MS-GWNN: MULTI-SCALE GRAPH WAVELET NEURAL NETWORK FOR BREAST CANCER DIAGNOSIS

Mo Zhang[1,2,3], *Bin Dong*[4,1], *Quanzheng Li*[5]

[1]Peking University, Center for Data Science, China;
[2]Peking University, Center for Data Science in Health and Medicine, China;
[3]Beijing Institute of Big Data Research, Laboratory for Biomedical Image Analysis, China;
[4]Peking University, Beijing International Center for Mathematical Research (BICMR), China;
[5]Harvard Medical School, Massachusetts General Hospital, MGH/BWH Center for Clinical Data Science, Center for Advanced Medical Computing and Analysis, Department of Radiology, USA.

ABSTRACT

Breast cancer is one of the most common cancers worldwide, and early detection can significantly reduce its mortality rate. It is crucial to take multi-scale information of tissue structure into account in the detection of breast cancer. And thus, it is the key to design an accurate computer-aided detection (CAD) system to capture multi-scale contextual features in a cancerous tissue. In this work, we present a novel graph convolutional neural network for histopathological image classification of breast cancer. The new method, named multi-scale graph wavelet neural network (MS-GWNN), leverages the localization property of spectral graph wavelet to perform multi-scale analysis. By aggregating features at different scales, MS-GWNN can encode the multi-scale contextual interactions in the whole pathological slide. Experimental results on two public datasets demonstrate the superiority of MS-GWNN. Moreover, ablation studies show that multi-scale analysis has a significant impact on the accuracy of cancer diagnosis.

***Index Terms*—** Graph wavelet neural network, breast cancer diagnosis, histopathological image.

1. INTRODUCTION

Breast cancer is the second leading cause of cancer-related death among women [1]. Effective treatment of breast cancer heavily relies on the accurate diagnosis at an early stage. In clinical practice, histopathological image analysis is the golden standard for detecting breast cancer, which is usually conducted manually by pathologists. However, this analysis is difficult even for skilled pathologists. Moreover, this diagnostic process is time-consuming due to its complexity. Therefore, in order to improve accuracy and efficiency, it is necessary to develop computer-aided diagnosis systems (CADs) for cancer recognition.

In the past decades, many automatic algorithms based on digital pathology have been proposed for tumor classification. By exploiting handcrafted features, a variety of machine learning methods have been used, such as support vector machines (SVM) [2], multi-layer perceptron (MLP) [3] and random forest (RF) [4]. Currently, convolutional neural networks (CNNs) have achieved remarkable success in this field [5, 6], benefiting from its advantage of extracting hierarchical features automatically. Typically, the whole pathological image is divided into small patches which are classified by a CNN, then these patch-wise predictions are integrated to obtain the final image-wise classification result. However, such patch-wise feature learning lacks the ability to capture global information.

To overcome this limitation, some researchers have attempted to use graph convolutional networks (GCNs) for pathological image classification [7, 8, 9]. Most of these works follow this workflow: Firstly, the pathological image is transformed into a graph representation, where detected cancer cells serve as nodes and edges are formed in terms of spatial distance. Secondly, they extract cell-level features as initial node embeddings. Thirdly, the cell-graph is fed into a GCN followed by a MLP to perform image-wise classification. In such setting, global features including spatial relations among cells are embedded in GCN.

However, the current pipeline still faces two challenges. Firstly, only cellular interactions is insufficient to completely represent the pathological structure. In fact, the tissue distribution is hierarchical with many substructures, such as stromal, gland, tumor, etc. To learn the intrinsic characteristic of cancerous tissue, it is necessary to aggregate multi-level structural information. Secondly, this multi-stage workflow is tedious and time consuming. The performance of GCN relies heavily on the previous steps such as cell detection and feature extraction.

Bin Dong is supported in part by the NSFC under Grant 12090022, 12090020, 11831002.

