IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 3, 2018, accepted January 6, 2019, date of publication January 14, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2892455

Content-Based Brain Tumor Retrieval for MR Images Using Transfer Learning

ZAR NAWAB KHAN SWATI[1,2], QINGHUA ZHAO[3], MUHAMMAD KABIR[1], FARMAN ALI[1], ZAKIR ALI[1], SAEED AHMED[1], AND JIANFENG LU[1]

[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[2]Department of Computer Science, Karakoram International University, Gilgit 15100, Pakistan
[3]College of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China

Corresponding author: Jianfeng Lu (lujf@njust.edu.cn)

This work is partly supported by the National Key Research and Development Program of China (No. 2018YFB1004), the 111 Project (No.B13022) and the Natural Science Foundation of Jiangsu Province of China under Grant (No.20131351).

ABSTRACT This paper presents an automatic content-based image retrieval (CBIR) system for brain tumors on T1-weighted contrast-enhanced magnetic resonance images (CE-MRI). The key challenge in CBIR systems for MR images is the semantic gap between the low-level visual information captured by the MRI machine and the high-level information perceived by the human evaluator. The traditional feature extraction methods focus only on low-level or high-level features and use some handcrafted features to reduce this gap. It is necessary to design a feature extraction framework to reduce this gap without using handcrafted features by encoding/combining low-level and high-level features. Deep learning is very powerful for feature representation that can depict low-level and high-level information completely and embed the phase of feature extraction in self-learning. Therefore, we propose a deep convolutional neural network VGG19-based novel feature extraction framework and apply closed-form metric learning to measure the similarity between the query image and database images. Furthermore, we adopt transfer learning and propose a block-wise fine-tuning strategy to enhance the retrieval performance. The extensive experiments are performed on a publicly available CE-MRI dataset that consists of three types of brain tumors (i.e., glioma, meningioma, and pituitary tumor) collected from 233 patients with a total of 3064 images across the axial, coronal, and sagittal views. Our method is more generic, as we do not use any handcrafted features; it requires minimal preprocessing, tested as robust on fivefold cross-validation, can achieve a fivefold mean average precision of 96.13%, and outperforms the state-of-the-art CBIR systems on the CE-MRI dataset.

INDEX TERMS Brain tumor retrieval, block-wise fine-tuning, closed-form metric learning, convolutional neural networks, feature extraction, transfer learning.

I. INTRODUCTION

Rapid advancements in medical imaging technology are useful for clinical diagnosis, treatment planning, decision-making, and patient health care. In hospitals, a large amount of medical imaging data is produced every day, which is helpful for clinical decision support and can be used for research and training in the field of medical science. Recent research has shown great interest in CBIR in medical imaging, such as MRI [1]–[5], X-ray [6]–[8], CT [9], and mammogram[10].

Manual MRI retrieval from a large archive of imaging data with similar structures or appearances is a difficult and challenging task for radiologists. It depends on the availability and expertise of the radiologist, who examines MR images and retrieves the relevant images from the archived data. This manual retrieval method is impractical, non-reproducible and time-intensive for a large amount of archived data. To address this problem, automatic CBIR is a possible solution for indexing archived images with minimum intervention by radiologists. In this research, we focus on CBIR for brain tumors. Specifically, when the radiologist presents a query image, the CBIR system retrieves the same pathological type of brain tumor images from the database; then, the radiologist selects the most closely related retrieved images and accesses the related diagnosis and treatment history to support the diagnosis and treatment of the current case. The CE-MRI dataset [11] utilized in this study consists of three types of brain tumors with the highest percentage among brain tumors. In clinical practice, the incident rates of glioma, meningioma,

and pituitary tumor are approximately 45%, 15%, and 15%, respectively, among all brain tumors.

A highly accurate CBIR system depends on the feature extraction method and distance metric learning (DML). Feature extraction is a core step in traditional machine-learning methods, which can be categorized into two types. The first type is local feature [2], [12]–[16], which is based on intensity and texture features such as first-order statistics (e.g., mean, standard deviation and skewness) and second-order statistics derived from gray level co-occurrence matrix (GLMC), shape, wavelet transform, and Gabor. These features are low level, and their representation power is limited because different types of brain tumors have a similar appearance, and the same type varies in appearance aspects such as boundary, texture, size, and shape, as shown in Fig. 1. The second type is global feature extraction, such as bag-of-words (BoW) [1], [3], [6], Fisher vector (FV) [4], [17], [18] and scale-invariant feature transformation (SIFT) [10]. The statistical features extracted from BoW, FV, and SIFT are high-level features that certainly ignore spatial information.

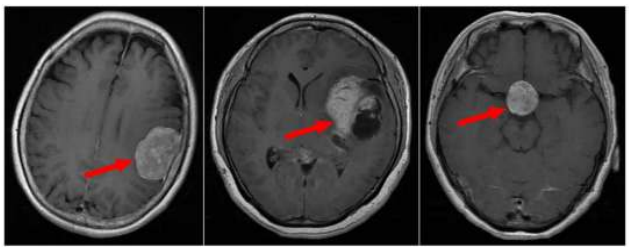


FIGURE 1. (a) and (b) are meningiomas with differing appearance, and (c) is a pituitary tumor with an appearance similar to that of (a).

Several methods have been proposed for content-based brain tumor retrieval on the T1-weighted CE-MRI dataset. Yang *et al.* [2] proposed a CBIR system using a margin information descriptor (MID) as a feature extractor. The maximum mean average precision projection (MPP) algorithm was designed to measure the similarity between the query tumor region and the database MR images. The authors achieved a *mAP* of 87.3%, which was comparatively better than that of the SIFT descriptor. Huang *et al.* [1] used a region-specific BoW model and closed-form metric learning (CFML). The BoW model was applied on the tumor boundary and tumor region separately and achieved better retrieval performance with a *mAP* of 91.0%. Huang *et al.* [3] improved the *mAP* to 91.8% by using brain tumors as a region of interest (ROI) in the region partition learning algorithm. They extracted local features from raw image patches of subregions, constructed a BoW histogram by pooling per region, and applied a DML called rank error-based metric learning (REML) for similarity measurement. Cheng *et al.* [4] used FV with adaptive spatial pooling and CFML and boosted the *mAP* to 94.68%. They used tumor region augmentation and division and extracted raw patches from subregions. PCA was applied to the subregion for dimension reduction, and the FV per subregion was then computed. The FVs of the subregions were combined to form the final single FV representation.

CBIR approaches [1]–[4], [6], [9], [10], [12]–[14] consist of several steps, including preprocessing, feature extraction (tumor region, tumor outline, feature selection, and dimension reduction) and DML. There are two problems in the feature extraction phase. First, it focuses only on either low-level or high-level features. In the CE-MRI dataset, the content of a specific category is distributed with intrinsic irregularity. There is a strong correlation between the layout of the tumor, edema, and surrounding normal tissues. Meningioma and pituitary tumor are similar in shape, as shown in Fig. 2 (a), (b), and these two tumor types are generally not related to edema. Meningioma is generally adjacent to the skull, gray matter, and cerebrospinal fluid. A pituitary tumor is near the sphenoidal sinus, internal carotid arteries, and optic chiasma. In appearance, glioma is dissimilar in shape and generally surrounded by edema, as shown in Fig. 2 (c) and (d). Second, the most important information and the discriminative features of brain tumors are related to the location/position of the tumor region in the MR image together with its boundary, texture, size, and shape. The CBIR system based on traditional machine learning uses handcrafted features (i.e., segmented the tumor region and outline), which require strong prior information (i.e., the position or location of the tumor in an image); thus, it is not a simple task and might cause inter- and intraoperator deviations [16]. Accordingly, there is a need to design such a feature extraction framework to encode/combine both low-level and high-level features.

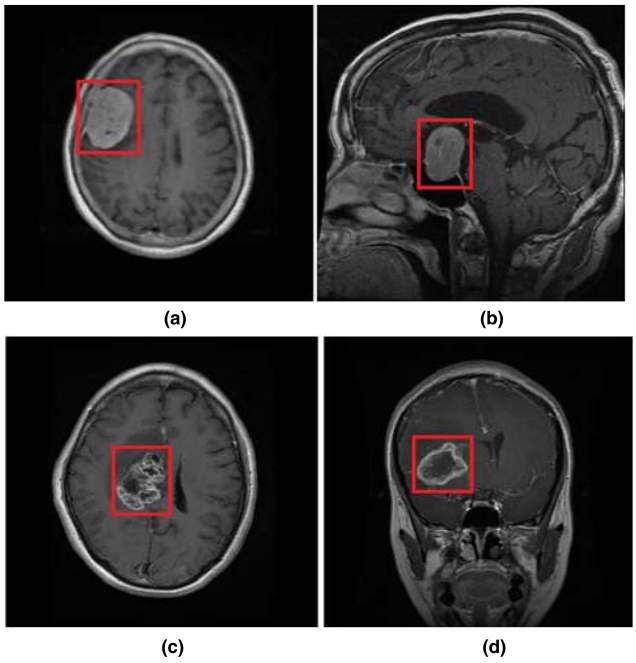


FIGURE 2. Four images of brain tumors in T1-weighted CE-MRI. The region inside the red rectangle contains a tumor. (a) Meningioma located near the skull, (b) pituitary tumor located near the sphenoidal sinus, (c) glioma containing edema and necrosis, and (c) glioma surrounded by edema.

