MRCG: A MRI Retrieval Framework with Convolutional and Graph Neural Networks for Secure and Private IoMT

Zhiri Tang, *Student Member, IEEE,* Zhao-Hui Sun, Edmond Q. Wu, *Senior Member, IEEE,* Chuan-Feng Wei, Dong Ming, *Senior Member, IEEE,* Sheng-Di Chen

***Abstract*—In the context of Industry 4.0, the medical industry is horizontally integrating the medical resources of the entire industry through the Internet of Things (IoT) and digital interconnection technologies. Speeding up the establishment of the public retrieval database of diagnosis-related historical data is a common call for the entire industry. Among them, the Magnetic Resonance Imaging (MRI) retrieval system, which is one of the key tools for secure and private the Internet of Medical Things (IoMT), is significant for patients to check their conditions and doctors to make clinical diagnoses securely and privately. Hence, this paper proposes a framework named MRCG that integrates Convolutional Neural Network (CNN) and Graph Neural Network (GNN) by incorporating the relationship between multiple gallery images in the graph structure. First, we adopt a Vgg16-based triplet network jointly trained for similarity learning and classification task. Next, a graph is constructed from the extracted features of triplet CNN where each node feature encodes a query-gallery image pair. The edge weight between nodes represents the similarity between two gallery images. Finally, a GNN with skip connections is adopted to learn on the constructed graph and predict the similarity score of each query-gallery image pair. Besides, Focal loss is also adopted while training GNN to tackle the class imbalance of the nodes. Experimental results on some benchmark datasets, including the CE-MRI dataset and a public MRI dataset from the Kaggle platform, show that the proposed MRCG can achieve 88.64% mAP and 86.59% mAP, respectively. Compared with some other state-of-the-art models, the MRCG can also outperform all the baseline models.**

***Index Terms*—Industry 4.0, Magnetic Resonance Imaging, the Internet of Medical Things, Convolutional Neural Network, Graph Neural Network**

This work was supported by the National Natural Science Foundation of China under Grant 81925020. Corresponding author: Dong Ming.

Zhiri Tang is with Department of Computer Science, City University of Hong Kong, Hong Kong, China. (E-mail: gerintang@163.com)

Zhao-Hui Sun is with School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. (E-mail: zh.sun@sjtu.edu.cn)

Edmond Q. Wu is with Department of Automation, Shanghai Jiao Tong University, Shanghai, China. (E-mail: edmondqwu@163.com)

Chuan-Feng Wei is with Human Space-Flight System Engineering Division, Institute of Manned Space System Engineering, CAST, Beijing, China.

Dong Ming is with Academy of Medical Engineering and Translation Medicine, Tianjin University, Tianjin, China. He is also with College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China. (E-mail:)

Sheng-Di Chen is with Department of Neurology and Institute of Neurology, Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China. (E-mail: chensd@rjh.com.cn)

Manuscript received April xx, 20xx; revised August xx, 20xx.

I. Introduction

Benefiting from the wave of digital transformation in the context of Industry 4.0, advanced general hospitals have integrated their medical resources through vertical integration [1], [2], [3]. Currently, plenty of medical alliances are also opening up the originally isolated information systems of the hospital through horizontal integration and end-to-end integration and promote the further utilization of medical resources in the form of more open and collaborative [4], [5], [6], [7]. Specifically, the hospitals that have established the alliance are contributing their originally closed medical data and building a larger and richer medical resource database through the alliance to serve doctors' diagnoses [8], [9], [10]. In addition, as the modern patient, what he faces is no longer a single doctor but a networked medical resource database. Patients have the opportunity to use mobile terminals to safely and actively participate in the decision-making of self-medical management and health management instead of giving all their private data permissions to medical institutions. As a typical standardized medical data, magnetic resonance imaging (MRI) is one of the important references for medical diagnosis. With the development of medical databases, a massive amount of MRI data is collected daily. To use MRI data effectively, it is important to develop a safe and private retrieval system for clinicians, patients, and even healthy people [11], [12], [13]. Retrieving the image with the highest similarity from the medical database is significant for successful clinical decision making and diagnostic opinion. At the same time, related patients and healthy people are also interested in checking their conditions timely and privately [14], [15]. Hence, how to build an effective MRI retrieval system for secure and private the Internet of Medical Things (IoMT) is necessary for both scientific research and medical applications.

With the development of the Convolutional Neural Network (CNN), it has been applied in many medical areas, including medical image/signal classification [16], [17], tumor localization [18], and clinical decision making [19], [20]. Due to its powerful feature extraction ability in Euclidean space, CNN can also extract much useful information from MRI data. To make use of the features from deep learning and improve the image retrieval performance, some similarity learning methods, such as Siamese [21] and triplet [22] architecture, is proposed, which aims to shorten the distance between similar images in the feature space. However, these methods

consider only a pair of query-gallery images during training and ignore the relationship between gallery images, which limits the performance of MRI retrieval. Recently, Graph Neural Network (GNN) has attracted attention from many researchers because it can learn a graph representation from non-Euclidean space. Some works [23] have proved that the GNN can effectively reveal the inner connections among data via a graph structure, which is an idealized tool to build a similarity learning method for global features. Although it is a powerful combination in theory, how to merge the CNN and GNN for constructing a powerful MRI retrieval system for IoMT is still a challenge.

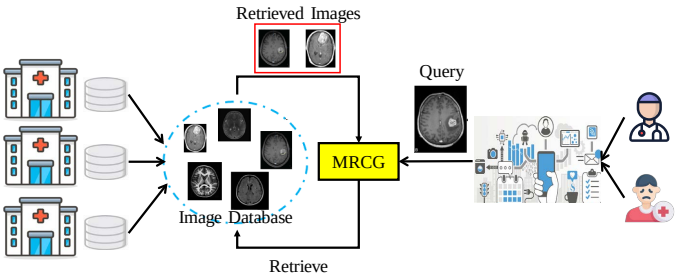

Fig. 1. An overview of the proposed MRCG system. The input includes query images from the Internet of Things (IoT) and an image database containing massive MRI data. Then the MRCG system will output the retrieval results.