To tackle the aforementioned problems, we propose a novel framework named multi-scale graph wavelet neural network (MS-GWNN) for histopathological image classification. Graph wavelet neural network (GWNN) [10] replaces the graph Fourier transform in spectral GCN as graph wavelet transform. Further, GWNN has good localization property in node domain, making it more flexible to adjust the receptive fields of nodes (via the scaling parameter s). Based on GWNN, we present multi-scale graph wavelet neural network (MS-GWNN), which takes advantage of spectral graph wavelets to make multi-scale analysis. More specifically, after converting pathological images into graph representations, we use multiple GWNNs with different scaling parameters in parallel to obtain the multi-scale contextual information in graph topology. Then, all these features are aggregated to produce the final image-level (i.e. graph-level) classification prediction.

The main contributions are summarized as follows:

1) We propose a novel framework (MS-GWNN) for breast cancer diagnosis in the way of mapping pathological images into graph domain. By exploiting multi-scale graph wavelets, MS-GWNN can obtain multi-level tissue structural information, which is exactly what the pathology analysis needs.

2) MS-GWNN can be trained in an end-to-end manner. Compared to the previous multi-stage workflow based on GCN, MS-GWNN simplifies the diagnostic process. To the best of our knowledge, MS-GWNN is the first end-to-end framework to apply GCN to pathological image classification.

3) MS-GWNN is evaluated on two public breast cancer datasets (BACH and BreakHis), and it achieves an accuracy of 93.75% and 99.67% respectively. The results outperform existing state-of-the-art methods, demonstrating the superiority of our proposed model. Through ablation studies, we verify that multi-scale structural features are crucial for characterizing cancers.

2. METHOD

Graph Representation. Let $G = \{V, E, \mathbf{H}\}$ be an undirected graph, where V is the set of nodes with $|V| = N$ and E is the set of edges. $\mathbf{H} \in R^{N\times C}$ is the node embedding matrix. $\mathbf{A}$ is a symmetric adjacency matrix defining the graph topology. We denote the graph's normalized Laplacian matrix as $\mathbf{L} = \mathbf{I_N} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, where $\mathbf{I_N}$ is the identity matrix and $\mathbf{D}$ is a diagonal matrix with $D_{i,i} = \sum_j A_{i,j}$. Further, we denote $\mathbf{U}$ as the eigenvector decomposition of the normalized Laplacian matrix $\mathbf{L} = \mathbf{U\Lambda U}^T$, where $\mathbf{\Lambda} = \mathrm{Diag}\left(\lambda_1, \ldots, \lambda_N\right)$ is the diagonal matrix formed by the eigenvalues of $\mathbf{L}$.

Spectral Graph Convolution. Spectral convolution was proposed by [11, 12], which defines the graph convolution operation in the Fourier domain. It can be represented as:

$$\mathbf{g}_\theta \star \mathbf{x} = \mathbf{U}\mathbf{g}_\theta(\mathbf{\Lambda})\mathbf{U}^T\mathbf{x}.$$

$\mathbf{x} \in R^N$ is the signal to be processed and $\mathbf{g}_\theta = \mathrm{Diag}(\boldsymbol{\theta})$ is the convolutional filter parameterized by $\boldsymbol{\theta} \in R^N$. This definition of spectral graph convolution is not spatially localized in node domain. In other words, the feature aggregation of one node depends on all the nodes, not only its neighbourhood nodes. Moreover, it is computationally expensive to compute the eigendecomposition of Laplacian matrix. To improve computational efficiency, some researchers have attempted to use a truncated expansion based on Chebyshev polynomials to approximate the convolutional kernel [12, 13, 14].