Recent studies [19]–[23] have verified that deep learning approaches do not need manually extracted (handcrafted) features and prior domain information. Deep learning embeds the phase of feature extraction in self-learning. It needs a dataset with only minimal preprocessing, if required, and then determines significant features in a self-learning way [24]. A key challenge in CBIR systems for MR images is to reduce the semantic gap between the low-level visual information captured by the MRI machine and the high-level semantic information perceived by the human evaluator. The effectiveness of such a feature extraction framework is more important in terms of feature representations that can depict low-level and high-level information completely. To justify the feasibility of the proposed CBIR system, CNNs automatically generate powerful discriminative features using a hierarchical learning approach. Feature maps in the earlier layers extract low-level features, and feature maps in higher layers extract high-level domain (content)-specific features. Lower-layer feature maps encode simple structural information, such as edges, shape, and textures, and higher layers build atop each other and combine these low-level feature maps to encode/construct abstract representation, which integrates local and global information.

Deep learning outperformed state-of-the-art methods in the field of machine learning. In particular, enhanced performance in computer vision encouraged the use of deep learning in medical image analysis [25], classification [26], segmentation [27], [28], fusion [29], computer-based diagnosis and prediction [30], [31], lesion/landmark detection [32]–[34], microscopic image examination [35], [36], and CBIR [5], [7], [8].

CNNs have been used for decades but were not popular until Krizhevsky *et al.* [37] employed the deep learning approach (AlexNet) and won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2012. Simonyan and Zisserman [38] introduced a similar but deeper CNN (VGG Net) and secured first place in the localization task and second place in the classification task in the ILSVRC in 2014. Deep learning, and especially CNNs, has achieved success due to the advancement of computational technologies such as powerful GPUs, the development of learning algorithms [39]–[43], and the availability of big data [44]–[46].

CNNs have shown good performance in computer vision on large labeled datasets, such as ImageNet [46], which contains more than one million labeled images. However, it is difficult to apply such deep CNNs in the medical imaging field. First, the sample size of the medical image dataset is usually small because such images require the availability of expert radiologists to manually examine and label them, which is time-consuming, laborious and costly. Second, training deep CNN is a complicated task for a small dataset because of overfitting and convergence problems. Third, domain expertise is required to repeatedly revise the model and adjust the learning parameters of the model to assure that all layers can learn at an equivalent rate. Therefore, training deep CNN from scratch is a challenging task that is tedious and time-consuming and demands much diligence and patience. For the small dataset scenario, a favorable substitute for training CNN from scratch is to use a pretrained CNN and adopt a transfer learning and fine-tuning strategy [47].

Pretrained CNN models have been effectively used in computer vision applications as feature extractors or as a baseline for transfer learning [48]–[50]. The main advantage of CNNs is the ''transferability'' of the knowledge learned in pretrained models. Several existing methods [5], [7], [8], [26], [47], [51], and [52] have proposed different transfer learning approaches for medical imaging CBIR using CNNs. Most of them used an off-the-shelf trained model over a large dataset of natural images, extracted features from a specified layer of a pretrained model for the new dataset, and trained a separate learning method for classification and retrieval.

In this research, we developed a content-based brain tumor retrieval system for MR images using a pretrained deep CNN model (VGG19). We adopt transfer learning, propose a block-wise fine-tuning strategy for feature extractions and use CFML to measure the similarity distance. The proposed method is evaluated on a publicly available CE-MRI dataset. We performed numerous experiments for brain tumor MR image retrieval, used a five-fold cross-validation test to ensure robustness, evaluated the performance, and compared our results with state-of-the-art brain tumor retrieval on the same CE-MRI dataset. To the best of our knowledge, this is the first deep learning-based work for brain tumor retrieval on a CE-MRI dataset.

The rest of the paper is organized as follows. Section II discusses the proposed research framework and methodology in detail. The experimental settings, parameter optimization, retrieval performance, results and comparisons are shown in Section III. A brief discussion is provided in Section IV, and the conclusion is presented in Section V.

II. PROPOSED METHOD

This paper proposed an automatic CBIR for retrieving similar brain tumor images from a database. Fig. 3 presents the detailed research framework of the proposed method. For feature extraction, we used the VGG19 [38] pretrained on a large ImageNet dataset (more than 1.2 million labeled images). CFML is applied to measure the similarity distance between the extracted features of the database images and the test/query image.

We extracted features from the fully connected layer (FC7) of VGG19 and fed them into CFML, as shown in Fig. 4. The contents of the CE-MRI and pretrained VGG19 datasets are different. Features extracted from the higher layer of the pretrained VGG19 did not produce satisfactory results because higher layers in the network are related to the content-specific features of the image learned from early layers in the network. Therefore, we fine-tuned the VGG19 on the CE-MRI dataset in a block-wise manner and observed the incremental performance improvement. This transfer learning and fine-tuning suggested an alternative to [5], [7], [8], [22], [51], and [52],

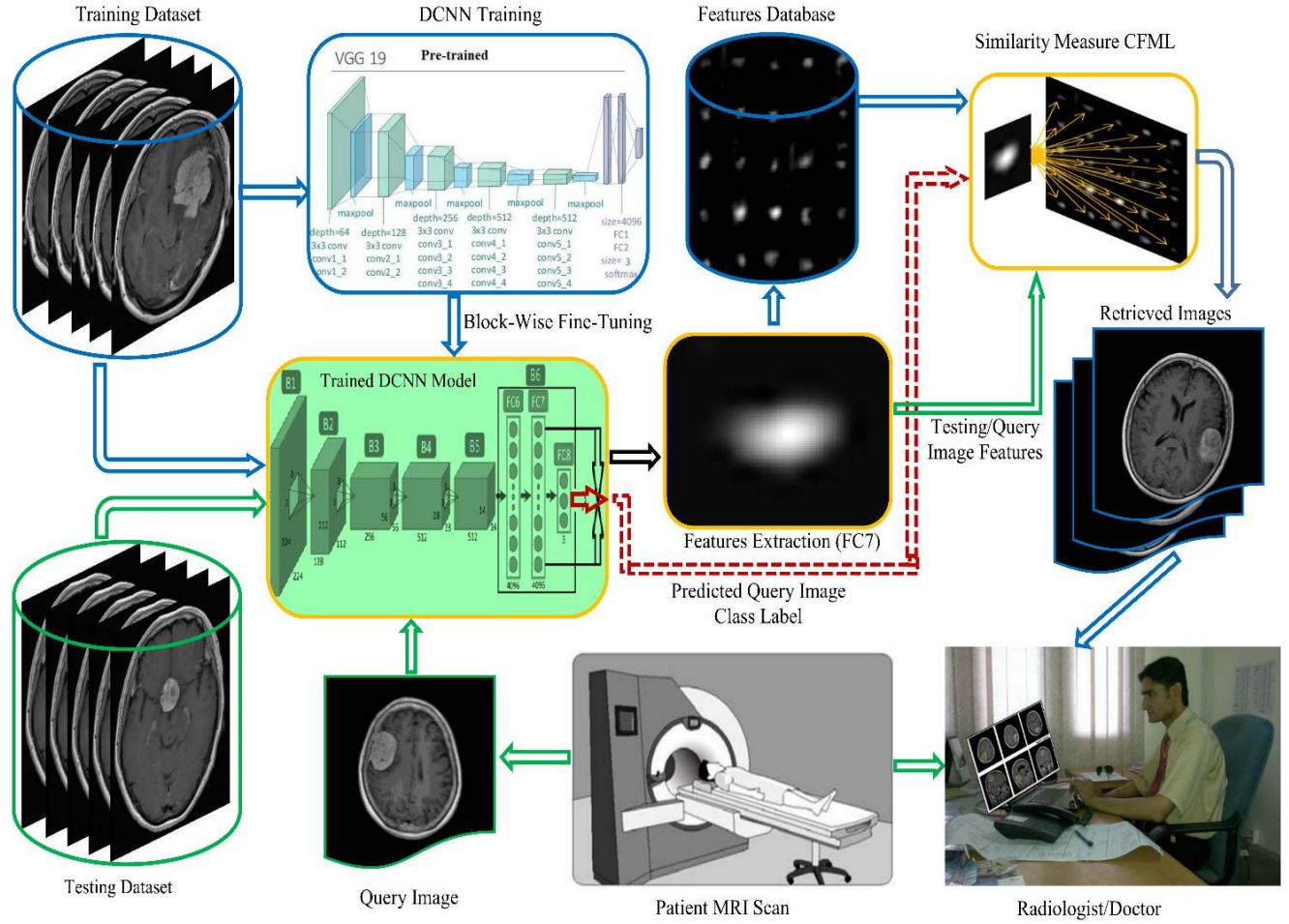


FIGURE 3. The proposed framework for CBIR using pretrained deep CNN (VGG19) and CFML. We trained the VGG19 by fine-tuning on the CE-MRI dataset. Once the model was successfully optimized and trained, we extracted features of the training and testing datasets from the trained model. We applied CFML to measure the similarity between the features of the database images and testing/query images. Predicting query image class labels is optional (dashed line). It will help the radiologist identify tumor types in uncertain cases and enable fast retrieval from the relevant search area of the database.

