Segmentation Aided Multiclass Tumor Classification in Ultrasound Images using Graph Neural Network

Iftekharul Islam Shovon[a], Ijaz Ahmad[b], and Seokjoo Shin[a]
[a]Dept. of Computer Engineering, Chosun University, Gwangju, Korea
[b]Dept. of Electrical and Computer Engineering, Korea University, Seoul, Korea
shovon@chosun.kr, ijaz@korea.ac.kr, sjshin@chosun.ac.kr (*Corresponding author*)

***Abstract*—The merit in adapting a Graph Neural Network (GNN) for image analysis is that it can capture long-range dependencies between distant parts of the image. This is particularly important in domains such as medical imaging, where it is crucial to distinguish between organs and/or normal and abnormal regions for a disease diagnosis. While GNNs offer significant benefits, they necessitate preprocessing techniques to effectively represent images as graphs. Several techniques are proposed in the literature to address this; however, their reliance on human intervention limits their applications. Therefore, this work proposes a data aided technique that complements the model with prior knowledge of the abnormal region location within the image. Specifically, we divide an image into patches and use a deep learning-based segmentation model to extract the mask of abnormal regions to learn patch, mask, and position embeddings for graph construction. This graph is fed into a GNN for multiclass classification of breast cancer in ultrasound images. Simulation analysis shows that the proposed segmentation-aided GNN model achieved better classification performance in terms of various evaluation metrics compared to existing models. For example, compared to existing GNN models that do not require additional data, our model achieved 4% better accuracy score.**

***Index Terms*—GNN, medical image, classification**

I. Introduction

Deep learning (DL) models keep improving for natural image classification in terms of their performance, number of parameters, floating-point operations, and training and inference speed. Consequently, their applications are extended to other computer vision-related domains. Several studies have implemented such DL models for medical image analysis. For example, ResNet [1] was used to detect colorectal cancer using colon glands images [2], MobileNet [3] was used for detecting brain tumours using MRI images [4], InceptionResNet [5] was used to detect COVID19 infection using X-ray and MRI images [6], VGG [7] was used to detect breast cancer in histopathology images [8], DenseNet [9] was used for medical image classification using mammography and osteosarcoma histology images [10], and EfficientNet [11] was used to detect tuberculosis in X-ray images [12].

In medical imaging, it is often challenging to train a DL model from scratch due to the scarcity of data. To deal with this, transfer learning and data augmentation techniques can be leveraged, as demonstrated in [13]. Besides data deficiency, medical image classification poses other challenges, such as medical images often being low-resolution grayscale images of internal body structure. In these images, organs and blood vessels frequently exhibit similar intensities, thus complicating the classification process. Therefore, Graph Neural Networks (GNN) are proposed that can distinguish different organs while capturing long-range dependencies between distant parts within the image. For example, [14] proposed a data-aided GNN model for chest X-ray image classification that uses gaze-point data to learn information relevant to a disease. Incorporating gaze-points data with their GNN model achieved better performance than their counterparts; however, generating such data requires expert intervention, thus limiting the applications of such data-aided GNN models.

In this paper, we introduce a data-aided method that uses a segmentation mask of an image to prepare it as an input for a GNN-based ultrasound image (USI) classification. The USI is initially processed to generate a masked image where some pixel intensity values are set to zero, while others remain non-zero [15]. The mask is for the tumor only, i.e., the USI that has tumor used segmentation to create the mask, the normal class has no mask. In our proposed system, the first step is patch embedding, in which the USI images and the segmented masks are converted into patches, which are then processed by a transformer model to extract features from them. Alongside the patch embedding, positional embedding and mask embedding are utilized to construct a graph. This graph is subsequently fed into a GNN, which utilizes the entire graph in what is known as a graph-level task to make predictions. Compared to our previous work [15], where we implemented a segmentation-aided GNN model for disease classification in chest X-ray images, we implemented it for a multiclass tumor classification in ultrasound images. Also, in our previous work, supplementary data was used to distinguish organs (that is, lungs) from the rest of the image, in this work, it is used to distinguish normal and abnormal regions (that is, tumor) to complement the GNN model with prior knowledge of the tumor location within the image. For comparison, we evaluate the performance of our model using metrics such as accuracy, precision, F1 score, recall, and specificity.

