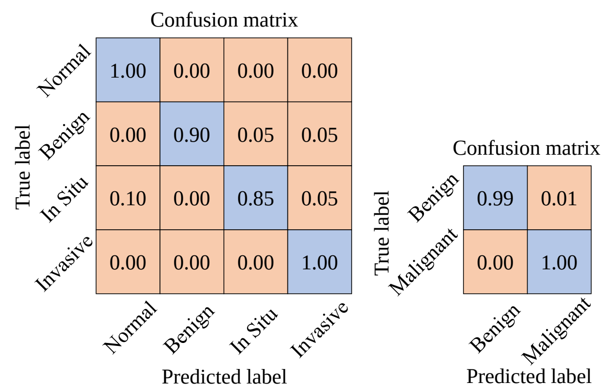
# SUPPLEMENTARY MATERIAL FOR MS-GWNN: MULTI-SCALE GRAPH WAVELET NEURAL NETWORK FOR BREAST CANCER DIAGNOSIS

*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.

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

| Model | Accuracy |
|---|---|
| Kumar[1] | 94.11% |
| Nahid[2] | 95.00% |
| Han[3] | 95.80% |
| Veeling[4] | 96.10% |
| Kausar[5] | 97.02% |
| Alom[6] | 97.95% |
| Bardou[7] | 98.33% |
| Gandomkar[8] | 98.60% |
| **Proposed MS-GWNN** | **99.67%** |




(a) BACH dataset
(b) BreakHis dataset

**Fig. 1**. Normalized confusion matrix results of MS-GWNN on two datasets.

## 1. EXPERIMENTS AND RESULTS

**Implementations.** To implement the proposed model, we use the Tensorflow platform on an NVIDIA GeForce GTX 1080 Ti GPU. Each GWNN model consists of three graph convolutional layers with the feature dimension of 256, 128 and 4 (2) respectively. The overall downsampling rate of the modified ResNet-50 is $1/16$. Moreover, we use the Adam optimizer to train the model, where the initial learning rate is $10^{-3}$ and $\beta_1 = 0.9, \beta_2 = 0.99$. All models are trained for 30000 epochs with a batch size of 16. The hyperparameters $\alpha, \lambda$ are set as 99 and 1 respectively.

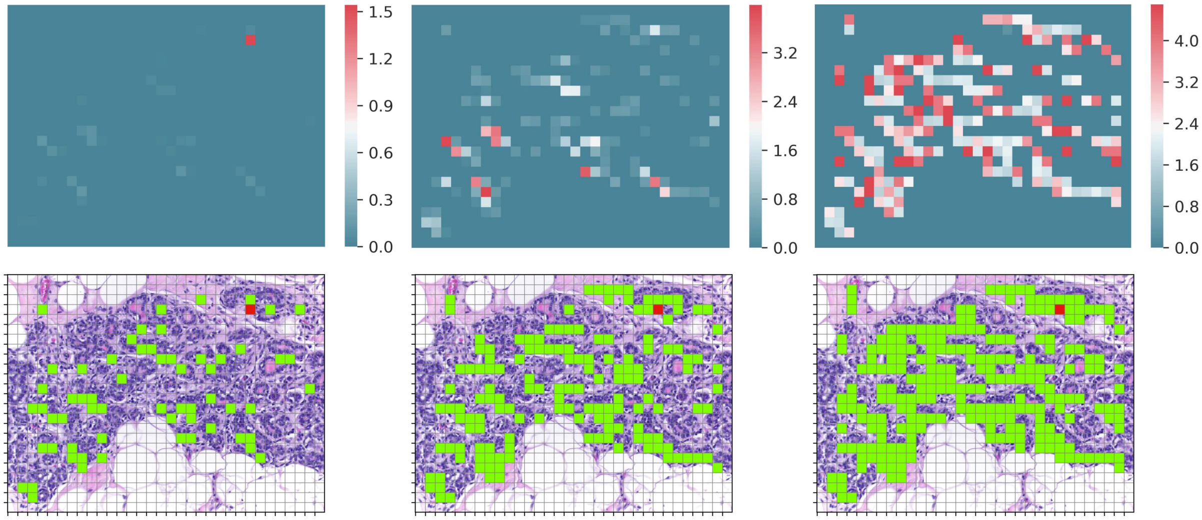
**Computational Complexity.** As the eigendecomposition of Laplacian matrix requires much computational cost, we utilize a fast algorithm to approximate the graph wavelets. Hammond et al. [9] suggested that $\mathbf{\Psi}_s$ and $\mathbf{\Psi}_s^{-1}$ can be efficiently approximated by a truncated expansion according to the Chebyshev polynomials with low orders. In this way, the computational complexity is $O(k \times |E|)$, where $|E|$ is the number of graph edges and $k$ is the degree of Chebyshev polynomials. In this work, we set $k = 2$.