Inspired by the above, an MRI retrieval system with CNN and GNN, named MRCG, is proposed in this work, where an overview of the proposed MRCG system is shown in Fig. 1. First, a pre-trained CNN with triplet architecture is adopted as the backbone, which is trained jointly for similarity learning and classification as an auxiliary task. Before the training, the query-gallery image sets are constructed by selecting the most similar images from the database as gallery images for each query image. The query-gallery sets are then fed into CNN for extract features, from which the graph is built. Through this process, a single graph encodes the relationship between a query image and N number of most similar gallery images. Next, the constructed graph is fed into GNN and trained for the node classification task, where the output of GNN is the similarity score for each query-gallery pair encoded in each node. Finally, k nodes with the highest similarity score are returned as retrieval results. Experimental results, including many evaluation metrics, on benchmark datasets show that the proposed MRCG can retrieve the image from MRI databases effectively. Compared with other latest methods, the proposed MRCG can also achieve better performance.

The main contributions of this work can be summarized as follows:

1) The query-gallery image sets are constructed and fed to a CNN trained jointly for similarity learning and classification, which helps to extract the global features and obtain more useful information;

2) A graph, which is built based on extracted features from CNN, is input to train the GNN and obtain the node classification and then retrieval results;

3) Experimental results on some benchmark datasets and comparisons with state-of-the-art works verify the performance of the proposed MRCG.

The rest of this paper is organized as follows: Section II reviews related work, Section III introduces the proposed method, Section IV presents the experimental setup and results, and Section V summarizes the work and suggests future work.

II. Related Work

Retrieval system plays an importance role in IoMT including not only image retrieval [24], [25] but also information retrieval [26]. Traditional image retrieval approaches extract hand-crafted features from images which are then used to calculate the similarity between MRI. Some classic works are listed as follows: Bag-of-words (BoW) [27] combines with metric learning using Mahalanobis distance; Fisher vector [28] combines local features and forms the representation of a single image; Margin information descriptor (MID) [29] is used to extract the local features of the tissues surrounding brain tumors; Huang et al. [30] used brain tumors as ROI to construct a BoW model and calculate the data similarity using rank error-based metric learning. Besides the above traditional methods, many other novel works have emerged in recent years [24], [26], [31], [32].

With the rapid development of deep learning, CNN has become the popular approach to extract features from Euclidean space. In addition, GNN attracts attention from many researchers for its powerful graph representation in non-Euclidean space. Hence, the following section reviews relevant studies on image retrieval and categorizes them into two types: CNN-based and GNN-based studies.

A. CNN-based Approaches

Feature extraction is the main task in CNN-based MRI retrieval. Various pre-trained CNN has been used for feature extraction on interstitial lung disease (ILD) datasets [33]. Besides, a 3D-CNN [34] is proposed to extract brain tumor from the images and later feed into pre-trained Vgg19 to get the feature representations of the brain tumor. Earlier, methods that combined CNN and other techniques are used in medical image retrieval tasks. Liu et al. [35] search images based on the feature extracted from CNN and the search results are refined with Radon barcodes. Furthermore, CNN is used jointly with edge histogram descriptor (EHD) to extract the global and local features of the images [36].

As an end-to-end approach to extract image features for image retrieval tasks, CNN is first trained for classification tasks and then the intermediate fully connected is later used to calculate the Euclidean distance between images [37], [38]. Deep Belief CNN [39] is also used to learn the image representations. In addition, Kruthika et al. [40] ensemble 3D Capsule Networks, 3D CNN, and 3D autoencoders to improve the learning of feature representations.

In recent years, Siamese CNN [21], [41] has been adopted for similarity learning using pre-trained CNN as the backbone. The aim of similarity learning is to minimize the distance between similar images and maximize the distance between different images in the feature space. Owais et al. [42] show

that image features extracted using transfer learning are more distinctive compared with training CNN from scratch. Besides, CNN-based MRI processing system [43] has also been applied to many areas, such as abnormal detection [44] and Alzheimer's disease classification [45].

B. GNN-based Approaches

Recently, GNN has attracted attention due to its ability to learn the representations of the graph that has a non-Euclidean structure. With the emergence of GNN, content-based retrieval approaches using GNN have been proposed with two common pipelines: (1) query and gallery image features are first extracted with CNN. A graph is then constructed from the extracted features where each node encodes the relationship between single query-gallery image pair [23]; (2) image is first segmented and fed into CNN for feature extraction. And then, a graph is constructed from the extracted features of the subregions where each node represents a subregion [46]. GNN, with application to retrieval task, has been receiving growing interest because it can learn the relationship between images that are linked with graph edges. It overcomes the limitations of similarity learning based on CNN that considers only the pairwise similarity of query-gallery image pairs. As such, a GNN-based retrieval system can learn more information compared with those approaches using CNN only.

To handle data in a graph structure, recent studies focus on GNN, which mainly consists of spectral-based [47], [48] and spatial-based [49], [50] graph convolutions. For example, Yan et al. [23] employ a relative attention module to filter useful context information in the image. Next, the graph is built based on the context information and trained with GNN.

Till now, the usage of GNN in the medical image is still in its infancy. Adnan et al. [46] represent each patch from the whole slide image (WSI) as graph nodes. ChebNet also outperforms GraphSage in the lung cancer classification task [47]. In addition, Gu et al. [51] conduct experiments on probe-based confocal laser endomicroscopy (pCLE) and histology images by building a multimodal graph for the cross-modal retrieval task. Hence, how to adopt the GNN for MRI retrieval system effectively is still a challenge. Furthermore, constructing a CNN and GNN-based retrieval system is also curial for IoMT.

III. Methodology

The main pipeline of the proposed method is shown in Fig. 2. First, a set of images is fed into CNN for feature extraction. Next, the graph is constructed from the feature vectors where each node feature embeds the relationship of a query-gallery image pair. The edge weight between two nodes is defined as the similarity between both gallery images. However, the convolution operator in CNN cannot be applied directly to the graph. To solve this issue, the graph convolutional layer (GCL) is adopted to handle the graph data to learn the node representations, which are then utilized for node classification. Finally, GNN classifies each node as 0 or 1 to decide whether the query-gallery pairs belong to the same class.

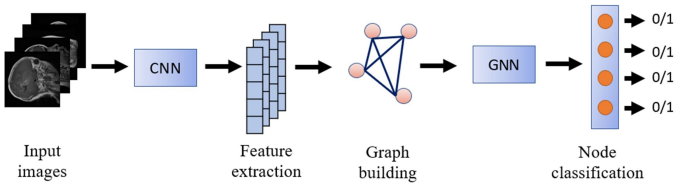

Fig. 2. The main pipeline of the proposed method. A set of query-gallery images are fed into CNN for feature extraction. Next, the graph is built using the extracted features where each node encodes the relationship between query and gallery images, whereas the graph edges are defined based on the similarity between gallery images. Finally, GNN is adopted to output the similarity score of each query-gallery image pair and therefore it can be formulated as a node classification task.