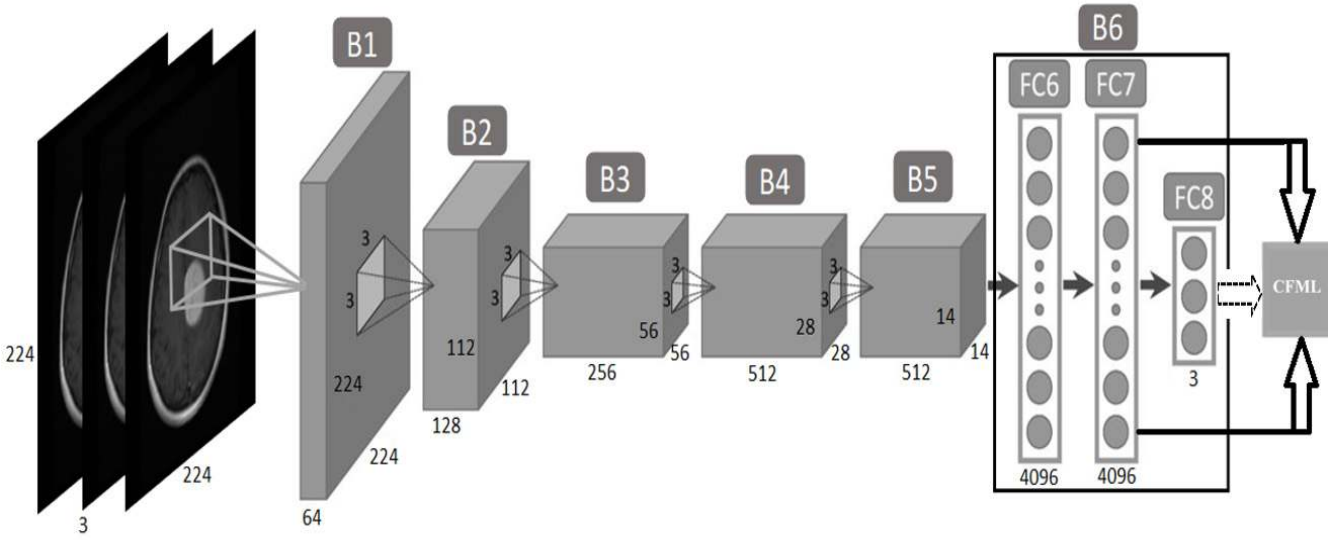


FIGURE 4. The proposed block-wise VGG19 architecture. B represents the group of layers placed in blocks (shown in Table 1). Bold lines with up-down arrows towards CFML indicate the feature extraction phase. Feature extraction starts after completion of the training process. The arrow between FC8 and CFML is optional and is used to predict the class label of the query image.

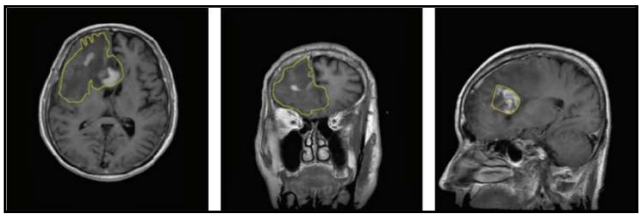


FIGURE 5. Three views of a glioma in a patient. The area inside the yellow outline indicates a tumor. (a) Axial (b) Coronal (c) Sagittal.

TABLE 2. Details of the CE-MRI dataset.

Tumor type	Number of patients	Number of MR images	MRI view	Number of MR images
Meningioma	82	708	axial	209
			coronal	268
			sagittal	231
Glioma	89	1426	axial	494
			coronal	437
			sagittal	495
Pituitary	62	930	axial	291
			coronal	319
			sagittal	320
Total	233	3064		3064

in which pretrained CNNs were used as off-the-shelf feature extractors. The working mechanism of the proposed research framework is presented in the following subsections.

A. DATASET

In this research, we used a publicly available CE-MRI dataset available at (https://figshare.com/articles/brain_tumor_dataset/1512427). The proposed brain tumor retrieval was based on two-dimensional images (2-D slices), not 3-D volume, because in Chinese clinical practice, the acquired and available MR images are 2-D slices with a large slice gap. Therefore, our brain tumor retrieval system based on 2-D MR images for clinical application is practical. The dataset was collected during 2005-2010 from Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjin Medical University, China. The dataset contains three types of tumors (i.e., glioma, meningioma, and pituitary tumor, as shown in Fig. 1 and Fig. 2) from 233 patients with a total of 3064 images across the axial, coronal, and sagittal views, as shown in Fig. 5. Table 2 shows the details of the CE-MRI dataset. The images in the dataset are provided in matrix form. The size of each image is 512×512 pixels, and the pixel size is 0.49 mm $\times$ 0.49 mm.

B. PREPROCESSING

The T1-weighted CE-MRI data are 2-D images of size 512×512. In this work, we provided MR images directly to the CNN, and the convolutional kernel is applied to the pixel intensity in the image. The result of a convolutional kernel relies heavily on these intensity values. However, intensity values in MR images do not have a fixed meaning, and it has been observed that intensity values across MR images vary greatly within or between subjects. These intensity values of MR images are also sensitive to the acquisition conditions. Data mining and especially CNN approaches need to normalize the inputs; otherwise, the network will be ill-conditioned. In principle, normalization is performed to obtain the same range of values for each inputs into the CNN model. This can guarantee a stable convergence of weight and biases. In this scenario, intensity normalization is necessary to preprocess MR images. Therefore, we normalize the CE-MR images by using min-max normalization to scale the intensity value to [0, 1], which is computed as follows:

$$y_i = (x_i - min(x))/(max(x) - min(x)) \tag{1}$$

where y_i is the normalized intensity value against position x_i (where $i = 1 \ldots .n$) and $min(x)$ and $max(x)$ are the minimum and maximum intensity values, respectively, across the entire image. After normalization, we resize the normalized image to 224×224 and duplicate it three times to create three channels according to the input size of the pretrained VGG19.

The intensity normalization brings the intensity values within a coherent range across all the MR images and facilitates learning in the training process. Resizing the images speeds up the training process and solves the memory issue.

C. DEEP CONVOLUTIONAL NEURAL NETWORKS (CNNs)

The training of CNNs starts from the first input layer to the final classification layer in a feed-forward fashion; then, error back-propagation starts from the classification layer towards the first convolutional layer. Neuron i in layer l receives input from neuron j of layer l-1 in a forward pass computed as follows:

$$In_i^l = \sum_{j=1}^{n} W_{ij}^l x_j + b_i \tag{2}$$

The output is computed by a nonlinearity ReLu function:

$$out_i^l = max(0, \mathrm{In}_i^l) \tag{3}$$

All neurons in the convolutional and fully connected layers use equations (2) and (3) to calculate the input and produce output in the form of nonlinear activation. The pooling layer uses a $K \times K$ square window sliding on the $N \times N$ feature map and takes the maximum or average value of the features inside the window. It decreases the spatial size of the feature map from $N \times N$ to $\frac{N}{K} \times \frac{N}{K}$ as it produces a single value from the $K \times K$ region.

The final layer computes the classification probability of each tumor type using the Softmax function:

$$out_i^l = \frac{e^{In_i^l}}{\sum_i e^{out_k^l}} \tag{4}$$

CNNs are trained with the back-propagation algorithm by minimizing the following cost function with respect to the unknown weights W:

$$C = -\frac{1}{m} \sum_i^m \ln \left(p \left(y^i \mid X^i \right) \right) \tag{5}$$

where m is the total number of training samples in the training set, X^i is the i^{th} sample in the training set with the label y^i

TABLE 1. Pretrained VGG19 (CNN) Architecture and Parameters.