Spectral Convolution Based on Graph Wavelets. Some researchers have defined wavelet transform on graphs[12, 15]. In this paper, we apply the spectral graph wavelet proposed in [12], which introduces a band-pass filter in the traditional graph Fourier domain mentioned above. We denote ψ_{si} as the wavelet centered at node i at the scale of s, thereby the graph wavelet basis is defined as:

$$\mathbf{\Psi_s} = (\psi_{s1}, \psi_{s2}, \ldots, \psi_{sN}) = \mathbf{U}\mathbf{H}_s\mathbf{U}^T.$$

$\mathbf{H}_s = \mathrm{Diag}(h(s\lambda_1), \ldots, h(s\lambda_N))$ is a scaling matrix with $h(s\lambda_i) = e^{\lambda_i s}$. Given a set of graph wavelets, the graph wavelet transform is defined as $\hat{\mathbf{x}} = \mathbf{\Psi}_s^{-1}\mathbf{x}$ and the inverse transform is $\mathbf{x} = \mathbf{\Psi}_s\hat{\mathbf{x}}$. $\mathbf{\Psi}_s^{-1}$ can be obtained by replacing $h(s\lambda_i)$ in $\mathbf{\Psi}_s$ with $h(-s\lambda_i)$. Further, in such setting, the graph convolution based on spectral wavelet bases reads as

$$\mathbf{g}_\theta \star \mathbf{x} = \mathbf{\Psi}_s g_\theta \mathbf{\Psi}_s^{-1}\mathbf{x}.$$

Similarly, $\mathbf{g}_\theta = \mathrm{Diag}(\boldsymbol{\theta})$ is the convolutional filter to be learned. As stated in the work of [10], the graph convolution using wavelet transform has excellent localization property, making it outperform the traditional spectral convolution in graph-related tasks such as node classification. In addition, the scaling parameter s controls the receptive fields of nodes in a continuous manner, different from the previous approach [13] using the discrete shortest path distance.

Graph Wavelet Neural Network (GWNN). More recently, Xu et al. [10] proposed graph wavelet neural network (GWNN) to introduce the graph wavelet transform to spectral GCN, offering high spareness and good localization for graph convolution. However, their experiments were performed only at one scale, which didn't leverage the localization property of graph wavelets to perform multi-scale analysis. In this paper, we consider a three-layer graph wavelet neural network (GWNN) for node classification. The structure of the m-layer can be divided into two steps:

feature transformation: $\mathbf{X}^{m'} = \mathbf{X}^m\mathbf{W}^m$,

graph convolution: $\mathbf{X}^{m+1} = \sigma\left(\mathbf{\Psi}_s\mathbf{F}^m\mathbf{\Psi}_s^{-1}\mathbf{X}^{m'}\right)$ [10].

$\mathbf{X}^m \in R^{N\times p}$ and $\mathbf{X}^{m+1} \in R^{N\times q}$ are the input and output tensor respectively. $\mathbf{W}^m \in R^{p\times q}$ is the matrix of filter parameters, $\mathbf{F}^m$ is the diagonal convolutional kernel, and σ is the activation function. More specifically, our forward model can be described as:

first layer: $\mathbf{X}^2 = relu\left(\mathbf{\Psi}_s\mathbf{F}^1\mathbf{\Psi}_s^{-1}\mathbf{X}^1\mathbf{W}^1\right)$,

