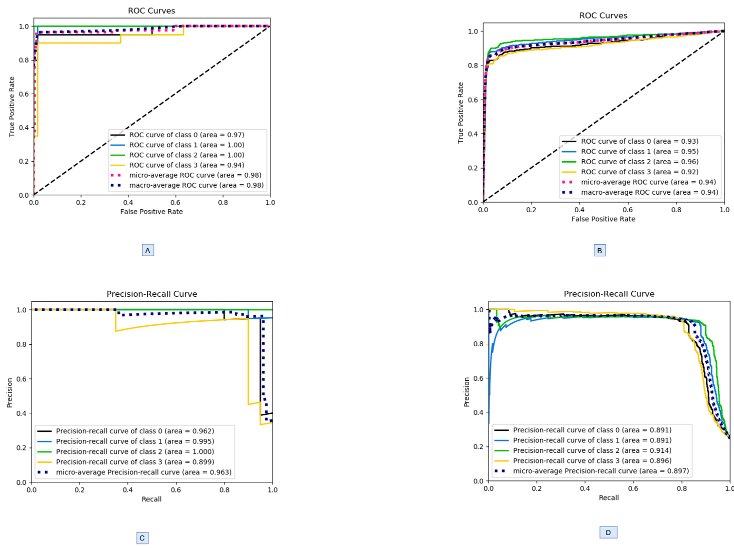

Fig. 2: Precision-Recall Curves for a)Image Wise and b)Patch-Wise classification and ROC curves for c)Image Wise and d)Patch-Wise classification for the four classes.0:Benign, 1:Insitu, 2:Invasive, 3:Normal

where, v, W_h, W_x, b are all trainable parameters. The MLP decoder takes in the weighted vector sum, r as input and outputs the softmax prediction of the image $\hat{y}$, as follows,

$$r^* = tanh(W_r r + b_r) \tag{6}$$

$$\hat{y} = softmax(W_s r^* + b_s) \tag{7}$$

where W_r, W_s, b_r, b_s are trainable parameters.

4 Experiments

Conforming to the standard practice, ICIAR 2018 [2] dataset, used here for experimentation, is split into three parts and used for training, validation and testing purposes. 70%, 10% and 20% of the images are used for training, validation and testing purposes respectively. Data augmentation is done to artificially boost number of training samples. The patch wise and image wise networks are trained with mini batch Stochastic Gradient Descent and Adam [4] optimizers respectively. We evaluate the classification performance of the model using metrics like accuracy, sensitivity, specificity, precision. Apart from 4-class classification, 2-class carcinoma vs non-carcinoma classification accuracies are also reported, where benign, normal and insitu, invasive classes are merged into non-carcinoma (benign and normal) and carcinoma(insitu and invasive) classes respectively. Our model achieves 4-class patch and image wise accuracies of 85.50% and 96.25% respectively, whereas 2-class patch and image wise accuracies are 96% and 98.75% respectively. From Table 1, we can see that the BLSTM model with our proposed AU is superior to the standard BLSTM as used in recent work by Yan et al. [12], by 2.50%. This demonstrates the efficacy of our proposed attention mechanism. From Table 3, we can see that our proposed model outperforms other state-of-the-art breast histology image classification techniques by a large margin, both for 4-class and 2-class scenarios.

Table 1: Comparison of image level accuracies with and without using attention mechanism. For fair comparison, the patch level feature extraction strategy is kept same in both cases (fine tuning VGG19). BLSTM without Attention model is similar to that proposed by Yan et al. [12].

Model	**Accuracy (%)**
BLSTM without Attention	93.75
BLSTM with Attention (Proposed)	96.25

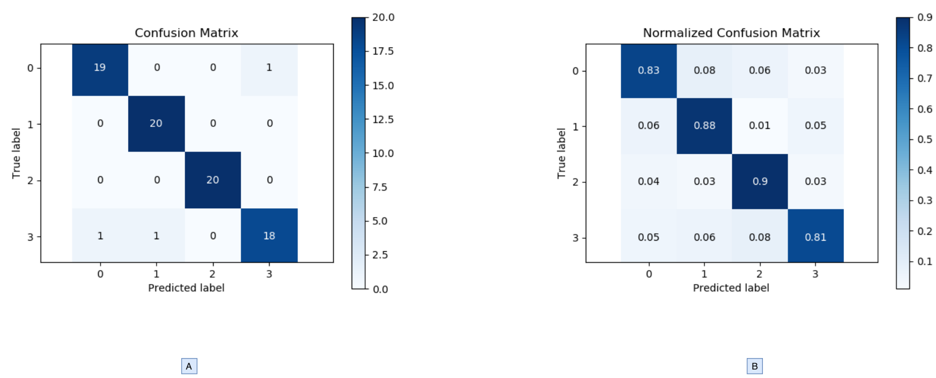

Fig. 3: Confusion Matrix for a)image wise and b)patch-wise classification. 0:Benign, 1:Insitu, 2:Invasive, 3:Normal