Block	Layer	Type	Input	Kernel	Stride	Pad	Output
	input	Image Input	3x224x224	N/A	N/A	N/A	3x224x224
B_1	conv1_1	Convolution	3x224x224	3x3	1	1	64x224x224
	conv1_2	Convolution	64x224x224	3x3	1	1	64x224x224
	pool1	Max Pooling	64x224x224	2x2	2	0	64x112x112
B_2	conv2_1	Convolution	64x112x112	3x3	1	1	128x112x112
	conv2_2	Convolution	128x112x112	3x3	1	1	128x112x112
	pool2	Max Pooling	128x112x112	2x2	2	0	128x56x56
B_3	conv3_1	Convolution	128x56x56	3x3	1	1	256x56x56
	conv3_2	Convolution	256x56x56	3x3	1	1	256x56x56
	conv3_3	Convolution	256x56x56	3x3	1	1	256x56x56
	conv3_4	Convolution	256x56x56	3x3	1	1	256x56x56
	pool3	Max Pooling	256x56x56	2x2	2	0	512x28x28
B_4	conv4_1	Convolution	512x28x28	3x3	1	1	512x28x28
	conv4_2	Convolution	512x28x28	3x3	1	1	512x28x28
	conv4_3	Convolution	512x28x28	3x3	1	1	512x28x28
	conv4_4	Convolution	512x28x28	3x3	1	1	512x28x28
	pool4	Max Pooling	512x28x28	2x2	2	0	512x14x14
B_5	conv5_1	Convolution	512x14x14	3x3	1	1	512x14x14
	conv5_2	Convolution	512x14x14	3x3	1	1	512x14x14
	conv5_3	Convolution	512x14x14	3x3	1	1	512x14x14
	conv5_4	Convolution	512x14x14	3x3	1	1	512x14x14
	pool5	Max Pooling	512x14x14	2x2	2	0	512x7x7
B_6	FC6	Fully Connected	512x7x7	7x7	1	0	4096x1
	FC7	Fully Connected	4096x1	1x1	1	0	4096x1
	FC8	Fully Connected	4096x1	1x1	1	0	1000x1

and $p\left(y^i \mid X^i\right)$ is the true classification probability. The cost function C is minimized by stochastic gradient descent over the mini-batches of size N and the training cost approximated by the mini-batch cost. Consider that W_l^t represents the weights at iteration t for convolutional layer l, and $\hat{C}$ represents the mini-batch cost. Then, the updated weights in the next iteration are computed as follows:

$$\begin{aligned} \gamma^t &= \gamma^{\lfloor tN/m \rfloor} \\ V_l^{t+1} &= \mu V_l^t - \gamma^t \alpha_l \frac{\partial \hat{C}}{\partial W_l} \\ W_l^{t+1} &= W_l^t + V_l^{t+1} \end{aligned} \tag{6}$$

where a_l is the learning rate of layer l, γ is the scheduling rate that decreases the initial learning rate α at the end of a specified number of epochs, and μ is the momentum that describes the influence of previously updated weights in the current iteration.

D. DISTANCE METRIC LEARNING (DML)

Good DML plays an important role in CBIR. The performance of the CBIR system depends on the standard used for similarity measurement between the query image and database images. Substantial research has inspired good DML from the training dataset. Cosine similarity and Euclidean distance are mostly used to measure the similarity of the features, but these techniques are simple, and their feature representation power is limited because of the complexity of the image content and the semantic gap between the high-level human interpretation and low-level visual features [53], [54]. Various algorithms of distance learning [2], [3], [55]–[58] are utilized to overcome the above problem and achieve better performance for CBIR. The Mahalanobis distance method is used to determine the optimum metric, which increases intraclass similarity while decreasing interclass similarity. The squared Mahalanobis distance (SMD), also called the generalized quadratic distance, can be defined as follows:

$$\begin{aligned} d_M\left(x_i, x_j\right) &= (x_i - x_j)^T M (x_i - x_j) \\ &= (x_i - x_j)^T L^T L (x_i - x_j) \\ &= (Lx_i - Lx_j)^T (Lx_i - Lx_j) \end{aligned} \tag{7}$$

where x_i and x_j represent the feature vectors of two images. The positive semidefinite matrix (PSD) is denoted by M, while L represents a linear transformation matrix. $x_i \in R^n and M \in R^{n \times n}$. In the literature, various DML algorithms have been proposed for the better projection of M or L to minimize the objective function. We used the DML named CFML proposed by Alipanahi *et al.* [57] and achieved significant results. The pathological classes of tumors in the dataset are already known. The same-class tumors share the same label and vice versa. The feature vectors of images with identical labels are categorized as similar, "S", and the remaining feature vectors with different labels are categorized as dissimilar, "D".

$$S \rightarrow (x_i, x_j) \in S \text{ if } x_i \text{ and } x_j \text{ are similar} \tag{8}$$

$$D \rightarrow (x_i, x_j) \in D \text{ if } x_i \text{ and } x_j \text{ are not similar} \tag{9}$$

The optimum transformation matrix L^* of CFML is expressed as follows:

$$\begin{aligned} f\left(L^*\right) = \arg\min\ & (Tr(L^T\left(M_S - M_D\right)L), \\ & s.t.\ L^T M_S L = I \end{aligned} \tag{10}$$

where I denotes the identity matrix, Tr represents the trace of the matrix, and

$$M_s = \frac{1}{|S|} \sum_{(x_i,x_j)\in S} (x_i - x_j)(x_i - x_j)^T, \quad (11)$$

$$M_D = \frac{1}{|D|} \sum_{(x_i,x_j)\in D} (x_i - x_j)(x_i - x_j)^T \quad (12)$$

The CFML attempts to reduce the SMD among intraclass and increase the SMD between interclass pairs. The matrix of eigenvectors produces a closed-form solution corresponding to the largest eigenvalues of the matrix $M_s^{-1}M_D$. The regularization form of CFML, i.e., $L\,(M_S + \lambda I)\,L^T = I$, is used, where λ is a small positive value (λ set to 1.5e-4).

E. TRANSFER LEARNING AND FINE-TUNING OF CNNs (VGG19)

During the training process, the weights of the CNN layers are updated after every iteration by equation (6). There are 19 layers and 144 million trainable parameters (weights) in the VGG19 architecture. To train such a deep network from scratch with randomly initialized weights and to determine the optimum weights requires a large dataset. For a small dataset, it is very difficult to determine the appropriate local minima for the cost function in equation (5), and the network will suffer from overfitting. Therefore, weights are initialized from the pretrained VGG19 model.

After the weights transfer, we extracted features from the activation of fully connected layer FC7 of the pretrained model on the CE-MRI dataset and fed it into CFML to measure the retrieval performance, achieving a *mAP* of 82.23%. To enhance the retrieval performance, we adopted the fine-tuning strategy of the pretrained model. The VGG19 consists of sixteen convolutional layers and three fully connected layers, as shown in Table 1. If we apply layer-wise fine-tuning by adding one layer each time, set the training parameters, train the network and measure retrieval performance, then it will need to fine-tune nineteen layers. Because five-fold cross-validation is under consideration, it will need to fine-tune ninety-five VGG19 architectures. If we estimate approximately thirty minutes for the training of each architecture, then it will take more than a week to complete the fine-tuning of the VGG19 in a layer-wise manner. Similarly, determining the optimum parameters for the layer-wise fine-tuning will be more time-intensive. A small improvement in the results was observed when adopting a layer-wise fine-tuning approach. Therefore, the VGG19 architecture is divided into six blocks based on pooling layers, as shown in Table 1. The block-wise architecture of the VGG19 is shown in Fig. 4. The final fully connected layer of VGG19 composed of 1000 neurons corresponds to classes in the ImageNet dataset, so the final fully connected layer here is changed to three neurons according to classes in the CE-MRI dataset.

The deep CNN is trained in a block-wise approach by starting fine-tuning from the final block and keeping all other blocks (layers) fixed by freezing their learning. Suppose B is the total number of blocks, α_b is the learning rate of block b, and we want to fine-tune only the final block (B) and then set $\alpha_b = 0$ for all blocks except block B. If fine-tuning the last two blocks, then set $\alpha_b = 0$ for b $\neq$ B, B-1. Similarly, set the learning parameters of all blocks for fine-tuning as shown in Table 3.

Earlier layers in the pretrained CNNs contain the generic feature, and higher layers contain the domain (content)-specific features of the natural images. The learning of the earlier layers can be frozen because of the low-level features in these layers. To learn the domain-specific features of MRI brain tumors, we start training from the higher layers by fine-tuning. That is why block-wise fine-tuning is initiated from the top block.