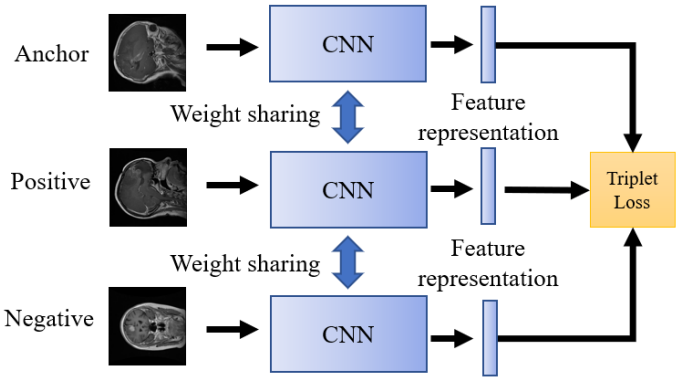

Fig. 3. An overview of the proposed triplet network, where anchor-positive-negative image pairs are constructed and fed into the network for feature extraction.

A. Convolutional Neural Network

First, the triplet networks architecture with a pre-trained Vgg16 [52] is adopted as the CNN backbone, as shown in Fig. 3. To train the triplet network, anchor-positive-negative image pairs are constructed and fed into the network for feature extraction. The network is penalized by triplet loss function during training which is defined as

$$L_{Triple} = \sum_{i}^{M} [||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) - f(x_i^n)||_2^2 + m]_+ \tag{1}$$

where M is the number of triplet pairs, $f(\cdot)$ represents the feature extraction CNN network. x^a, x^p, x^n denote the anchor, positive, and negative image, respectively. m controls the margin between positive samples and negative samples to maintain the distance between positive and negative samples so that they do not converge to the same spot in the feature space.

To improve the representation ability of CNN, the model is trained jointly on similarity learning and classification task. Therefore, the CNN exists as a branched network with the first branch responsible for similarity learning, whereas the second branch is responsible for the classification task, as shown in Fig. 4. We adopt a transfer learning and fine-tuning strategy by freezing the parameters of blue blocks, training only the green block in Fig. 4. Hence, we define the total loss function of the CNN model L_{CNN} as

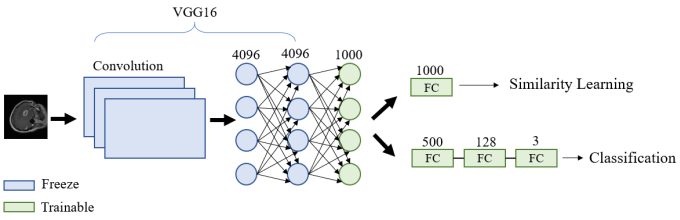

Fig. 4. The proposed CNN architecture with pre-trained Vgg16 as the backbone. It consists of convolution layers and branched fully connected layers with the first branch responsible for similarity learning, whereas the second branch responsible for the classification task. To implement transfer learning, we freeze the blue layers and train only the parameters in green layers. FC in the figure represents fully connected layers, and the number on top of it denotes the dimension of the outputs. Specifically, the output of the last FC depends on the categories of dataset, i.e. 3 for the CE-MRI and 4 for the Kaggle dataset.

$$L_{CNN} = L_{Triple} + \lambda L_{CE} \tag{2}$$

where λ controls the weights of cross-entropy loss on the total loss. $L_{Triplet}$ and L_{CE} denote triplet loss and cross-entropy loss, respectively.

B. Graph Neural Network

1) Graph Construction: Define graph as $G = (V, E)$, where V and E represent vertices and edges, respectively. Before constructing the graph, query-gallery pairs are formed by sampling N images from the database with the highest similarity score to the query image as gallery images. Therefore, hard negative samples are included in the gallery images, which allow the network to achieve more discriminative node representation. As such, a single graph with N nodes embeds the relationship between a single query image and N gallery images. Each node is associated with node features v_i, which is defined as

$$v_i = q - g_i, i = 1, 2, ..., N \tag{3}$$

where q and g_i are query and gallery feature vectors extracted from the CNN network. Nodes with query-gallery pair being the same class are labeled as 1, whereas nodes with query-gallery pair from different class are labeled as 0.

The edges between nodes are represented by an adjacency matrix A where its elements represent the edge weights between nodes. In this paper, we use cosine similarity to define the edge weights as the similarity between two gallery images. Given two gallery image feature vectors, g_i and g_j, cosine similarity between them c_{ij} is defined as

$$c_{ij} = \frac{g_i^T g_j}{||g_i||||g_j||} \tag{4}$$

However, the above equation returns a value in the range $[-1, 1]$. To ensure the training process is stable and the constructed adjacency matrix is symmetric, we define the similarity using the angle distance so that the edge weight is in the range $[0, 1]$. The symmetric property is necessary because we are using spectral GNN, which receives undirected graphs as input. Hence, the edge weight between node i and node j, a_{ij} is defined as

$$a_{ij} = 1 - \frac{\arccos c_{ij}}{\pi} \tag{5}$$

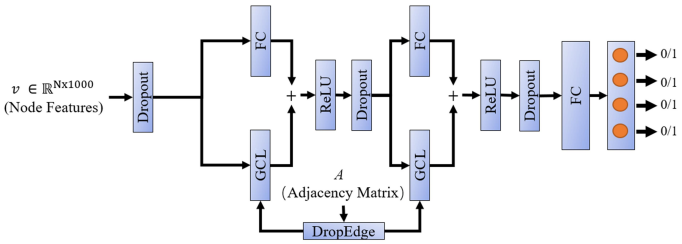

Fig. 5. The proposed GNN model. Adjacency matrix A and node features v_i are fed into the model as input, passing through GCL and FC to learn the graph structure. Finally, the model output the probability of the query-gallery image pairs being the same class. Skip connection is also adopted and ReLU is used as an activation function. DropEdge and Dropout are implemented to avoid overfitting.

2) Graph Convolutional Layers: The goal of the GNN model is to predict whether each gallery image is positive or negative samples. Therefore, the problem is formulated as a binary node classification task, where the GCL is adopted to perform convolution in the graph structure.