II. Proposed Method

This section provides an overview of the architecture of our proposed segmentation aided classification model for disease diagnosis in BUSI using GNNs. The overall schematic of our proposed model is outlined in Fig. 1, which is divided into two main modules: the graph generation module and GNN-based

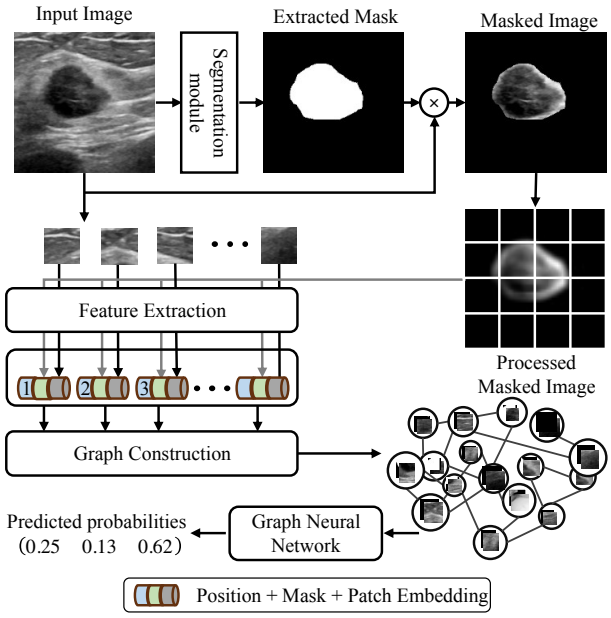


Fig. 1: An illustration of the proposed segmentation-aided classification scheme that utilizes image segmentation mask to give it a graph representation for GNN-based disease prediction in ultrasound images.

classification module. The first module construct a graph by using the actual image and its corresponding mask while the second module uses this generated graph for classification. The GNN updates and aggregates data from the node to create an intricate feature matrix of the entire graph. The whole graph is used for the prediction task, which is known as graph level classification. Each stage of this process is outlined below.

A. Graph Representation

For graph representation, the proposed method utilizes two types of input data: the BUSI and its corresponding mask. The mask data, which isolates the tumor area, differs from the typical regular grid structure of an image. Additionally, this mask serves as the source of attention information. To construct a graph, the mask and image data are transformed into feature vectors using the following embedding system.

1) Patch Embedding: In an image $J_{h,w}$, the dimensions, comprising the number of rows and columns, can be expressed as the products of two integers, with $h = p \times q$ for rows and $w = r \times q$ for columns. Consequently, the image can be divided into $P = p \times r$ square blocks, where each block contains q^2 pixels [15]. An image $J \in \mathbb{R}^{w \times h}$ is divided into P patches, denoted as $X = \{x_1, x_2, \ldots, b_P\}$, with each patch $x_j \in \mathbb{R}^{q \times q}$ for $j = 1, 2, \ldots, P$. For each patch x_j, a feature vector $f_j^J \in \mathbb{R}^D$ is derived to encapsulate local image characteristics [16]:

$$Y_j^J = A(x_j), \tag{1}$$

where $A(\cdot)$ is a function designed to extract features from image patches, as outlined in [16]. For enhanced computational efficiency, we consider each patch as a graph node rather than treating individual pixels as nodes.

2) Mask Embedding: This subsection elaborates on the mask creation and embedding methods employed in our system. In the domain of medical image analysis, segmentation distinguishes pixels that represent lesions or organs from those that form the background [15]. The isolation of these regions focuses attention on the areas of interest, ensuring that only tumor regions are emphasized, thereby reducing the likelihood of erroneous identifications by the model. For this purpose, deep learning-based automatic segmentation models can be leveraged to eliminate the experts intervention. Similar to the input image, the masked image is also partitioned into patches of size $M \times M$. Following [14], each patch P_i in the masked image is processed in (2) to represent its attention features.

$$Y_j^T = \sum_{(s_j, t_j) \in P_i} m(s_j, t_j), \tag{2}$$

where, $m_{(s_j,t_j)}$ is a pixel value at position (s_j, t_j) in the masked image and Y_j^T is the processed mask.