III. EXPERIMENTS AND RESULTS

To test the performance of the proposed approach, we adopted the same experimental setup as in [1]–[4] and randomly divided the CE-MRI data of 233 patients into five subsets of approximately equal size. We ensured no overlap and equal ratios of the different types of tumors in the five subsets for the CE-MRI dataset. Dividing according to the patients ensured that images from the same patient did not exist simultaneously in the training and testing set. We used five-fold cross-validation to evaluate the performance. In five-fold cross-validation, one subset is used as the test dataset, and the remaining four subsets are sequentially used as the training dataset (database). Each image in the test dataset is considered the query image. The final result, called the *mAP,* is the average retrieved precision of the five-fold test dataset. The proposed CBIR architecture was implemented in MATLAB R2017b and trained and tested on GPU NIVIDA TITAN X (Pascal) with 12 GB onboard memory.

A. TRAINING AND PARAMETERS OPTIMIZATION

The training and fine-tuning of each CNN takes approximately 20 to 30 minutes, but it depends on the choice of the training and fine-tuning parameters, proper convergence, training and validation accuracy, and error. To find the optimum convergence of each CNN, we properly monitor the improvement of training, validation accuracy and error. Training stops automatically if there is no improvement with respect to validation accuracy and error.

TABLE 3. Fine-tuning parameters of the VGG19. FT represents fine-tuning, *B* represents the block, α_b is the learning rate of specified block *B*, and μ, γ and α are training options.

Fine-tuning	Training options			Learning rate of each Block (α_b)					
	μ	α	γ	α_{B1}	α_{B2}	α_{B3}	α_{B4}	α_{B5}	α_{B6}
FT: B_1-B_6	0.9	0.01	0.9	0.1	0.1	0.1	0.1	0.1	0.1
FT: B_2-B_6	0.9	0.01	0.9	0	0.1	0.1	0.1	0.1	0.1
FT: B_3-B_6	0.9	0.01	0.9	0	0	0.1	0.1	0.1	0.1
FT: B_4-B_6	0.9	0.01	0.9	0	0	0	0.1	0.1	0.1
FT: B_5-B_6	0.9	0.01	0.9	0	0	0	0	0.1	0.1
FT: B_6	0.9	0.01	0.9	0	0	0	0	0	0.1

Table 3 presents the optimum value of each parameter used in the experiments for fine-tuning. We discovered these values using a trial-and-error-based approach. We performed experiments with different values of these parameters and found during the training process that the proper convergence depends on the initial learning rate α, the learning rate α_b of each layer and the scheduling rate γ. The optimum value for $\alpha = 0.01$ and $\alpha_b = 0.10$ ensured proper convergence. If we set α and α_b to be large, then the CNN fails to converge properly, suffering from overfitting and resulting in low performance on the testing and validation data. If we set α and α_b to be very small, then the convergence process slows down. The value of γ is related to convergence speed. If the convergence is very slow, then γ should be large enough to keep the learning rate high. If the convergence is very fast, then γ should be small enough to decrease the learning rate and prevent the network from overfitting. Our initial choices of the learning rates α and α_b start the convergence relatively fast. We control α by γ to decrease the learning rate after every five epochs to prevent the network from overfitting. During our experimental investigation, the suitable value of $\gamma = 0.90$. Nesterov's momentum μ describes the influence of previously updated weights in the current iteration, and its most commonly used values are [0.5, 0.9, 0.95 and 0.99]. The reasonable value found during our experimental analysis for μ is 0.9.

We set the base-learning rate of each layer twice as the α_b, the mini-batch size for training at 64 (the maximum mini-batch size supported by our GPU for the VGG19) and the maximum epochs at 50 for fine-tuning throughout the experiments. However, most of our fine-tuned CNNs converge between 25 and 35 epochs. The validation test prevents the network from overfitting and helps to monitor proper convergence. We validate the training process after every epoch and stop the training process automatically if there is no improvement on the validation test over 15 epochs.

After training the VGG19, we extract the features from the FC7 layer of the trained model and apply CFML to measure the similarity between the testing dataset/query image and the database images. One of the important parameters of CFML is reduced dimensionality (D) derived from the projection matrix (L). Fig. 6 shows the effect of different values of D. The best results are achieved when D is 2, and the retrieval performance remains unchanged for D greater than 2. These stable results for D show the robustness of CFML and reduce the computational cost in the retrieval phase.

B. PERFORMANCE METRICS

Suppose N is the total number of images in the database; then, $k = 1, 2, \ldots, N$ is the number of images retrieved from the database and T_j is the relevance of the two images x_j and x_i, where $T_j \epsilon \{0, 1\}$. Following the same performance metrics as in [1]–[4], the retrieval performance is evaluated based on the mAP and top n retrieval precision ($Prec@n$). Precision and recall are computed as follows:

$$Precision = \sum_{j=1}^{k} T_j / K \tag{13}$$

$$Recall = \sum_{j=1}^{k} T_j \Big/ \sum_{j=1}^{N} T_j \tag{14}$$

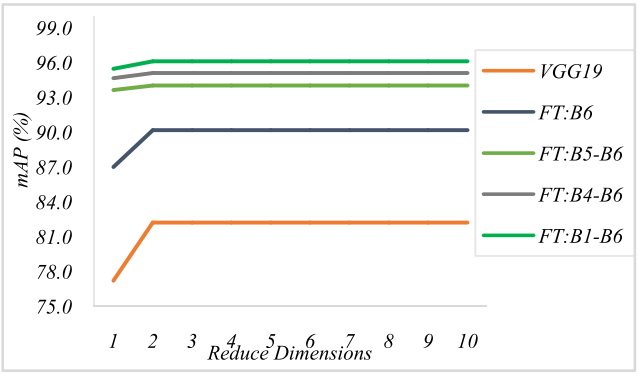


FIGURE 6. Evaluation of the retrieval performance of CMFL with D = 1, 2, . . . , 10.

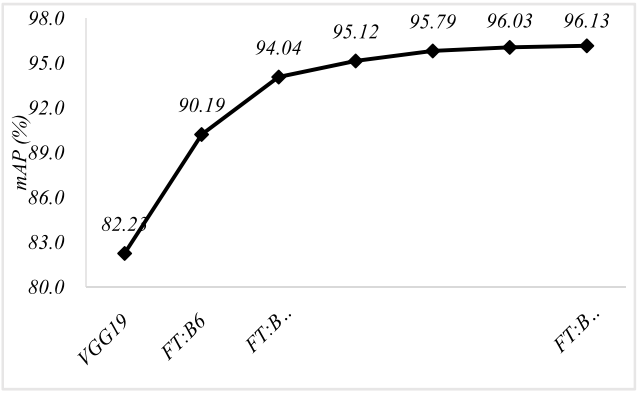


FIGURE 7. Brain tumor retrieval performance of the pretrained VGG19 and fine-tuned models.

If query image x_j and retrieved database image x_i are of the same tumor type then $T_j = 1$; otherwise, $T_j = 0$. Presenting x_j to CBIR, the retrieved images are ranked in ascending order based on their relevance to the x_j. Let $\pi(x_j)$ denote the rank of retrieved image x_j in the ranking list. The top n most similar retrieved images are represented by $Prec@n$, which is the precision at the position where n is the most similar database images returned. $Prec@n$ is calculated as follows:

$$Prec@n = \frac{1}{n} \sum_{j=1}^{N} T_j 1\left\{\pi(x_j) \leq n\right\} \tag{15}$$

where $1\{\cdot\}$ is the indicator function. The average precision (AP) is the average of the precision at the positions where a relevant image exists in the ranking list. AP is calculated as follows:

$$AP = \frac{1}{\sum_{j=1}^{N} T_j} \sum_{j=1}^{N} T_j \times Prec@j \tag{16}$$

TABLE 4. Retrieval performance of the proposed method for specific types of tumors (Mean ± STD).

Tumor Type	*mAP*	*Prec@10*	*Prec@20*
Meningioma	91.81±3.77	89.52±4.94	89.52±4.94
Glioma	97.45±1.45	95.54±2.64	95.54±2.64
Pituitary	97.47±1.70	96.39±2.47	96.39±2.47

TABLE 5. Retrieval performance of the proposed method on five-fold test datasets and its Mean ± STD.

Metric	Set 1	Set 2	Set 3	Set 4	Set 5	Mean ± STD
mAP	96.77	94.38	97.17	96.31	96.04	96.13± 0.96
Prec@10	95.44	92.21	95.91	93.96	94.43	94.39±1.29

The *mAP* is the mean of *AP* over all the query images and is used to calculate the overall retrieval performance given by

$$mAP = \frac{1}{M}\sum_{q=1}^{M} AP_q \tag{17}$$

where M is the number of queries.

