



Brain Tumor MR Image Classification Using Convolutional Dictionary Learning With Local Constraint

Xiaoqing Gu¹, Zongxuan Shen¹, Jing Xue², Yiqing Fan³ and Tongguang Ni^{1*}

¹ School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China, ² Department of Nephrology, Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China, ³ Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

OPEN ACCESS

Edited by:

Mohammad Khosravi,
Persian Gulf University, Iran

Reviewed by:

Min Shi,
Fuzhou University of International
Studies and Trade, China
Yanhui Zhang,
Hebei University of Chinese Medicine,
China

*Correspondence:

Tongguang Ni
ntg@cczu.edu.cn

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 12 March 2021

Accepted: 09 April 2021

Published: 28 May 2021

Citation:

Gu X, Shen Z, Xue J, Fan Y and
Ni T (2021) Brain Tumor MR Image
Classification Using Convolutional
Dictionary Learning With Local
Constraint.
Front. Neurosci. 15:679847.
doi: 10.3389/fnins.2021.679847

Brain tumor image classification is an important part of medical image processing. It assists doctors to make accurate diagnosis and treatment plans. Magnetic resonance (MR) imaging is one of the main imaging tools to study brain tissue. In this article, we propose a brain tumor MR image classification method using convolutional dictionary learning with local constraint (CDLLC). Our method integrates the multi-layer dictionary learning into a convolutional neural network (CNN) structure to explore the discriminative information. Encoding a vector on a dictionary can be considered as multiple projections into new spaces, and the obtained coding vector is sparse. Meanwhile, in order to preserve the geometric structure of data and utilize the supervised information, we construct the local constraint of atoms through a supervised k -nearest neighbor graph, so that the discrimination of the obtained dictionary is strong. To solve the proposed problem, an efficient iterative optimization scheme is designed. In the experiment, two clinically relevant multi-class classification tasks on the Cheng and REMBRANDT datasets are designed. The evaluation results demonstrate that our method is effective for brain tumor MR image classification, and it could outperform other comparisons.

Keywords: brain tumor image classification, magnetic resonance imaging, dictionary learning, local constraint, convolutional neural network

INTRODUCTION

Brain tumors are abnormal cell aggregations that grow inside the brain tissues. Brain tumors can be divided into benign tumors and malignant tumors. Brain benign tumors can be cured by surgery, while malignant brain tumors are one of the most deadly types of cancer and can lead directly to death (Yang et al., 2018; Sun et al., 2019; Ge et al., 2020). Brain tumors can also be divided into primary tumors formed in the brain or derived from the brain nerves and metastatic brain tumors metastasized from other parts of the body to the brain. The most common primary brain tumors in adults are primary central nervous system lymphoma and gliomas, of which gliomas originate from the periglial tissue and account for more than 80% of malignant brain tumors. Different symptoms appear with different lesion areas, such as headache, vomiting, visual decline, epilepsy, and confusion. A more detailed classification divides brain tumors into four grades, the higher the grade, the more malignant the brain tumor is. According to the Global cancer statistics 2020 (Sung et al., 2021), there are about 308,000 new cases of brain cancers in 2020, accounting for about 1.6%

of all new cases of cancers, and about 251,000 deaths from brain cancers, accounting for about 2.5% of all cancer deaths.