3) Position Embedding: GNN considers features as unordered nodes during graph processing, thus we employ the position embedding technique described in [17] to retain the positional information of the original image. The positional embedding approach involves two steps. Initially, a learnable absolute positional encoding vector $(e_i \in \mathbb{R}^D)$ is added to the feature vector $(Y_j^J + Y_j^T)$. Subsequently, the relative positional distance between nodes, computed as $e_i^T e_j$, is utilized as a distance metric within the *k*-nearest neighbor algorithm to identify adjacent nodes for graph construction.

B. Graph Construction

A graph $G = \{V, E\}$ is constructed using the set of vertices (V) and edges (E) defined in (3) and (4), respectively. The vertices V are composed of the mask embedding Y_i^J, position embedding Y_i^T and graph feature vector v_i

$$v_i = Y_i^J + Y_i^T + e_i. \tag{3}$$

To define the edges of the graph, we use k-nearest neighbors

$$E = \{(v_i, v_j) \mid v_j \in K(v_i)\}. \tag{4}$$

The $K(v_i)$ represents the neighbors of v_i.

C. Graph Neural Network

The proposed model includes L graph processing blocks, which draw inspiration from [17]. These blocks incorporate an average pooling layer along with a graph classification head, enhanced by multiple fully connected (FC) layers and a graph convolution layer (GCN) as detailed in [18]. If a graph is denoted as N, with D-dimensional feature vectors, and the input of the graph at block t is $A^t = [a_1^t, a_2^t, \ldots, a_P^t] \in \mathbb{R}^{P \times D}$, the graph processing block outputs $A^t \in \mathbb{R}^{P \times D}$ as:

$$U^t = \phi_2(\gamma(\phi_1(Z^t))) + Z^t, \tag{5}$$

$$A^t = \phi_4(\phi_3(U^t)) + U^t. \tag{6}$$

Here, ϕ denotes the graph convolution and γ denotes FC layers. U^t represents the intermediate output after the first shortcut connection, and $M^t = \gamma_1(V^t)$ is the input for the graph convolution layer. Hence, the graph convolution $S^t = \gamma(M^t)$ is constructed as:

$$f_i^t = \mathbf{W} \cdot \max(\{m_i^t - m_j^t \mid j \in K(m_i^t)\}). \tag{7}$$

Here, $\mathbf{W}$ is the learnable weight matrix for updating features of nodes. The aggregation used here is the max function, which aggregates the maximum features from the i-th node's neighbors. Thus, the graph convolution aggregates neighbors' feature information into the node feature, and finally, the classification head, which is a series of FC layers with a softmax function, predicts the probability of each category.

III. Simulation Results

A. Experimental Setup

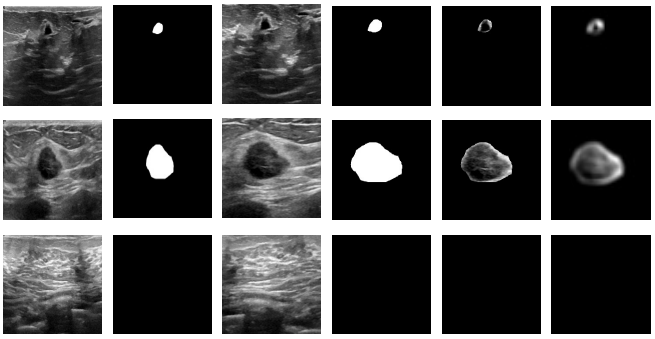
The experiments were conducted on a Windows PC equipped with an Intel i5 CPU and an NVIDIA RTX 3060 GPU using PyTorch. The AdamW optimizer [19] was selected for the experiments. For model hyperparameters, we followed the training setup of [14]. Also, during training, the model with the highest accuracy was saved as our best model. For baselines, we considered a [14]'s GNN model and the implementation of different CNN models in [13].

B. Dataset

For the experiments, we used a publicly available BUSI dataset [20], which comprises 780 samples and their masks, divided into three classes: Normal, Benign, and Malignant. The original images are of size 933×571, all in grayscale, which we resized to 224×224 using random center crop before partitioning them into patches. Besides random cropping, we applied random flip and rotation to the training set as our data augmentation technique. Fig. 2. shows example images from our dataset for each class along with their corresponding masks and preprocessing to obtain mask embedding.