C. RETRIEVAL PERFORMANCE AND RESULTS COMPARISON

We evaluated the retrieval performance of features extracted from each block-wise fine-tuned model on five-fold cross-validation. We summarized our results in the form of tables and graphical figures. Fig. 7 summarizes the five-fold average retrieval performance of the pretrained VGG19 and the proposed block-wise fine-tuned models. Our experimental results show that retrieval performance increases gradually with incremental block-wise fine-tuning.

Table 4 shows the average retrieval performance for individual/specific tumor types. The retrieval performance for meningioma is lower than that for glioma and pituitary tumor. The reason is an imbalance of data. Table 5 shows the five-fold retrieval performance on the test datasets and its average and standard deviation. Fig. 8 describes the retrieval performance of the proposed CBIR in comparison with the state-of-the-art CBIR on the same dataset. The retrieval results of the four compared methods are extracted directly from the corresponding original papers. Our proposed CBIR achieved the highest retrieval performance *mAP* of 96.13% and *Prec@10* of 94.39% with the deep fine-tuned model FT: B_1-B_6.

We also examined the transferability of knowledge from natural images to medical brain MR images. To observe the visual effect of the low-level and high-level features, we took feature maps from the deep fine-tuned model FT: B_1-B_6. B_1 describes the low-level features, while B_5 describes the high-level features. Fig. 9 visualizes the concept of low-level general features, while Fig. 10 visualizes the concept of high-level content-specific features of MR images.

In Fig. 9, each square in the montage is the activation output of a channel in block B_1 (the second convolutional and the ReLu layer), as shown in (a) and (b). In (a), white pixels represent strong positive activation, and black pixels represent strong negative activation. A channel that is mostly gray does not activate strongly on the input image. The position of a pixel in the activation of a channel corresponds to the same position in the input image (c). A white pixel at some location in a channel indicates that this channel is strongly activated at that position. When (d) is compared with the input image (c), it is clear that this channel activates on the edges. It activates positively on the light edges and negatively on the dark edges. However, only the positive activation is used because of the ReLu that follows the convolutional layer, as shown in (b).

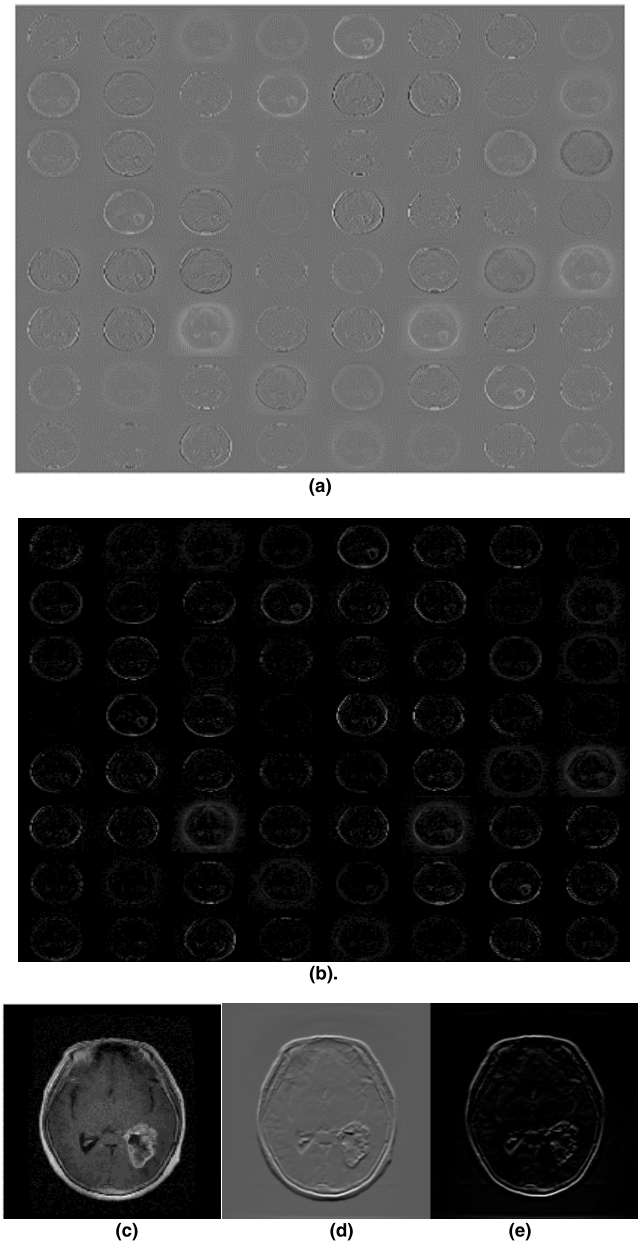
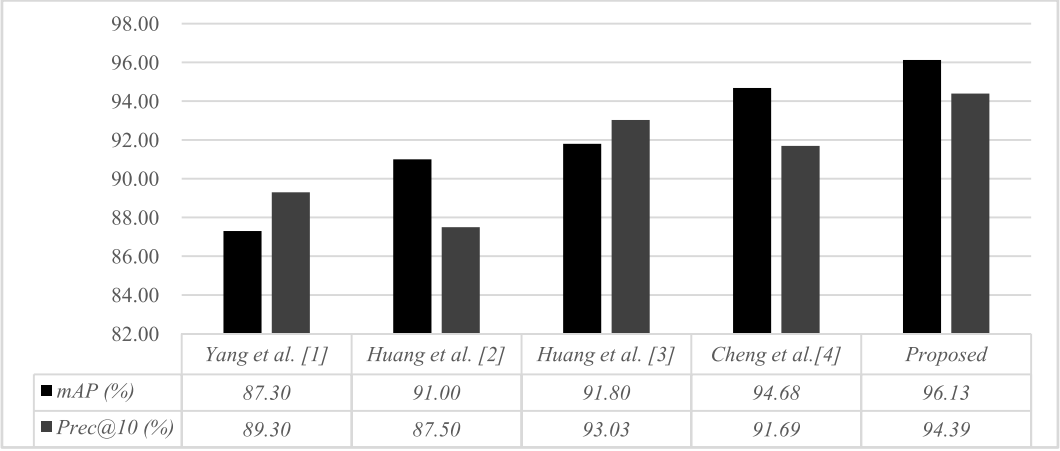


FIGURE 9. Visual results of low-level features learned in the first block B1 of fine-tuned model FT: B1-B6. (c) is the input image. (a) and (b) are a montage of 64 images on an 8-by-8 grid, one for each channel, showing the activation output of the second convolutional and ReLu layers in block B1, respectively. (d) and (e) are the strongest activation channels of the convolutional and ReLu layers, respectively.



	Yang et al. [1]	Huang et al. [2]	Huang et al. [3]	Cheng et al.[4]	Proposed
mAP (%)	87.30	91.00	91.80	94.68	96.13
Prec@10 (%)	89.30	87.50	93.03	91.69	94.39

FIGURE 8. Result comparison of the proposed and state-of-the-art CBIR systems on the CE-MRI dataset.

If we compare the strongest activation channel of the convolutional and ReLu layers, as shown in (d) and (e), respectively, then (e) clearly identifies parts of the image that have strong brain edge features.

In Fig. 10, the deep CNN learns domain (content)-specific features in higher convolutional layers in a self-learning manner by combining features of earlier layers. We explored the fourth convolutional and ReLu layer in block B5 of the deep fine-tuned model (FT: B_1-B_6) in the same way as in Fig. 9. There are 512 feature maps, but for simplicity and owing to space limitations, we explore only the 25 strongest activation channels, as shown in (a) and (b). These activation channels illustrate the interesting structure and focus on the brain area containing the tumor. Many of the channels in (a) contain areas of activation that are both light and dark. However, only the positive activation is used because of the ReLu layer that follows the convolutional layer, as shown in (b). When the strongest activation channel of the convolutional layer (d) is compared with the input image (c), it represents parts of the image that have the tumor structure. Similarly, if we compare (e) and (c), then (e) clearly shows that this strongest activation channel in the ReLu layer activates on the tumor region. We have not provided the tumor information (such as tumor location, tumor segment or tumor boundary) to the CNN, but it has learned that the tumor region is a useful feature to distinguish between classes of brain tumors in MR images. Conventional machine learning methods often use handcrafted features specific to the problem, but these deep CNNs can learn useful features by themselves.

We further investigated the usefulness of transfer learning and fine-tuning for smaller training datasets. For this purpose, we used validation set one for testing and the remaining four sets as training data from the five-fold cross-validation datasets. We reduced the training data by randomly selecting 25, 50, and 75%. We trained FT: B_1-B_6 on the reduced training dataset. As described in Table 6, the features extracted from model FT: B_1-B_6 show a minor decrease in retrieval performance even with 25% training data. The relatively high performance of the CNN, even with smaller datasets, indicates the power of the feature representation of convolutional neural networks.