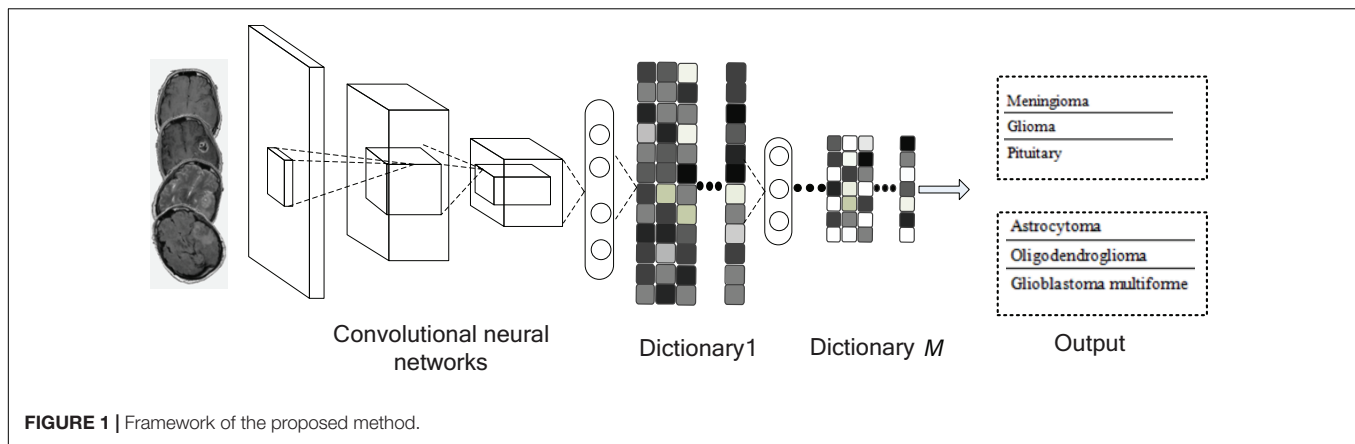
Early detection is important for effective treatment of brain tumors (Gumaei et al., 2019). With the development of medical imaging, imaging techniques play an important role in brain tumor diagnosis and treatment evaluation and can provide doctors with a clear human brain structure. These imaging techniques can provide information on the shape, size, and location of brain tumors, assisting doctors to make an accurate diagnosis and develop a treatment plan. Magnetic resonance (MR) imaging is one of the most commonly used scanning methods in neurology. MR imaging uses radiofrequency signals to excite the target tissue under the influence of a very strong magnetic field to produce an image of its interior. It has the advantages of high soft tissue contrast and zero ionizing radiation exposure. Therefore, MR imaging is more suitable for the detection of brain lesions (Zeng et al., 2018; Mittal et al., 2019; Bunevicius et al., 2020).

In recent years, artificial intelligence has attracted more and more attention due to its achievements in the field of intelligent medicine. The classification and segmentation of MR images using artificial intelligence methods has become a hot topic in the research of medical image processing (Mohan and Subashini, 2018; Anaraki et al., 2019). The application of brain tumor classification falls into two main types: classification of brain images into normal and abnormal, that is, whether the brain image contains a tumor or not; and classification within abnormal brain images, that is, differentiation between different classes of brain tumors. Classifying brain tumors into different pathological classes is more challenging than a binary classification. The challenge lies in brain tumors being permeable, their appearance is highly heterogeneous, their location is random, and the number of voxels in each subregion varies widely (Chahal et al., 2020).

Brain tumor classification includes two procedures: feature extraction and classification. In some previous studies, traditional manual extraction of features was widely used, such as intensity and texture features of brain tumor images. However, traditional feature extraction methods require the professional knowledge and experience in specific fields. Manual feature extraction will also reduce the efficiency of the system. Deep learning techniques overcome this disadvantage (Sajjad et al., 2019). Feature extraction methods based on deep learning have demonstrated successful results in real-world medical image processing applications (Deepak and Ameer, 2019). Among various classification methods, dictionary learning (DL) is a powerful tool in image processing and machine vision, making sparse coding tasks efficient and robust (Al-Shaikhli et al., 2016; Ni et al., 2020). The sparse coding can approximate high-dimensional image features into a linear combination of a few atoms from the learned dictionary (Li et al., 2018; Gu et al., 2020b). Ghasemi et al. (2020, 2021) developed fuzzy dictionaries to deal with the uncertainty in brain tumor image classification. The classic fuzzy inference is embedded into the dictionary learning process and fuzzy membership functions are used to model uncertainty and improve sparse representation. Wu et al. (2017) developed a parse representation