GCL learns the node representations of a particular node by aggregating features from its neighbors and its own features. This can be done by adding self-loop using the formula, $\widehat{A} = A + I$. Besides, the adjacency matrix $\widehat{A}$ is normalized with the formula $\widehat{D}^{-1/2}\widehat{A}\widehat{D}^{-1/2}$ to ensure numerical stability and avoid the exploding/vanishing gradient issues. $\widehat{D}$ denotes the degree matrix of the normalized adjacency matrix $\widehat{A}$. The propagation rule is defined as

$$X^{l+1} = \sigma(\widehat{D}^{-\frac{1}{2}}\widehat{A}\widehat{D}^{-\frac{1}{2}}X^l W^l) \tag{6}$$

where X^l represents the graph signal of the l-th layer, W^l denotes the trainable weight matrix, $\sigma(\cdot)$ denotes the activation function.

In addition, skip connection is introduced in the GCL to improve the learning of the model, as it has been used in CNN [53] as an effective approach to tackle the vanishing gradient problem. Hence, the graph signal can be updated by

$$X^{l+1} = \sigma(\widehat{D}^{-\frac{1}{2}}\widehat{A}\widehat{D}^{-\frac{1}{2}}X^l W^l + F(X^l)) \tag{7}$$

where $F(\cdot)$ denotes the fully connected layers with trainable parameters.

The architecture of the proposed GNN is illustrated in Fig. 5. The input to the GNN model consists of adjacency matrix A and node features $v \in \mathbb{R}^{N\times 1000}$. A two-layer GCL architecture is adopted with a fully connected layer as the final layer. The output of the final layer is fed into the sigmoid function to represent the probability of each node having the same class as the query image. A node with an output probability closer to 1 indicates the corresponding gallery image is similar to the query image and vice versa. ReLU is used as the activation of the GCN layer. Dropout [54] is commonly used in CNN to prevent overfitting. Similarly, we

utilized DropEdge [55] to randomly remove the graph edges as well as Dropout to alleviate the overfitting problem and improve the generalization ability of the model. DropEdge removes some edges of the graphs randomly and reduces the convergence speed of over-smoothing, which is also very convenient to be adopted in any GNN backbones.

As mentioned in the section of graph construction, the gallery images are chosen based on their similarity to the query images and, therefore, contain imbalance positive and negative samples. Hence, we train the GNN with Focal loss [56] as a loss function to handle the class imbalance of the nodes, which is as follows:

$$FL(p) = \begin{cases} -\alpha(1-p)^{\gamma}\log(p), & y=1 \\ -(1-\alpha)p^{\gamma}\log(1-p), & \text{otherwise} \end{cases} \tag{8}$$

where p is the predicted probabilities of each node, α and γ are hyperparameters. Adjusting α controls the ratio of loss caused by positive or negative samples. On the other hand, a larger value of γ will allow the model to focus more on hard samples and pay less attention to the easy samples.

3) Content-Based Retrieval: The final step of content-based retrieval (CBR) is to select top k images from the ranking list and return them as results. Therefore, a reliable ranking list is necessary to achieve good performance. In this paper, we utilize the outputs from the sigmoid function in the final layer as a ranking list representing the similarity score of each query-gallery pair. We return top k images in the ranking list as retrieved images after sorting the list into descending order so that most similar images ranked top.

IV. Experiment

A. Dataset

Experiments are conducted on two datasets, including the CE-MRI dataset [57] and a public dataset [58] from the Kaggle platform (Kaggle dataset for short), to verify the merits of the proposed MRCG. The MRI image size of the CE-MRI dataset is 512×512, which consists of axial, coronal and sagittal views. The MRI images are collected from 233 patients with the total images reaches 3049 images. The dataset consists of 3 brain tumor classes, namely meningioma, glioma and pituitary tumors, where the number of samples for each class is listed in Table I.

TABLE I
DETAILS OF THE CE-MRI DATASET.

Tumor type	Number of images
Meningioma	708
Glioma	1426
Pituitary	915
Total	3049

In addition, the Kaggle dataset is a combination of T1, T2, and FLAIR types MRI scans with 3264 brain MRI images in total. It contains 4 classes, including the same 3 tumor types from CE-MRI and images without tumors, where the number of samples for each class is listed in Table II.

TABLE II
DETAILS OF THE KAGGLE DATASET.

Tumor type	Number of images
Meningioma	937
Glioma	926
Pituitary	901
No tumor	500
Total	3264

B. Implementation

The main framework of the proposed MRCG combines the training of both the triplet CNN network and GNN. Input images to CNN are resized to 224×224. CNN and GNN are both trained together for 100 epochs. The batch size used to train CNN and GNN are 32 and 64, respectively. Adam is used as an optimizer, with a learning rate of 0.0001 on training CNN and a learning rate of 0.0005 on training GNN. The weight of cross-entropy loss on CNN, λ equal to 1 following [59]. The margin of triplet loss is set to be 50. The hyperparameters of focal loss are set as $\gamma = 2$ and $\alpha = 0.25$ following [56]. The drop rate used for both Dropout and DropEdge is 0.5. The random seed is set to be 1 for each experiment. Both datasets are divided into the training set and test set in the ratio of 7:3. Our approach is implemented in the Pytorch and Pytorch Geometric [60] frameworks.

C. Evaluation Metrics

We evaluate the models with mean average precision (MAP). Firstly, precision is defined as the ratio of retrieved images that are relevant to the query image over the retrieved images:

$$precision = \frac{|\{relevant\ images\} \cap \{retrieved\ images\}|}{|\{retrieved\ images\}|} \tag{9}$$

Average precision is defined as

$$\text{AP@}n = \frac{1}{\text{TTP}} \sum_{k=1}^{n} P@k \times rel@k \tag{10}$$

where TTP denotes the number of ground truth positives, $P@k$ denotes precision at rank k, $rel@k = 1$ if the image at rank k is relevant and n is the maximum number of ranks considered.

Besides, the MAP is formulated as:

$$\text{MAP} = \frac{1}{Q} \sum_{i=1}^{Q} \text{AP}_i \tag{11}$$

where Q denotes the number of query images. Therefore, MAP represents the mean of AP over total queries made.

D. Competing Methods

To evaluate the performance of the proposed MRCG, we compare our approach with five baseline models, which include a model using solely Siamese CNN [61] and four models that combine Siamese CNN and classic GNN methods. The Siamese CNN uses the same pre-trained Vgg16 as the

backbone, which is only used for similarity learning with the contrastive loss [62] as a loss function. Furthermore, the input graph of the baseline GNNs is constructed the same way as the proposed MRCG, with 50 nodes in each graph. All baseline models are trained for 100 epochs using Adam optimizer.