TABLE 6. Retrieval performance on the reduced training dataset.

Training Data	*mAP*	*Prec@10*	*Prec@20*
25%	91.89	88.25	88.25
50%	95.72	93.86	93.82
75%	96.52	94.74	94.82
100%	96.79	95.44	95.45

To show the practical results, we present three retrieval examples for the three categories of brain tumors, as shown in Fig. 11-13. Due to space limitations, we present only the *Prec@5* (top 5) retrieval results of the query image. Among these figures, the first image is the query image, and the remaining images are those retrieved by the proposed CBIR system. The tumor region is roughly outlined by a yellow boundary. All the *Prec@5* retrieved images are relevant to the given query image.

IV. DISCUSSION

We designed an automatic content-based brain tumor retrieval system. The performance of CBIR depends on good feature representation and suitable distance metric learning. We extracted the features using the potential of transfer learning and block-wise fine-tuning of CNN, and similarity was computed using CFML. The details of these two components are described in the section on the proposed method. Most state-of-the-art methods [1]–[4] (details are provided in the introduction section) use handcrafted features such as tumor region and outline in the feature extraction phase. Our feature extraction method is more generic, as we do not use any handcrafted features. Our proposed deep CNN-based feature extraction framework learned these discriminative features in a self-learning way. We performed this research on the VGG19 model because its architecture is deeper and is suit-

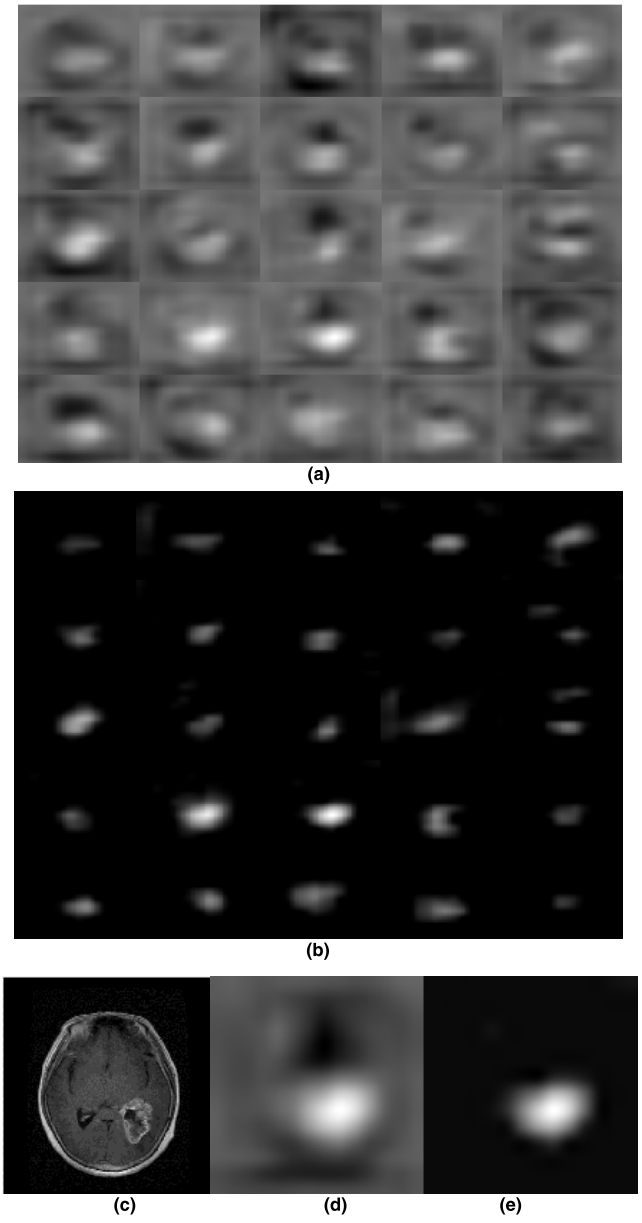


FIGURE 10. Visual results of high-level features learned in the fifth block of fine-tuned model FT: B1-B6. (c) is the input image. (a) and (b) are a montage of 25 images on a 5-by-5 grid, one for each channel, displaying the top 25 strongest activations of the fourth convolutional and ReLu layers in block B5, respectively. (d) and (e) are the strongest activation channels of the convolutional and ReLu layers, respectively.

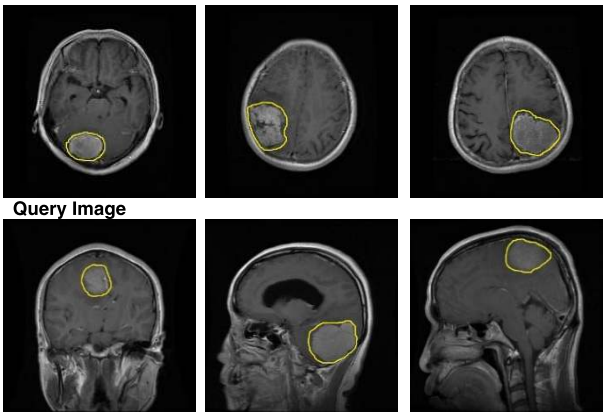


FIGURE 11. Prec@5 (top 5) retrieval results for the query image meningioma.

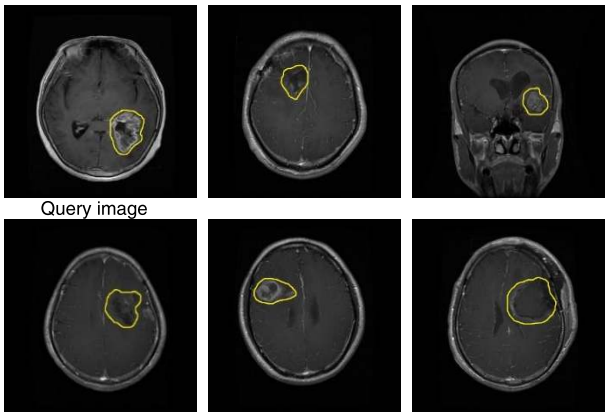


FIGURE 12. Prec@5 (top 5) retrieval results for the query image glioma.

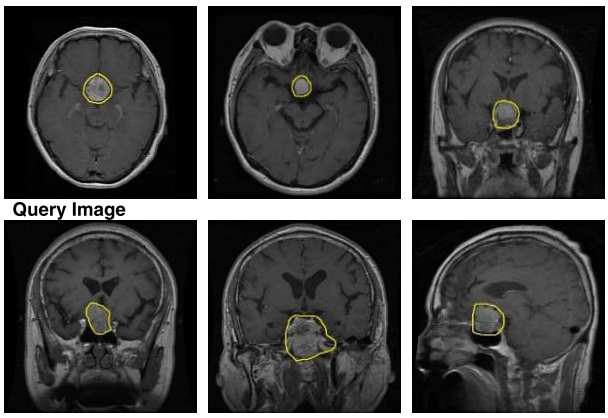


FIGURE 13. Prec@5 (top 5) retrieval results for the query image pituitary.

able for feature representation in terms of localization or detection of specific content in an image. Furthermore, due to the limitations of time and space, the objective of our research focused only on tumor retrieval on brain MR images. Our results reveal that pretrained deep learning models with transfer learning and fine-tuning are the best strategy in the scenario of small datasets, especially in the field of medical imaging.

The second important component of this research is distance metric learning. We obtained efficient retrieval performance using CFML with a *mAP* of 96.13% compared to 70% using simple Euclidean distance. Furthermore, we achieved the best results by applying the CFML approach to project the feature representations into a new space of two dimensions. Therefore, CFML is very efficient in terms of computation and memory due to its low dimensionality. An additional benefit of low-dimension feature vectors is that they can support indexing techniques [59] (e.g., KD-tree, R-tree, and quad-trees). Indexing techniques compare the query image

with only a portion of the relevant database images, thus improving the retrieval efficiency for a large-scale database.

V. CONCLUSION

In this research, we developed a new CBIR approach to brain tumor retrieval based on transfer learning and fine-tuning, which can serve as a helpful tool for clinical diagnosis. The proposed feature extraction framework suggests an alternative approach to pretrained CNN off-the-shelf feature extraction (without training) and training the separate method for retrieval, and it also demonstrates the transferability of learning from natural images to medical brain MR images. This approach may be used to develop CBIR for other body organ MRI images and other medical imaging domains, such as X-rays, PET, and CT. Our CBIR is more generic because it requires only MR images as a query to retrieve the relevant tumor images from the database. The experimental results revealed that the proposed CBIR outperformed state-of-the-art methods on CE-MRI dataset.