C. Performance Analysis

Table I summarizes a detailed performance of proposed model in terms of parameters, data size, accuracy, precision, recall, F1-score, and specificity. For the baseline, we considered conventional CNN methods (such as DenseNet-121, ResNet-50, MobileNet-V2, InceptionResNet-V2, VGG-16, and Inception-V3 models) and [13] implemented on the same dataset. The metric scores referenced are directly reported from source [13], except for [14], which we obtained by running their open-source code. Since gaze data was unavailable for this dataset, we executed the code without incorporating gaze data. Following the model architecture of [14], our proposed model employs a transformer model to learn patch embedding, and a GNN model to process graph representation of an image for the classification. From Table I, it is evident that our proposed method surpasses all the CNN and GNN models except [9], whose accuracy is 1% better than ours. This can attributed to their augmented dataset, which is $5\times$ larger than the one used in this study, and the large number of parameters used in their model. Also, compared to [14]'s GNN model, our proposed model achieved better performance across all metrics. It is worth mentioning that their GNN model is a data-aided model, which requires gaze-points data to learn the abnormal region locations for better performance. Supplementing this model with gaze may improve their model performance. However, in practice, it is difficult to acquire such gaze-points data without experts intervention. On the other hand, proposed method can be readily extended to any dataset by leveraging deep learning based segmentation model. In addition, to analyze performance of proposed model for different types tumors in comparison with existing GNN model [14], Fig. 3. plots the confusion matrix and Receiver Operating Characteristics (AUC) curves. It can be observed that proposed method efficiently differentiated the benign and malignant classes from the normal class.



(a) (b) (c) (d) (e) (f)

Fig. 2: Example images for each label from the dataset with preprocessing results: (a) original ultrasound image, (b) corresponding masks of the tumor region, (c) and (d) are center cropped images and masks, respectively, (e) combination of segmented regions of the cropped image and mask, and (f) mask embeddings. Rows correspond to Benign, Malignant, and Normal class examples, respectively.

IV. Conclusion

This paper proposed a novel segmentation-aided medical image classification framework leveraging Graph Neural Networks (GNN). Our approach utilized ultrasound images and their corresponding masked images to construct a graph, which the GNN processes for disease classification. A fundamental limitation of existing data-aided GNN techniques is that they require gaze points data, which is not always readily available and rely on experts intervention to generate such data. Proposed method deal with it by using deep learning-based automatic mask generation to aid the GNN model in classification. Results showed that proposed model is effective, outperforming several deep learning and existing data-aided techniques in terms of accuracy, precision, recall, F1-score, specificity and average AUC scores. In the future, we are interested to implement a more efficient technique to process the mask data for graph representation.

TABLE I: Performance analysis of proposed segmentation-aided classification model with existing deep learning techniques on the same dataset using different evaluation metrics.

Methods	**Parameters** ($\times 10^6$)	**Data Size**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Specificity (%)**
ResNet-50 [1]	26.6	3,900	85.08	80.92	81.16	81.00	89.54
MobileNet-V2 [3]	3.5		89.03	88.83	90.50	89.30	94.61
InceptionResNet-V2 [5]	55.9		92.61	88.17	92.00	89.50	95.82
VGG-16 [7]	138.4		93.97	92.83	94.83	90.50	97.59
Inception-V3 [21]	22.9		89.90	90.77	89.83	89.83	94.35
DenseNet-121 [9]	8.1		95.48	95.00	94.67	94.8	97.28
CNNTF [13]	2.789		92.46	94.06	92.59	92.59	92.15
GNN [14]	9.69	780	90.44	88.20	90.51	89.26	95.04
Proposed Method	2.2		94.50	92.62	91.73	92.16	94.88