The GNN models used for comparisons include four classic spectral-based GNN methods: (1) ChebNet [47]; (2) GCN [48] as well as spatial-based GCN; (3) GraphSage [49]; (4) Graph Attention Network (GAT) [50]. For easy understanding, we use "+" to represent the baseline model that integrates both CNN and GNN, e.g., Siamese + GAT indicates a model that combines Siamese CNN and GAT.

Both Siamese + ChebNet and Siamese + GCN baseline models have similar settings, which consist of two graph convolution layers with L2 regularization only on the first layer. The models are trained with a learning rate of 0.0005 on CNN and a learning rate of 0.0001 on GNN. Besides, ChebNet is implemented using second-order Chebyshev polynomial by setting $K = 2$.

On the other hand, the Siamese + GraphSage baseline has two graph convolution layers and is trained with a learning rate of 0.0005 on CNN and a learning rate of 0.001 on GraphSage. Besides, Siamese + GAT adopts similar settings as the inductive learning task setting in [50]. The model consists of three GAT layers with skip connections. In addition, multi-head mechanism is adopted where the first two layers consist of 4 attention heads and 6 attention heads for the final layer. The model is trained with a learning rate of 0.0005 on CNN and a learning rate of 0.0001 on GAT.

E. Results

The MAP in the CE-MRI dataset achieved by the proposed MRCG and other state-of-the-art models is reported as shown in Table III, whose results are calculated by averaging the MAP over the last 10 epochs. Similarly, the experimental results are shown in Table IV. Table V and Table VI demonstrate the performance of the proposed method on each class of the CE-MRI and the Kaggle datasets, respectively. Fig. 6 and Fig. 7 show examples of query and retrieve results on the CE-MRI and the Kagge datasets, respectively, where we demonstrate the retrieval process by selecting an image from each tumor class as query images. The top 10 images in the ranking list ranked by our MRCG method are returned as retrieved images. The false retrievals are marked with a red bounding box.

F. Discussions

From Table III, it shows no significant improvements from baseline models that combine CNN and GNN compared with the model using Siamese CNN only. Siamese + ChebNet achieves the best performance among the baseline models, scoring 87.04% MAP, 84.15% precision@5, and 83.75% precision@10. This indicates the spectral GCN achieves better performance on the CE-MRI dataset in this retrieval task. Nevertheless, our method achieved the best performance, scoring 88.64% MAP, 87.17% precision@5, and 86.60% precision@10. Compared with the second-best results from Siamese + ChebNet, the proposed MRCG can succeed by about 1.6% MAP, 3% precision@5, and 3% precision@10, which shows a significant improvement from our work.

Table IV shows experiments conducted on the Kaggle dataset, where baseline models that combine CNN and GNN achieve 1-4% improvement compared with the model using Siamese CNN only. The Siamese + GAT baseline model achieves the best 82.05% MAP results among the baseline models, which shows the attention mechanism and skip connection implemented in the GAT baseline is beneficial to achieve a good retrieval performance. However, our method outperforms all baseline models significantly, showing an improvement of 6-8% MAP compared with the baseline models without the attention mechanism. Furthermore, our model achieves 6-11% better performance in terms of precision@5 and precision@10 compared with all the baseline models. The proposed MRCG is the only model scores above 80% in terms of precision metrics. Although the Kaggle dataset, which consists of 4 classes, is a challenging task for counterparts, our proposed MRCG can still maintain a good performance.

TABLE III
COMPARISON WITH BASELINE MODELS ON THE CE-MRI DATESET.

Models	MAP (%)	Prec@5 (%)	Prec@10 (%)
Siamese	86.01 ± 0.44	82.65 ± 0.75	82.79 ± 0.59
Siamese + ChebNet	87.04 ± 0.69	84.15 ± 0.94	83.75 ± 0.65
Siamese + GCN	86.42 ± 0.96	83.58 ± 1.10	83.24 ± 0.77
Siamese + GraphSage	86.50 ± 0.84	83.46 ± 0.99	83.44 ± 0.81
Siamese + GAT	86.27 ± 0.73	82.87 ± 0.79	82.98 ± 0.81
MRCG (Ours)	$\mathbf{88.64 \pm 0.64}$	$\mathbf{87.17 \pm 0.73}$	$\mathbf{86.60 \pm 0.65}$

TABLE IV
COMPARISON WITH BASELINE MODELS ON THE KAGGLE DATESET.

Models	MAP (%)	Prec@5 (%)	Prec@10 (%)
Siamese	78.15 ± 0.58	73.28 ± 0.84	72.91 ± 0.66
Siamese + ChebNet	80.51 ± 1.17	76.13 ± 1.60	76.03 ± 1.35
Siamese + GCN	79.62 ± 1.07	74.85 ± 1.33	74.80 ± 1.18
Siamese + GraphSage	80.75 ± 0.93	76.17 ± 1.06	75.67 ± 0.98
Siamese + GAT	82.05 ± 0.96	78.34 ± 1.28	77.70 ± 1.17
MRCG (Ours)	$\mathbf{86.59 \pm 0.35}$	$\mathbf{84.38 \pm 0.67}$	$\mathbf{83.68 \pm 0.70}$

TABLE V
PERFORMANCE OF THE PROPOSED METHOD ON EACH CLASS OF THE CE-MRI DATESET.

Tumor type	MAP (%)	Prec@5 (%)	Prec@10 (%)
Meningioma	79.82	77.65	75.52
Glioma	88.09	86.33	86.52
Pituitary	95.34	94.57	94.66

TABLE VI
PERFORMANCE OF THE PROPOSED METHOD ON EACH CLASS OF THE KAGGLE DATESET.