method to extract important features and key feature index across different class images. Then, the learned feature weights and classification dictionary are used in a radiomics system for the diagnosis of brain tumors. Al-Shaikhli et al. (2014) developed a coupled dictionary learning method, which designs one dictionary of brain tumor image patches and one dictionary of image labels. The label dictionary is used to present the foreground and background multiple labels. Then, Al-Shaikhli et al. (2016) extended this work by using the information of brain topology and texture to develop a multi-class brain tumor classification method. Chen et al. (2017) proposed a kernel sparse representation method for multi-label brain tumor segmentation. This method consists of three main parts as principal component analysis—split for dictionary learning initialization, second for kernel sparse representation processing of kernel dictionary learning and kernel sparse coding, and third for making brain image segmentation using graph-cut method. Adebileje et al. (2017) applied two dictionary learning methods for classifying proton MR spectroscopy of brain gliomas tumor, i.e., one is discriminate sub-dictionary learning method, and the other is projective dictionary pair learning. These two methods were tested on many H-MRS patients signal samples selected in an Iran hospital and evaluated to be noise insensitive. Tong et al. (2019) proposed a kernel dictionary learning method that segmented MR brain tumor images after the noise removal and contrast enhancement and then extracted the nonlinear features by the learned kernel dictionary for healthy and pathologically tissues. Finally, the segmentation is done by kernel-clustering method. Vu et al. (2015) proposed a feature-oriented dictionary learning method. This method incorporated feature extraction discriminative into dictionary learning. In addition, it built discriminative class-specific dictionaries that emphasized the small intra-class differences and large inter-class differences. Finally, this method had been evaluated using brain cancer dataset Cancer Genome Atlas.

Our goal in this study is to build an automatic and effective brain tumor MR image classification method to assist physicians in decision-making. In order to capture the better discriminative feature representations of brain tumor MR images, we propose convolutional dictionary learning with local constraint (CDLLC) method to seek sparse feature representation and dictionary simultaneously by using a convolutional neural network (CNN) framework. In addition, we employ the locality constraint term on codes in the last layer. The locality constraint term is used to enforce the manifold structure of the codes to preserve the locality information. Various CNN structures can be used in CDLLC; in this study, we use AlexNet (Krizhevsky et al., 2012) and softmax classifier loss in the last layer. The framework of the proposed method is shown in **Figure 1**. The advantages of the proposed CDLLC method are as follows: (1) CDLLC learns a multi-layer convolutional dictionary for feature representation and encoding in the nonlinear space, so that the nonlinear latent information of data is employed. (2) Encoding a vector on a multi-layer dictionary can be considered as multiple projections into new spaces. The projection is nonlinear and the obtained coding vector is sparse. Simultaneously, the resulting coding vectors of different classes can give the discriminative



approximation and delete the redundant information. For example, the coding vectors in single layer dictionary learning may be nonlinear separable, and they will be transformed into linear separable in our method. (3) By considering the supervised information and graph Laplacian regularization term, the learned coding vectors are more discriminative. Simultaneously, graph Laplacian regularization preserves the locality structure information of the learned dictionary in the last layer. (4) The proposed CDLLC method is conducted on two public brain tumor datasets. The performance and usefulness of CDLLC are validated in terms of accuracy, recall, precision, F1-score, and balance loss.

The rest of the article is organized as follows: the related work is introduced in section “Backgrounds.” The proposed method is given in section “Convolutional Dictionary Learning With Local Constraint,” and experiments are reported in section “Experiments.” Finally, a conclusion is summarized in section “Conclusion.”

BACKGROUNDS

Dataset

The brain tumor datasets used in this article are provided by Cheng et al. (2015, 2016) and the Repository of Molecular Brain Neoplasia Data (REMBRANDT) (Clark et al., 2013). The images provided by Cheng are 3064 T1-weighted contrast-enhanced images, containing 708 meningiomas, 1426 gliomas, and 930 pituitary tumors. All images are digitized at a resolution of 512×512 pixels. The REMBRANDT dataset contains 110,020 pre-surgical MR multi-sequence images from 130 brain tumor patients. The dataset contains astrocytoma (AST), oligodendroglioma (OLI), glioblastoma multiforme (GBM), and other unidentified tumor types. All images are digitized at a resolution of 256×256 pixels. Each image in the Cheng and REMBRANDT datasets is labeled with one type of brain tumor. The example samples of the Cheng and REMBRANDT datasets are shown in Figures 2, 3, respectively. The challenge of these two datasets lies in some factors, such as high variability in shape, size, and the similar presentation of different pathological types.