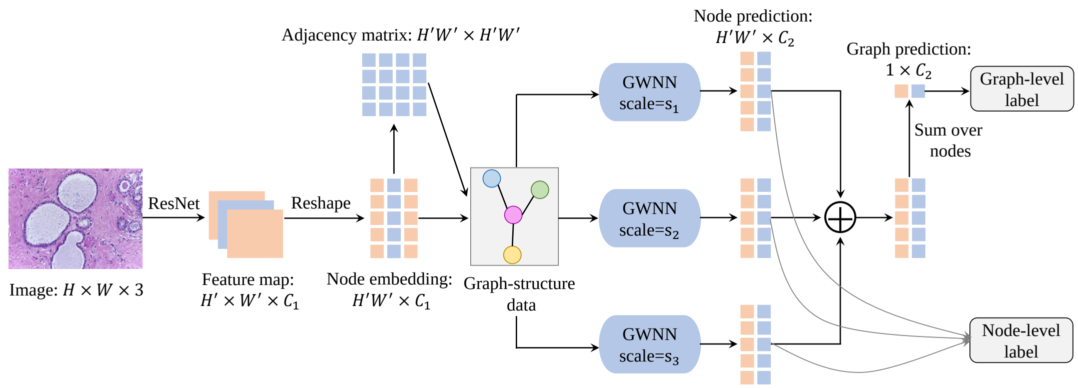

Fig. 1. The architecture of MS-GWNN. Firstly, pathological images are transformed into graph-structure data. Then, node classification is performed via GWNNs at different scales. Finally, the multi-level node embeddings are incorporated to yield the graph-level (image-level) classification result. $H^{'}W^{'} = N$ is the number of nodes. C_1 is the dimension of the initial node embeddings and C_2 is the number of cancer type. Both graph-level and node-level labels are used to train the model in an end-to-end way.

second layer: $\mathbf{X}^3 = relu\left(\mathbf{\Psi}_s\mathbf{F}^2\mathbf{\Psi}_s^{-1}\mathbf{X}^2\mathbf{W}^2\right)$,
third layer: $\mathbf{X}^4 = softmax\left(\mathbf{\Psi}_s\mathbf{F}^3\mathbf{\Psi}_s^{-1}\mathbf{X}^3\mathbf{W}^3\right)$.

2.1. Multi-Scale Graph Wavelet Neural Network (MS-GWNN)

In this section, we present a new framework called multi-scale graph wavelet neural network (MS-GWNN), for the task of pathological image classification. The architecture of MS-GWNN is shown in Fig.1, which consists of three parts: graph construction, node classification based on GWNN and graph classification using feature aggregation.

Graph Construction. In this work, we transform pathological images into graph representations. Nodes are nonoverlapping image patches and edges are generated in terms of the intrinsic relationships between these patches. Firstly, we use the modified ResNet-50 (we remove block3,4 and use average-pooling for downsampling) to learn discriminative features. In this way, we can obtain a series of feature maps ($H^{'} \times W^{'} \times C_1$), where each pixel corresponds to a $r \times r$ ($\frac{1}{r}$ is the downsampling rate) square patch in the raw image. Therefore, the feature vector at one pixel can be regarded as the node embedding of the corresponding patch (node).

Secondly, we make use of the similarity between node embeddings to form edges, which is defined by

$$f\left(\mathbf{x}_i, \mathbf{x}_j\right) = \theta\left(\mathbf{x}_i\right)^{\mathrm{T}} \phi\left(\mathbf{x}_j\right),$$

where $\mathbf{x}_i, \mathbf{x}_j$ are the node embeddings of node i and j, while $\theta(\mathbf{x}), \phi(\mathbf{x})$ are two transformation functions implemented via 1×1 convolutions. The formula to form graph edges is as follow:

$$e_{ij} = \begin{cases} 1, & \text{if } i = j \text{ or } f\left(\mathbf{x}_i, \mathbf{x}_j\right) \geq Q_\alpha(f) \\ 0, & \text{otherwise} \end{cases}$$

$Q_\alpha(f)$ is the α-th percentile ($0 < \alpha \leq 100$) of $\{f\left(\mathbf{x}_i, \mathbf{x}_j\right) | i, j = 1, 2, \ldots, N\}$. N is the number of nodes (i.e. $H^{'}W^{'}$ here).

Node Classification via GWNN. Given the constructed graph-structure data, we utilize multiple GWNNs with different scaling parameters s in parallel to conduct node classification. Each branch is supervised by the node-level label via the cross entropy loss function. Different parameters s mean different receptive fields, enabling the model to extract multi-scale contextual information.

Feature Aggregation. After the process of node classification, each branch can obtain a node prediction probability map with the size of $H^{'}W^{'} \times C_2$ (C_2 is the number of cancer type). In this work, we sum these probability maps to aggregate multi-scale structure representations. To perform graph-level (i.e. image-level) classification, we sum the node class probabilities to yield the graph-level prediction probability. Then, this graph-level prediction is passed to a softmax layer, which is supervised by the graph-level label.