Tumor type	MAP (%)	Prec@5 (%)	Prec@10 (%)
Meningioma	77.37	71.47	73.33
Glioma	80.38	77.07	76.80
Pituitary	92.84	91.73	91.80
No tumor	93.70	92.93	92.00

In addition, Table V and Table VI show that both meningioma and glioma tumors achieve a lower score. The reason might be because meningioma and glioma tumors have a

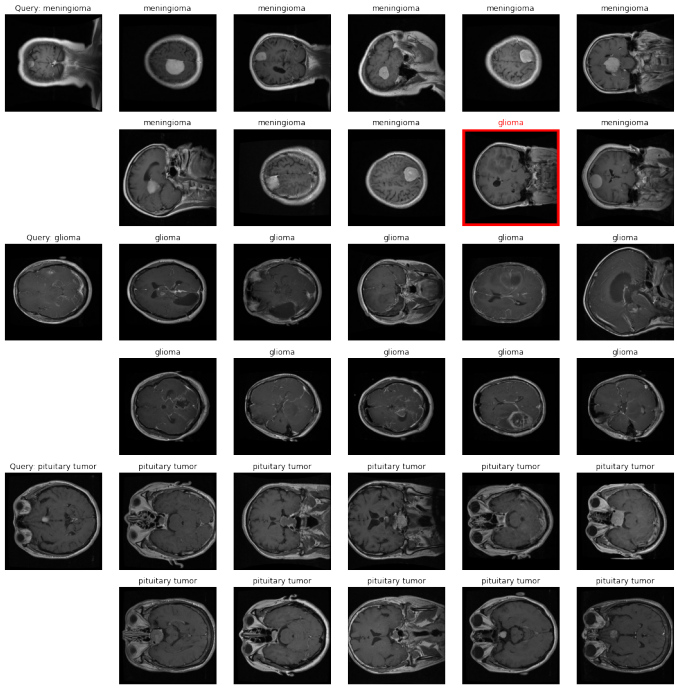

Fig. 6. Retrieval results for the CE-MRI dataset on 3 classes. 10 images are retrieved for each query. The first column represents the query image, whereas the rest are the retrieved images for each query. False retrieval is marked with a red bounding box.

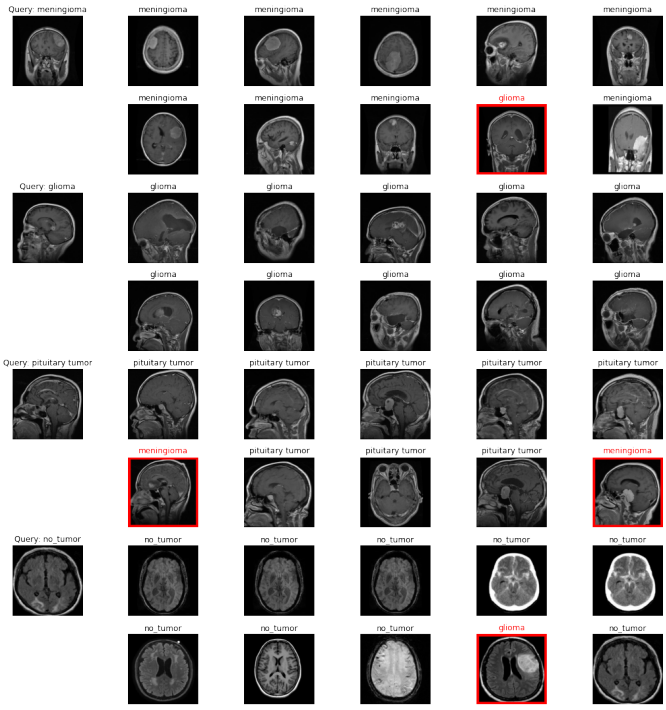

Fig. 7. Retrieval results for the Kaggle dataset on 4 classes. 10 images are retrieved for each query. The first column represents the query image, whereas the rest are the retrieved images for each query.

similar appearance. Therefore, the retrieval performance is affected on the two tumors. Specifically, for the pituitary tumors in the CE-MRI dataset, the proposed MRCG can achieve over 90% in terms of MAP and precision metrics.

Table VI shows images with no tumor score highest, although the sample size is smallest among other classes. This implies that brain images without tumors have a distinctive appearance. Specifically, the proposed MRCG can achieve 92.84% MAP, 91.73% precision@5, and 91.80% precision@10 for Pituitary tumors. Furthermore, it has 93.70% MAP, 92.93% precision@5, and 92.00 precision@10 for no tumor, which is even higher than that of pituitary tumors and shows the outstanding retrieval performance of the proposed MRCG.

From the Fig. 6, both the retrieval results of glioma and pituitary tumors show no error in the top 10 images. For meningioma tumors, only the ninth in the ranking list gives the wrong one, which shows a good performance on the CE-MRI dataset.

Fig. 7 shows examples of query and retrieve results on the Kaggle dataset, which contains four classes and is a relatively hard dataset compared with the CE-MRI dataset. From the figure, retrieval results of pituitary tumors have two wrong results in the sixth and tenth, both of which should belong to meningioma. For meningioma and no tumors, only one wrong result emerges, which is in the ninth of the ranking list and should be glioma tumors. Besides, it shows no error in the top 10 retrieval results for glioma tumors, which verifies the good performance of the proposed MRCG on the Kaggle dataset.

Besides the overall performance of the proposed MRCG, the node class distributions of the CE-MRI and the Kaggle datasets are shown in Fig. 8 and Fig. 9, respectively, to verify the class imbalance problems. In Fig. 8, 10 queries are chosen from each class and the top 50 most similar gallery images are selected for each query. For meningioma tumors which including a total of 500 gallery images, only 136 samples belong to different classes and the other 364 belong to the same one. Similar to meningioma tumors, it has 392 and 465 samples belonging to the same class for glioma and pituitary tumors, respectively, which is much higher than that belonging to the different classes from a total of 500 gallery images. Hence, the class imbalance problem in node classification exists, which is why the Focal loss is adopted in the GNN.

Similar to the CE-MRI dataset, the node class distribution of the Kaggle dataset is shown in Fig. 9. The class imbalance problem still exists, especially for glioma, pituitary tumors and no tumor. Specifically, only 71, 98, and 57 samples from 500 gallery images belong to different classes for glioma, pituitary tumors and no tumor, respectively. Most samples belong to the same class for all the four classes in the Kaggle dataset. Hence, the Focal loss in this work can help combat the node class imbalance problem and effectively improve retrieval performance.

V. Conclusion

In this paper, an MRI retrieval system for IoMT utilizing the relationship between multiple gallery images to overcome

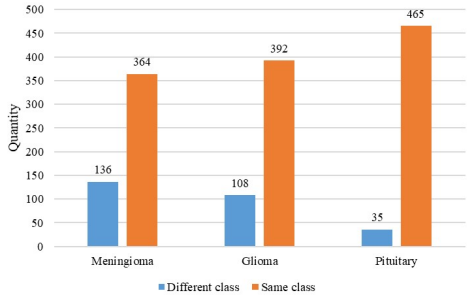

Fig. 8. Node class distribution on the CE-MRI dataset. 10 queries are selected from each class and the top 50 most similar gallery images are chosen for each query, e.g., 10 queries of meningioma tumors are selected and therefore, a total of 500 gallery images are chosen, with 136 samples being different classes and 364 samples being same class.