5 Conclusion

In this paper, we have proposed a novel attention mechanism for classification of high resolution breast histology images. The attention model focuses on the most relevant regions of the image for predicting its class. We first extract features from all the constituent patches of an image and feed them to a BLSTM encoder

Table 2: Image Wise and Patch Wise metrics of 4 class classification

Metric Type	Metrics	Classes				Avg
		Benign	InSitu	Invasive	Normal	
Patch Wise	**Precision**	84.94	84.38	85.25	87.63	85.55
	Sensitivity	83	88	90	81	85.50
	Specificity	95.10	94.57	94.81	96.19	95.17
	F1 Score	83.95	86.15	87.56	84.18	85.46
Image Wise	**Precision**	95	95.24	100	94.74	96.24
	Sensitivity	95	100	100	90	96.25
	Specificity	98.33	98.33	100	98.33	98.75
	F1 Score	95	97.56	100	92.30	96.20

Table 3: Comparision of proposed method with some state of the art methods

Method	Dataset Used	Number of Classes	Patch Wise Accuracy(%)	Image Wise Accuracy(%)
Araujo et al. [1]	Bioimaging 2015	4	66.7	77.8
Araujo et al. [1]	Bioimaging 2015	2	77.6	83.3
Rakhlin et al. [7]	ICIAR 2018	4	-	87.5
Rakhlin et al. [7]	ICIAR 2018	2	-	93.8
Vang et al. [11]	ICIAR 2018	4	-	87.5
Golatkar et al. [3]	ICIAR 2018	4	79	85
Golatkar et al. [3]	ICIAR 2018	2	-	93
Roy et al. [8]	ICIAR 2018	4	77.4	90
Roy et al. [8]	ICIAR 2018	2	84.7	92.5
Nazeri et al. [6]	ICIAR 2018	4	-	95
Yan et al. [12]	ICIAR 2018	4	82.10	91.30
Proposed Method	ICIAR 2018	4	85.50	96.25
Proposed Method	ICIAR 2018	2	96	98.75

network. The AU lies on top of the BLSTM network, and learns a weighted representation of the hidden states of the same. This weighted representation captures the relative importance that our model gives to every constituent patch of the image. It is then fed to an MLP decoder which then predicts the class of the image. Experimental results show that our proposed model outperforms other state-of-the-art methods. Although our model proved to be superior to the previously proposed models in terms of classification accuracy, we believe it is still limited due to the dearth of training data. Hence, in future, we would attempt to use Generative Adverserial Network (GAN) based approaches, where we can generate synthetic images for robust data augmentation.

References

1. Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., Campilho, A.: Classification of breast cancer histology images using convolutional neural networks. PloS one **12**(6), e0177544 (2017)
2. Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al.: Bach: Grand challenge on breast cancer histology images. Medical image analysis (2019)
3. Golatkar, A., Anand, D., Sethi, A.: Classification of breast cancer histology using deep learning. In: International Conference Image Analysis and Recognition. pp. 837–844. Springer (2018)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv preprint arXiv:1312.4400 (2013)
6. Nazeri, K., Aminpour, A., Ebrahimi, M.: Two-stage convolutional neural network for breast cancer histology image classification. In: International Conference Image Analysis and Recognition. pp. 717–726. Springer (2018)
7. Rakhlin, A., Shvets, A., Iglovikov, V., Kalinin, A.A.: Deep convolutional neural networks for breast cancer histology image analysis. In: International Conference Image Analysis and Recognition. pp. 737–744. Springer (2018)
8. Roy, K., Banik, D., Bhattacharjee, D., Nasipuri, M.: Patch-based system for classification of breast histology images using deep learning. Computerized Medical Imaging and Graphics **71**, 90–103 (2019)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
10. Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE transactions on medical imaging **35**(8), 1962–1971 (2016)
11. Vang, Y.S., Chen, Z., Xie, X.: Deep learning framework for multi-class breast cancer histology image classification. In: International Conference Image Analysis and Recognition. pp. 914–922. Springer (2018)
12. Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., Rao, X., Zheng, C., Zhang, F.: Breast cancer histopathological image classification using a hybrid deep neural network. Methods (2019)