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

## 2. ABLATION STUDIES

**The Function of Multi-Scale Feature Learning.** In this section, we aim at analyzing the effect of the multi-scale features fusion. As shown in Table 2, as aggregating features at more scales, the model performance becomes better accordingly. The specific scaling parameter $s$ is given in brackets. When using only one-level features ($s = 0.5$), MS-GWNN obtains an accuracy of 88.75%. After introducing the features at another scale ($s = 1.0$), the accuracy increases to 91.25%. Further, MS-GWNN with three branches ($s = 0.5, 1.0, 1.5$) can achieve a 93.75% accuracy. This comparison confirms that extracting multi-scale context information is very important for pathological image analysis.

In addition, we also perform an experiment which re-




(a) Small scale (b) Medium scale (c) Large scale

**Fig. 2**. Visualization of the wavelet bases at different scales. The small, medium and large scale correspond to $s = 1, 3, 5$ respectively. Top row: multi-scale graph wavelets $\Psi_{si}$ centered at node $i$. Bottom row: the receptive fields corresponding to the above wavelets. Each square denotes a node (patch) in the graph (image). The red point is the center node $i$ and the other green points are the neighborhood of node $i$. As the scaling parameter $s$ gets larger, the receptive field of node $i$ becomes wider accordingly.

**Table 2**. Ablation study of the multi-scale feature learning on BACH dataset. MS-GWNN-1 ($s = 0.5$) denotes the MS-GWNN model with one branch ($s = 0.5$).

| Model | Accuracy |
|---|---|
| GCN | 88.75% |
| MS-GWNN-1 (s=0.5) | 88.75% |
| MS-GWNN-1 (s=1.0) | 88.75% |
| MS-GWNN-1 (s=1.5) | 87.50% |
| MS-GWNN-2 (s=0.5,1.0) | 91.25% |
| MS-GWNN-2 (s=0.5,1.5) | 92.50% |
| MS-GWNN-2 (s=1.0,1.5) | 91.25% |
| **MS-GWNN-3 (s=0.5,1.0,1.5)** | **93.75%** |

places the multi-branch GWNNs in MS-GWNN as the traditional graph convolutional network (GCN) [10]. As shown in Table 2, GCN achieves a similar result to MS-GWNN with one branch. However, due to the high spareness of graph wavelets, GWNN is much more computationally efficient than GCN, which has also been discussed in the previous work [11]. **Visualization of the wavelet bases at different scales.** To further investigate the working mechanism of MS-GWNN, we show the graph wavelet bases at different scales in the top row of Fig. 2. Each small square represents a node in the constructed graph, and the node-to-patch correspondence relation is drawn in the raw image as shown in the second row of Fig. 2. Specifically, the red square denotes the node $i$ which the wavelet $\Psi_{si}$ centered at, and the green squares denote the neighbourhood of node $i$ whose values are greater than 0 in the wavelet $\Psi_{si}$. As the scale gets larger, the scope of neighborhood becomes wider accordingly, meaning that the receptive field of node $i$ is gradually expanding. At the small scale ($s = 1$), only a minority of nodes contribute to the embedding updating of node $i$. However, at the large scale ($s = 5$), the neighborhood nodes increase a lot which nearly spread on the entire tissue. In such setting, information can be propagated among the tissue structures at different levels, enabling MS-GWNN to acquire multi-scale contextual features.

## 3. REFERENCES


[1] Abhinav Kumar, Sanjay Kumar Singh, Sonal Saxena, K Lakshmanan, Arun Kumar Sangaiah, Himanshu Chauhan, Sameer Shrivastava, and Raj Kumar Singh, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Information Sciences*, vol. 508, pp. 405–421, 2020.

[2] Abdullah-Al Nahid and Yinan Kong, "Histopathological breast-image classification using concatenated r–g–b histogram information," *Annals of Data Science*, vol. 6, no. 3, pp. 513–529, 2019.

[3] Zhongyi Han, Benzheng Wei, Yuanjie Zheng, Yilong

Yin, Kejian Li, and Shuo Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.

[4] Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling, "Rotation equivariant cnns for digital pathology," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2018, pp. 210–218.

[5] Tasleem Kausar, MingJiang Wang, Muhammad Idrees, and Yun Lu, "Hwdcnn: Multi-class recognition in breast histopathology with haar wavelet decomposed image based convolution neural network," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 4, pp. 967–982, 2019.

[6] Md Zahangir Alom, Chris Yakopcic, Mst Shamima Nasrin, Tarek M Taha, and Vijayan K Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *Journal of digital imaging*, vol. 32, no. 4, pp. 605–617, 2019.

[7] 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.

[8] Ziba Gandomkar, Patrick C Brennan, and Claudia Mello-Thoms, "Mudern: Multi-category classification of breast histopathological image using deep residual networks," *Artificial intelligence in medicine*, vol. 88, pp. 14–24, 2018.

[9] 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.

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

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