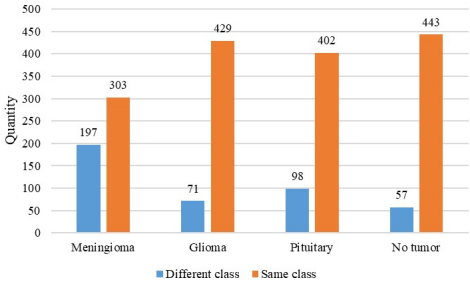

Fig. 9. Node class distribution on the Kaggle dataset. 10 queries are selected from each class and the top 50 most similar gallery images are chosen for each query, e.g., 10 queries of meningioma tumors are selected and therefore, a total of 500 gallery images are chosen, with 197 samples being different classes and 303 samples being same class.

the limitations of traditional similarity learning. By modeling the relationship between images using graph structure, a GNN is utilized to learn the node representations. In addition, some modifications have been made to the network architecture of the triplet CNN and GNN to improve the retrieval performance, of which Focal loss is adopted to handle the node class imbalance problem and focus on the loss induced by hard samples during GNN training. Experimental results on benchmark datasets show that the proposed method outperforms all other counterparts, with a significant improvement in the experiments on the CE-MRI and Kaggle datasets, which lays a good foundation for secure and private IoMT applications. Further work can be conducted on cross-modality medical image retrieval for more applications. Besides, learning discriminative node representations can be included to further improve retrieval performance in the future. And how to build an end-to-end system based on the proposed MRCG is also a challenge. One of the feasible ways might be to fuse the graph construction process into the CNN or GNN.

REFERENCES

[1] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, and T. Sato, "Deep-learning-empowered breast cancer auxiliary diagnosis for 5gb remote e-health," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 54–61, 2021.

[2] L. Tan, K. Yu, N. Shi, C. Yang, W. Wei, and H. Lu, "Towards secure and privacy-preserving data sharing for covid-19 medical records: A blockchain-empowered approach," *IEEE Transactions on Network Science and Engineering*, 2021.

[3] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for iiot devices," *IEEE Transactions on Industrial Informatics*, 2021.

[4] H. Ghayvat, S. N. Pandya, P. Bhattacharya, M. Zuhair, M. Rashid, S. Hakak, and K. Dev, "Cp-bdhca: Blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[5] Y. Sun, J. Liu, K. Yu, M. Alazab, and K. Lin, "Pmrss: Privacy-preserving medical record searching scheme for intelligent diagnosis in iot healthcare," *IEEE Transactions on Industrial Informatics*, 2021.

[6] K. Dev, S. A. Khowaja, A. S. Bist, V. Saini, and S. Bhatia, "Triage of potential covid-19 patients from chest x-ray images using hierarchical convolutional networks," *Neural Computing and Applications*, pp. 1–16, 2021.

[7] H. Li, K. Yu, B. Liu, C. Feng, Z. Qin, and G. Srivastava, "An efficient ciphertext-policy weighted attribute-based encryption for the internet of health things," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[8] L. Zhang, M. Peng, W. Wang, Z. Jin, Y. Su, and H. Chen, "Secure and efficient data storage and sharing scheme for blockchain-based mobile-edge computing," *Transactions on Emerging Telecommunications Technologies*, p. e4315, 2021.

[9] E. Q. Wu, M. Zhou, D. Hu, L. Zhu, Z. Tang, X.-Y. Qiu, P.-Y. Deng, L.-M. Zhu, and H. Ren, "Self-paced dynamic infinite mixture model for fatigue evaluation of pilots' brains," *IEEE Transactions on Cybernetics*, 2020.

[10] E. Q. Wu, L.-M. Zhu, G.-J. Li, H.-J. Li, Z. Tang, R. Hu, and G.-R. Zhou, "Nonparametric hierarchical hidden semi-markov model for brain fatigue behavior detection of pilots during flight," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[11] L. Zhang, T. Huang, X. Hu, Z. Zhang, W. Wang, D. Guan, C. Zhao, and S. Kim, "A distributed covert channel of the packet ordering enhancement model based on data compression," *CMC-COMPUTERS MATERIALS & CONTINUA*, vol. 64, no. 3, pp. 2013–2030, 2020.

[12] E. Q. Wu, C.-T. Lin, L.-M. Zhu, Z. Tang, Y.-W. Jie, and G.-R. Zhou, "Fatigue detection of pilots' brain through brains cognitive map and multilayer latent incremental learning model," *IEEE Transactions on Cybernetics*, 2021.

[13] W. Wang, C. Qiu, Z. Yin, G. Srivastava, T. R. Gadekallu, F. Alsolami, and C. Su, "Blockchain and puf-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, 2021.

[14] W. Wang, H. Huang, L. Zhang, and C. Su, "Secure and efficient mutual authentication protocol for smart grid under blockchain," *Peer-to-Peer Networking and Applications*, vol. 14, no. 5, pp. 2681–2693, 2021.

[15] E. Q. Wu, P. Xiong, Z.-R. Tang, G.-J. Li, A. Song, and L.-M. Zhu, "Detecting dynamic behavior of brain fatigue through 3-d-cnn-lstm," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

[16] J. Zhang, Y. Xie, Q. Wu, and Y. Xia, "Medical image classification using synergic deep learning," *Medical image analysis*, vol. 54, pp. 10–19, 2019.

[17] E. Q. Wu, P.-Y. Deng, X.-Y. Qu, Z. Tang, W.-M. Zhang, L.-M. Zhu, H. Ren, G.-R. Zhou, and R. S. Sheng, "Detecting fatigue status of pilots based on deep learning network using eeg signals," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[18] W. Zhao, L. Shen, B. Han, Y. Yang, K. Cheng, D. A. Toesca, A. C. Koong, D. T. Chang, and L. Xing, "Markerless pancreatic tumor target localization enabled by deep learning," *International Journal of Radiation Oncology* Biology* Physics*, vol. 105, no. 2, pp. 432–439, 2019.

[19] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, no. 141, p. 20170387, 2018.

[20] D. A. Kaji, J. R. Zech, J. S. Kim, S. K. Cho, N. S. Dangayach, A. B. Costa, and E. K. Oermann, "An attention based deep learning model of clinical events in the intensive care unit," *PloS one*, vol. 14, no. 2, p. e0211057, 2019.