In this framework, the final loss is formed by a weighted sum of the node-level loss and graph-level loss:

$$\mathcal{L}_{final} = \lambda \cdot \mathcal{L}_{node} + \mathcal{L}_{graph}, \tag{1}$$

where λ is the weight. Note that both $\mathcal{L}_{node}$ and $\mathcal{L}_{graph}$ are implemented with the classic cross entropy loss function, and $\mathcal{L}_{node}$ is the sum of the node-level losses from three branches. By minimizing the final loss, the error can be easily back-propagated through the whole MS-GWNN system in an end-to-end way.

3. EXPERIMENTS AND RESULTS

Datasets. We validate the proposed MS-GWNN on two public datasets: ICIAR 2018 breast cancer histology (BACH) grand challenge dataset [16] and BreakHis dataset [17]. The

Table 1. Comparisons with state-of-the-art methods on BACH dataset.

Model	Accuracy
Golatkar[20]	85.00%
Mahbod[21]	88.50%
Roy[5]	90.00%
Meng[22]	91.00%
Yao[23]	92.00%
Wang[24]	92.00%
Sitaula[25]	92.20%
Kassani [26]	92.50%
Proposed MS-GWNN	**93.75%**

BACH dataset consists of 400 histopathological images with size of $2048 \times 1536 \times 3$, which aims to predict breast cancer type as 1) normal, 2) benign, 3) in situ carcinoma, and 4) invasive carcinoma. We randomly select 320 samples for training and the rest 80 images for testing. In BreakHis dataset, there are 7909 samples classified as either benign or malignant. The 700×460 images are collected at different magnification factors ($40\times$, $100\times$, $200\times$, $400\times$). Experiments are conducted on the samples obtained at $40\times$ magnification (1995 images in total). In line with the previous works [18], the entire dataset is randomly divided into a training set with 70% samples and a testing set with 30% data. In this work, all images on BACH (BreakHis) dataset are resized to 512×384 (350×230) to reduce the memory workload of GPU. For image preprocessing, we use the classic H&E color normalization approach described in [19]. To avoid overfitting, extensive data augmentation is performed including flip, rotation, translation, shear and linear contrast normalization. Moreover, the performance of model is evaluated according to the average accuracy at the image level. More implementation details are in the supplementary material.

Results and Comparisons. As for breast cancer classification, the new MS-GWNN obtains an accuracy of 93.75% and 99.67% on BACH (as listed in Table 1) and BreakHis dataset (as listed in Table 1 of the supplementary material) respectively. The normalized confusion matrices are shown in Fig.1 of the supplementary material. On BACH dataset, both normal and invasive categories yield a high accuracy of 100%, while the accuracy of predicting in situ type is relatively low. This could be due to the similar structures appeared in inter-class images. In addition, we compare the results with existing state-of-the-art models. Obviously, MS-GWNN outperforms the other methods on both datasets, especially on BreakHis dataset where the error rate is only 0.33%. The results demonstrate the strong capacity of the proposed MS-GWNN model to tackle pathological image classification. More ablation studies are in the supplementary material.

4. CONCLUSION

In this work, we propose multi-scale graph wavelet neural network (MS-GWNN) for histopathological image classification. The MS-GWNN leverages the localization property of graph wavelets to perform multi-scale analysis with different scaling parameters s. For the task of breast cancer diagnosis, MS-GWNN outperforms the state-of-the-art approaches, mainly resulting from its powerful ability to integrate multi-scale contextual interactions. Through ablation studies, we verify the importance of exploiting multi-scale features. More broadly, the MS-GWNN model provides a novel solution to extract multi-scale structural information using graph wavelets, which can also be applied to other tasks.

5. REFERENCES

[1] Rebecca, L., Siegel, MPH, Kimberly, D., Miller, MPH, Ahmedin, and Jemal, "Cancer statistics, 2016," *Ca A Cancer Journal for Clinicians*, 2016.

[2] Paweł Filipczuk, Thomas Fevens, Adam Krzyżak, and Roman Monczak, "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE transactions on medical imaging*, vol. 32, no. 12, pp. 2169–2178, 2013.