VI. ACKNOWLEDGMENT

The authors are thankful to Prof. Lu for providing financial support and supervision for this research. They are also thankful to the anonymous reviewers for their insightful comments.

REFERENCES

[1] 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," *Comput. Math. Methods Med.*, vol. 2012, Oct. 2012, Art. no. 280538.

[2] W. Yang *et al.*, "Content-based retrieval of brain tumor in contrast-enhanced MRI images using tumor margin information and learned distance metric," *Med. Phys.*, vol. 39, no. 11, pp. 6929–6942, 2012.

[3] M. Huang *et al.*, "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.

[4] J. Cheng *et al.*, "Retrieval of brain tumors by adaptive spatial pooling and Fisher vector representation," *PLoS ONE*, vol. 11, no. 6, p. e0157112, 2016.

[5] A. Shah, S. Conjeti, N. Navab, and A. Katouzian, "Deeply learnt hashing forests for content based image retrieval in prostate MR images," *Proc. SPIE*, vol. 9784, p. 978414, Mar. 2016.

[6] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 733–746, Mar. 2011.

[7] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "Visualizing and enhancing a deep learning framework using patients age and gender for chest X-ray image retrieval," *Proc. SPIE*, vol. 9785, p. 978510, Jul. 2016.

[8] X. Liu, H. R. Tizhoosh, and J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and Radon transform," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2872–2878.

[9] W. Yang, Z. Lu, M. Yu, M. Huang, Q. Feng, and W. Chen, "Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single- and multiphase contrast-enhanced CT images," *J. Digit. Imag.*, vol. 25, no. 6, pp. 708–719, 2012.

[10] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 783–792, Feb. 2015.

[11] J. Cheng, *Brain Tumor Dataset*, document, Figshare, 2017. [Online]. Available: https://doi.org/10.6084/m9.figshare.1512427.v5

[12] H. Greenspan and A. T. Pinhas, "Medical image categorization and retrieval for PACS using the GMM-KL framework," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 190–202, Mar. 2007.

[13] M. M. Rahman, B. C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," *Comput. Med. Imag. Graph.*, vol. 32, no. 2, pp. 95–108, 2008.

[14] D. K. Iakovidis, N. Pelekis, E. E. Kotsifakos, I. Kopanakis, H. Karanikas, and Y. Theodoridis, "A pattern similarity scheme for medical image retrieval," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 442–450, Jul. 2009.

[15] P. John, "Brain tumor classification using wavelet and texture based neural network," *Int. J. Sci. Eng. Res.*, vol. 3, no. 10, pp. 1–7, 2012.

[16] J. Jiang, Y. Wu, M. Huang, W. Yang, W. Chen, and Q. Feng, "3D brain tumor segmentation in multimodal MR images based on learning population- and patient-specific feature sets," *Comput. Med. Imag. Graph.*, vol. 37, pp. 512–521, Oct./Dec. 2013.

[17] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1704–1716, Sep. 2012.

[18] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, Sep. 2014.

[19] H. Greenspan, B. V. Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, Mar. 2016.

[20] J. Wan *et al.*, "Deep learning for content-based image retrieval: A comprehensive study," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 157–166.

[21] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object detection networks on convolutional feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1476–1481, Jul. 2017.

[22] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[23] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[25] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[26] F. Ciompi *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Med. Image Anal.*, vol. 26, no. 1, pp. 195–202, Dec. 2015.

[27] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, Mar. 2015.

[28] J. Kleesiek *et al.*, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, Apr. 2016.

[29] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014.

[30] H.-I. Suk, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Deep learning in diagnosis of brain disorders," in *Recent Progress in Brain and Cognitive Engineering*. Dordrecht, The Netherlands: Springer, 2015, pp. 203–213.

[31] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," *NeuroImage*, vol. 129, pp. 292–307, Apr. 2016.

[32] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.

[33] G. van Tulder and M. de Bruijne, "Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1262–1272, May 2016.

[34] Q. Dou *et al.*, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, May 2016.

[35] D. C. Cire an, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2013, pp. 411–418.

[36] H. Chen, X. J. Qi, J. Z. Cheng, and P. A. Heng, "Deep contextual networks for neuronal structure segmentation," in *Proc. AAAI*, 2016, pp. 1167–1173.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[38] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[39] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[40] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[41] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[43] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[44] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. (2011). *The Pascal Visual Object Classes Challenge 2012 (VOC2012) Results (2012)*. [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html
[45] L. Roux *et al.* (2014). *MITOS-ATYPIA-14*. [Online]. Available: http://mitos-atypia-14.grand-challenge.org/
[46] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[47] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[48] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 806–813.
[49] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, "From generic to specific deep representations for visual recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 36–45.
[50] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 44–51.
[51] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," *Proc. SPIE*, vol. 9414, p. 94140V, Mar. 2015.
[52] B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 286–289.
[53] A. Mojsilovic and B. Rogowitz, "Capturing image semantics with low-level descriptors," in *Proc. Int. Conf. Image Process.*, 2001, pp. 18–21.
[54] H. Guan, S. Antani, L. R. Long, and G. R. Thoma, "Bridging the semantic gap using ranking SVM for image retrieval," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro (ISBI)*, Jun./Jul. 2009, pp. 354–357.
[55] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.
[56] K. Simonyan, A. Vedaldi, and A. Zisserman, "Learning local feature descriptors using convex optimisation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1573–1585, Aug. 2014.
[57] B. Alipanahi, M. Biggs, and A. Ghodsi, "Distance metric learning vs. Fisher discriminant analysis," in *Proc. 23rd Nat. Conf. Artif. Intell.*, 2008, pp. 598–603.
[58] E. P. Xing, M. I. Jordan, S. J. Russell, and A. Y. Ng, "Distance metric learning with application to clustering with side-information," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, pp. 521–528.
[59] P. Wu, S. C. H. Hoi, P. Zhao, C. Miao, and Z.-Y. Liu, "Online multi-modal distance metric learning with application to image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 454–467, Feb. 2016.

ZAR NAWAB KHAN SWATI received the B.S. degree in computer science from Hazara University, Mansehra, in 2007, and the master's degree in computer science from COMSATS University, Abbottabad, Pakistan, in 2009. He is currently pursuing the Ph.D. degree with the Nanjing University of Science and Technology, China. He was a Research Associate with COMSATS University, during 2007–2010. In 2010, he joined Karakoram International University, Pakistan, as an Assistant Professor. He was the Chairman of the Department of Computer Science, from 2011 to 2015. He is a member of the Deep Learning Group. His research interests include image processing, machine learning, deep learning, computer vision, and pattern recognition.

QINGHUA ZHAO received the B.S. degree from Xingtai University, Xingtai, China, in 2006, the M.S. degree from Northwest Minzu University, Lanzhou, China, in 2011, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, Nanjing, China, in 2018. From 2013 to 2015, he was a Visiting Scholar with the University of Georgia, Athens, USA. He is currently an Assistant Professor with the College of Information Engineering, Nanjing University of Finance and Economics. His current research interests include pattern recognition, medical image analysis, and optimization algorithm.

MUHAMMAD KABIR received the bachelor's degree in computer science from the Islamia College Peshawar, in 2012, and the master's degree in computer science from Abdul Wali Khan University Mardan, Pakistan, in 2016. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. He is a member of the Pattern Recognition and Bioinformatics Group. His current research interests include machine learning and deep learning in bioinformatics and image processing.

FARMAN ALI received the B.S. degree in computer science from the University of Peshawar, in 2009, and the M.S. degree in computer science from Abdul Wali Khan University Mardan, in 2016. He is currently pursuing the Ph.D. degree in computer science with research area bioinformatics and machine intelligence with the Nanjing University of Science and Technology. He is a member of the CSBIO Group.

ZAKIR ALI received the B.S. and M.S. degrees in information technology from IBMS, The University of Agriculture Peshawar, Pakistan, in 2012. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanjing University of Science and Technology. He is a member of the Deep Learning Group. He received gold medal in Master degree (Master of Science in Information Technology).

SAEED AHMED received the B.S. degree in telecommunication from the University of Science and Technology, Bannu, Pakistan, in 2011, and the master's degree in computer science from Abdul Wali Khan University Mardan, Pakistan, in 2016. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanjing University of Science and Technology. He is a member of the Pattern Recognition and Bioinformatics Group.

JIANFENG LU received the B.S. degree in computer software and the M.S. and Ph.D. degrees in pattern recognition and intelligent system from the Nanjing University of Science and Technology, Nanjing, China, in 1991, 1994, and 2000, respectively, where he is currently a Professor and a Vice Dean of the School of Computer Science and Engineering. His research interests include image processing, pattern recognition, and data mining.