[21] Y.-A. Chung and W.-H. Weng, "Learning deep representations of medical images using siamese cnns with application to content-based image retrieval," *arXiv preprint arXiv:1711.08490*, 2017.

[22] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu, "Learning fine-grained image similarity with deep ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1386–1393.

[23] Y. Yan, Q. Zhang, B. Ni, W. Zhang, M. Xu, and X. Yang, "Learning context graph for person search," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2158–2167.

[24] K. Dinakaran, K. Adinadh, K. Sanjuna, and P. Valarmathie, "Quality of service (qos) and priority aware models for adaptive efficient image retrieval in wsn using tbl routing with rlbp features," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 4137–4146, 2021.

[25] K. Greeshma and J. Viji Gripsy, "A review on classification and retrieval of biomedical images using artificial intelligence," *The Fusion of Internet of Things, Artificial Intelligence, and Cloud Computing in Health Care*, pp. 47–66, 2021.

[26] R. O. Ogundokun, J. B. Awotunde, E. A. Adeniyi, and F. E. Ayo, "Crypto-stegno based model for securing medical information on iomt platform," *Multimedia tools and applications*, pp. 1–23, 2021.

[27] M. Huang, W. Yang, M. Yu, Z. Lu, Q. Feng, and W. Chen, "Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced mri images," *Computational and mathematical methods in medicine*, vol. 2012, 2012.

[28] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou, R. Yang, J. Zhao, Y. Feng, Q. Feng *et al.*, "Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation," *PloS one*, vol. 11, no. 6, p. e0157112, 2016.

[29] W. Yang, Q. Feng, M. Yu, Z. Lu, Y. Gao, Y. Xu, and W. Chen, "Content-based retrieval of brain tumor in contrast-enhanced mri images using tumor margin information and learned distance metric," *Medical physics*, vol. 39, no. 11, pp. 6929–6942, 2012.

[30] M. Huang, W. Yang, Y. Wu, J. Jiang, Y. Gao, Y. Chen, Q. Feng, W. Chen, and Z. Lu, "Content-based image retrieval using spatial layout information in brain tumor t1-weighted contrast-enhanced mr images," *PloS one*, vol. 9, no. 7, p. e102754, 2014.

[31] E. Q. Wu, D. Hu, P.-Y. Deng, Z. Tang, Y. Cao, W.-M. Zhang, L.-M. Zhu, and H. Ren, "Nonparametric bayesian prior inducing deep network for automatic detection of cognitive status," *IEEE transactions on cybernetics*, 2020.

[32] Z. Tang, R. Zhu, R. Hu, Y. Chen, E. Q. Wu, H. Wang, J. He, Q. Huang, and S. Chang, "A multilayer neural network merging image preprocessing and pattern recognition by integrating diffusion and drift memristors," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[33] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[34] A. Rehman, M. A. Khan, T. Saba, Z. Mehmood, U. Tariq, and N. Ayesha, "Microscopic brain tumor detection and classification using 3d cnn and feature selection architecture," *Microscopy Research and Technique*, vol. 84, no. 1, pp. 133–149, 2021.

[35] X. Liu, H. R. Tizhoosh, and J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 2872–2878.

[36] M. Chowdhury, S. R. Bulo, R. Moreno, M. K. Kundu, and Ö. Smedby, "An efficient radiographic image retrieval system using convolutional neural network," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 3134–3139.

[37] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Content-based brain tumor retrieval for mr images using transfer learning," *IEEE Access*, vol. 7, pp. 17 809–17 822, 2019.

[38] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, 2017.

[39] S. K. Sundararajan, B. Sankaragomathi, and D. S. Priya, "Deep belief cnn feature representation based content based image retrieval for medical images," *Journal of medical systems*, vol. 43, no. 6, pp. 1–9, 2019.

[40] K. Kruthika, H. Maheshappa, A. D. N. Initiative *et al.*, "Cbir system using capsule networks and 3d cnn for alzheimer's disease diagnosis," *Informatics in Medicine Unlocked*, vol. 14, pp. 59–68, 2019.

[41] S. Deepak and P. Ameer, "Retrieval of brain mri with tumor using contrastive loss based similarity on googlenet encodings," *Computers in Biology and Medicine*, vol. 125, p. 103993, 2020.

[42] M. Owais, M. Arsalan, J. Choi, and K. R. Park, "Effective diagnosis and treatment through content-based medical image retrieval (cbmir) by using artificial intelligence," *Journal of clinical medicine*, vol. 8, no. 4, p. 462, 2019.

[43] G. S. Tandel, A. Balestrieri, T. Jujaray, N. N. Khanna, L. Saba, and J. S. Suri, "Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm," *Computers in Biology and Medicine*, vol. 122, p. 103804, 2020.

[44] D. J. Hemanth and J. Anitha, "Modified genetic algorithm approaches for classification of abnormal magnetic resonance brain tumour images," *Applied Soft Computing*, vol. 75, pp. 21–28, 2019.

[45] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images," *Cognitive Systems Research*, vol. 57, pp. 147–159, 2019.

[46] M. Adnan, S. Kalra, and H. 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.

[47] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in neural information processing systems*, vol. 29, pp. 3844–3852, 2016.

[48] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[49] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 1025–1035.

[50] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[51] Y. Gu, K. Vyas, M. Shen, J. Yang, and G.-Z. Yang, "Deep graph-based multimodal feature embedding for endomicroscopy image retrieval," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 2, pp. 481–492, 2020.

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

[53] 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*, 2016, pp. 770–778.

[54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[55] Y. Rong, W. Huang, T. Xu, and J. Huang, "Dropedge: Towards deep graph convolutional networks on node classification," in *International Conference on Learning Representations*, 2019.

[56] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[57] J. Cheng, "Brain tumor dataset," [Online], https://figshare.com/articles/dataset/brain_tumor_dataset/1512427 Accessed April 2, 2017.

[58] B. Bhuvaji, "Brain tumor classification (mri)," [Kaggle], https://www.kaggle.com/dsv/1183165 Accessed May 12, 2020.

[59] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *International workshop on similarity-based pattern recognition*. Springer, 2015, pp. 84–92.

[60] M. Fey and J. E. Lenssen, "Fast graph representation learning with pytorch geometric," *arXiv preprint arXiv:1903.02428*, 2019.

[61] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 539–546.

[62] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1735–1742.