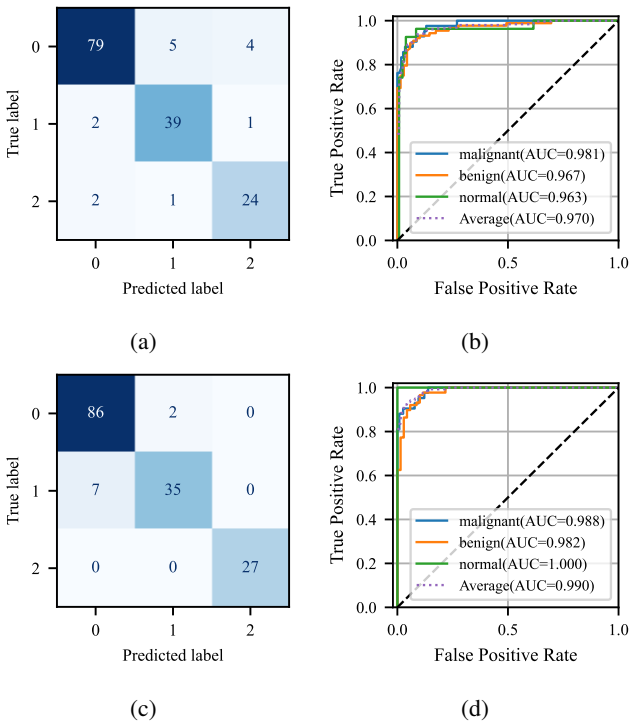


Fig. 3: Evaluation of proposed and [14]'s GNN model on the BUSI dataset. For each model, the confusion matrix and ROC curves are in the first and second columns, respectively. The metrics in the first and second rows are for [14] and proposed model, respectively. The labels 0, 1, and 2 represents the Benign, Malignant and Normal classes in the dataset.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (RS-2023-00278294).

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[2] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer," *Procedia Computer Science*, vol. 179, pp. 423–431, 2021.

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[4] T. H. Arfan, M. Hayaty, and A. Hadinegoro, "Classification of brain tumours types based on MRI images using MobileNet," in *2021 2nd ICITech*, pp. 69–73, IEEE, 2021.

[5] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-Resnet and the impact of residual connections on learning," in *Proceeding of the AAAI conference on artificial intelligence*, vol. 31, 2017.

[6] Y. Chen, Y. Lin, X. Xu, J. Ding, C. Li, Y. Zeng, W. Liu, W. Xie, and J. Huang, "Classification of lungs infected covid-19 images based on Inception-ResNet," *CMAPIB*, vol. 225, p. 107053, 2022.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. Ahmad Almansour, "Deep CNN model based on VGG16 for breast cancer classification," pp. 805–810, 2021.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[10] Z. Huang, X. Zhu, M. Ding, and X. Zhang, "Medical image classification using a light-weighted hybrid neural network based on pcanet and densenet," vol. 8, pp. 24697–24712, IEEE, 2020.

[11] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in *International conference on machine learning*, pp. 10096–10106, PMLR, 2021.

[12] I. Ahmad and S. Shin, "A perceptual encryption-based image communication system for deep learning-based tuberculosis diagnosis using healthcare cloud services," *Electronics*, vol. 11, no. 16, p. 2514, 2022.

[13] A. Ciobotaru, M. A. Bota, D. I. Goța, and L. C. Miclea, "Multi-instance classification of breast tumor ultrasound images using convolutional neural networks and transfer learning," *Bioengineering*, vol. 10, no. 12, p. 1419, 2023.

[14] B. Wang, H. Pan, A. Aboah, Z. Zhang, E. Keles, D. Torigian, B. Turkbey, E. Krupinski, J. Udupa, and U. Bagci, "GazeGnn: A gaze-guided graph neural network for chest x-ray classification," in *Proceedings of the IEEE/CVF Winter*, pp. 2194–2203, 2024.

[15] I. I. Shovon, I. Ahmad, and S. Shin, "Segmentation aided multiclass classification of lung disease in chest x-ray images using graph neural networks," in *ICOIN*, IEEE, 2025. (in press).

[16] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pvt v2: Improved baselines with pyramid vision transformer," *Computational Visual Media*, vol. 8, pp. 415–424, Sep 2022.

[17] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, "Vision gnn: an image is worth graph of nodes," NIPS '22, 2024.

[18] G. Li, M. Muller, A. Thabet, and B. Ghanem, "Deepgcns: Can gcns go as deep as cnns?," in *IEEE/CVF*, pp. 9267–9276, 2019.

[19] I. Loshchilov, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[20] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104863, 2020.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.