[3] Yasmeen Mourice George, Hala Helmy Zayed, Mohamed Ismail Roushdy, and Bassant Mohamed Elbagoury, "Remote computer-aided breast cancer detection and diagnosis system based on cytological images," *IEEE Systems Journal*, vol. 8, no. 3, pp. 949–964, 2013.

[4] Cuong Nguyen, Yong Wang, and Hanam Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *Journal of Biomedical Science and Engineering*, vol. 2013, no. 5, pp. 551–560, 2013.

[5] Kaushiki Roy, Debapriya Banik, Debotosh Bhattacharjee, and Mita Nasipuri, "Patch-based system for classification of breast histology images using deep learning," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 90–103, 2019.

[6] Laith Alzubaidi, Omran Al-Shamma, Mohammed A Fadhel, Laith Farhan, Jinglan Zhang, and Ye Duan, "Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model," *Electronics*, vol. 9, no. 3, pp. 445, 2020.

[7] Yanning Zhou, Simon Graham, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng, and Nasir Rajpoot, "Cgc-net: Cell graph convolutional network for grading of colorectal cancer

histology images," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[8] Jingwen Wang, Richard J Chen, Ming Y Lu, Alexander Baras, and Faisal Mahmood, "Weakly supervised prostate tma classification via graph convolutional networks," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 239–243.

[9] Mohammed Adnan, Shivam Kalra, and Hamid R Tizhoosh, "Representation learning of histopathology images using graph neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 988–989.

[10] Bingbing Xu, Huawei Shen, Qi Cao, Yunqi Qiu, and Xueqi Cheng, "Graph wavelet neural network," *arXiv preprint arXiv:1904.07785*, 2019.

[11] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.

[12] David K Hammond, Pierre Vandergheynst, and Rémi Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[13] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in neural information processing systems*, 2016, pp. 3844–3852.

[14] Thomas N Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[15] Bin Dong, "Sparse representation on graphs by tight wavelet frames and applications," *Applied and Computational Harmonic Analysis*, vol. 42, no. 3, pp. 452–479, 2017.

[16] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al., "Bach: Grand challenge on breast cancer histology images," *Medical image analysis*, vol. 56, pp. 122–139, 2019.

[17] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2015.

[18] Dalal Bardou, Kun Zhang, and Sayed Mohammad Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018.

[19] Abhishek Vahadane, Tingying Peng, Shadi Albarqouni, Maximilian Baust, Katja Steiger, Anna Melissa Schlitter, Amit Sethi, Irene Esposito, and Nassir Navab, "Structure-preserved color normalization for histological images," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 1012–1015.

[20] Aditya Golatkar, Deepak Anand, and Amit Sethi, "Classification of breast cancer histology using deep learning," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 837–844.

[21] Amirreza Mahbod, Isabella Ellinger, Rupert Ecker, Örjan Smedby, and Chunliang Wang, "Breast cancer histological image classification using fine-tuned deep network fusion," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 754–762.

[22] Zhu Meng, Zhicheng Zhao, and Fei Su, "Multi-classification of breast cancer histology images by using gravitation loss," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1030–1034.

[23] Hongdou Yao, Xuejie Zhang, Xiaobing Zhou, and Shengyan Liu, "Parallel structure deep neural network using cnn and rnn with an attention mechanism for breast cancer histology image classification," *Cancers*, vol. 11, no. 12, pp. 1901, 2019.

[24] Zeya Wang, Nanqing Dong, Wei Dai, Sean D Rosario, and Eric P Xing, "Classification of breast cancer histopathological images using convolutional neural networks with hierarchical loss and global pooling," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 745–753.

[25] Chiranjibi Sitaula and Sunil Aryal, "Fusion of whole and part features for the classification of histopathological image of breast tissue," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1–12, 2020.

[26] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, Michal J Wesolowski, Kevin A Schneider, and Ralph Deters, "Breast cancer diagnosis with transfer learning and global pooling," *arXiv preprint arXiv:1909.11839*, 2019.