Dictionary Learning

Let $X = [x_1, x_2, \dots, x_N] \in R^{d \times N}$ be the labeled training images and $A \in R^{K \times N}$ be the sparse coding vector matrix. Denote the dictionary to be learned by $D \in R^{d \times K}$. Consider the linear representation, $DA \approx X$. The dictionary D can be learned as

$$\begin{aligned} \min_{D, A} \quad & \|X - DA\|_F^2 + \lambda \Theta(A), \\ \text{s.t.} \quad & \|d_i\|_2 \leq 1, \forall i \end{aligned} \quad (1)$$

where the first term is reconstruction error and $\Theta(A)$ represents the constraints of coding vector matrix A , such as ℓ_0 , ℓ_1 , ℓ_2 and Frobenius norm. The parameter λ is a positive scalar, which controls the sparsity. $\|d_i\|_2 \leq 1$ is used to control the complexity of model. Its purpose is to prevent dictionary D from being arbitrarily large, as it will result in very small values for coding matrix A . Parameters D and A are often optimized by alternating iterations until convergence.

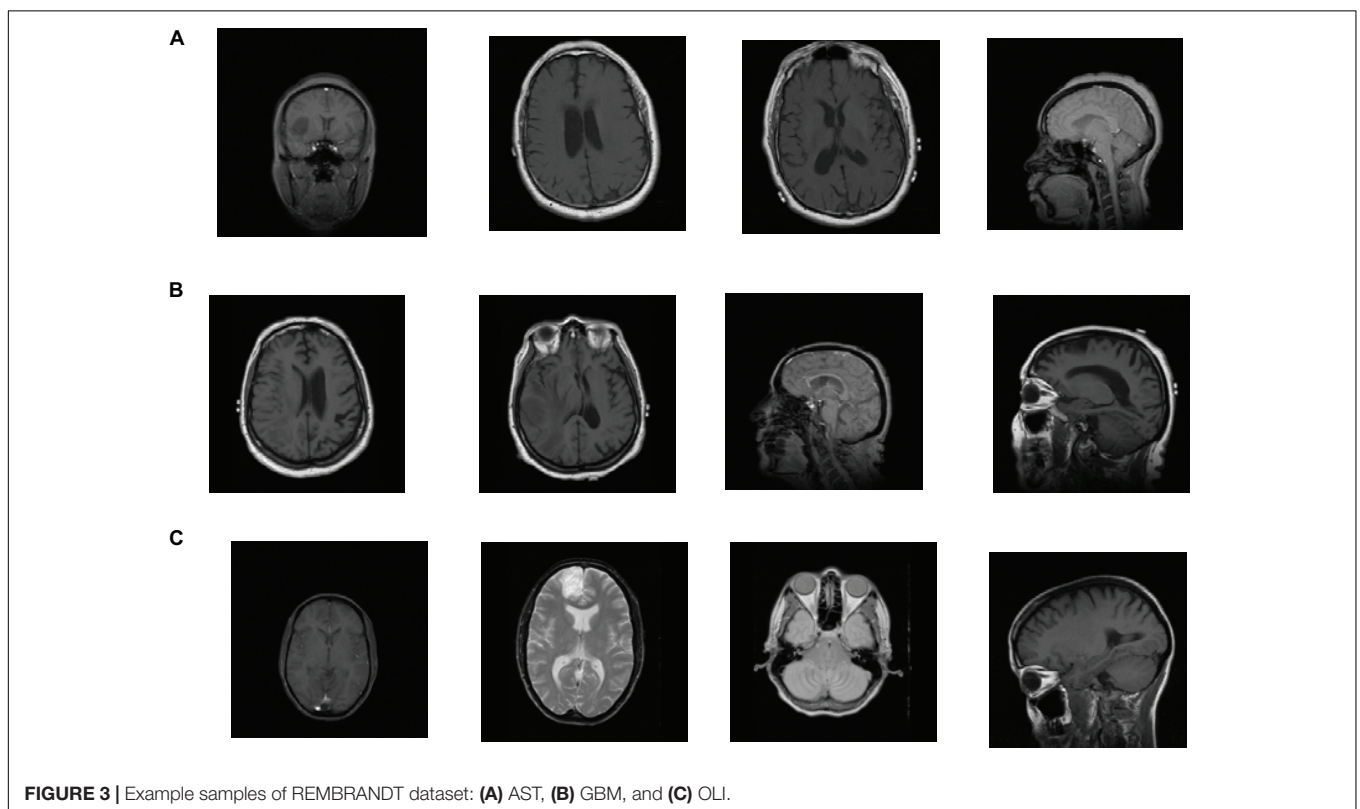
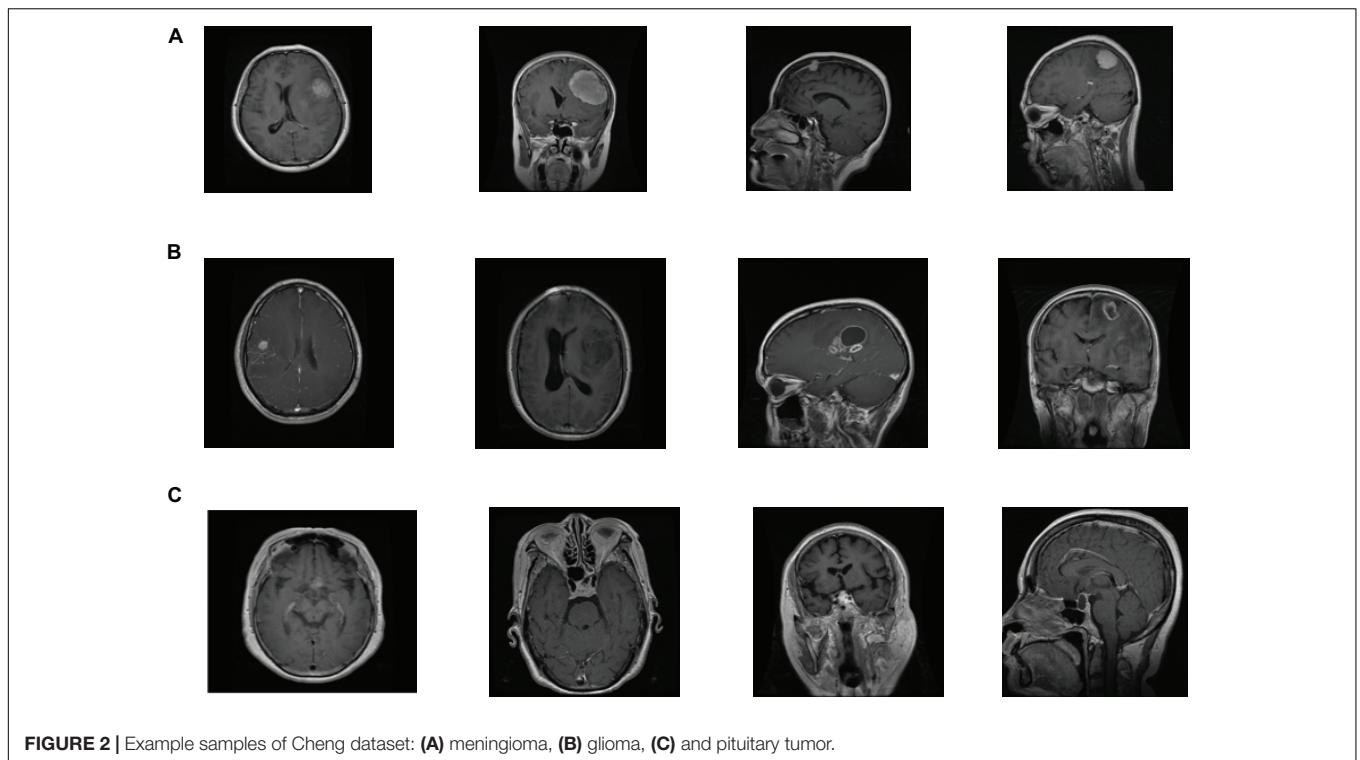
Equation 1 is an unsupervised learning framework. It can obtain good performance in reconstruction tasks and can also be used in some classification tasks, as it is good at mining potential patterns in the data. In order to make better use of supervised information in classification tasks, different kinds of loss functions are considered in dictionary learning. The supervised dictionary learning can be presented as

$$\begin{aligned} \min_{\theta, D} \quad & \sum_{x \in X} L(l_x, a_x, D, \theta), \\ a_x = \arg \min \quad & \|l_x - Da\|_2^2 + \lambda \Theta(a), \end{aligned} \quad (2)$$

where l_x is the class label of training sample x . Determining the suitable classification loss function L and its parameter θ are critical to classification tasks. The joint learning of parameters D and θ allows D to have discriminative capability with minimal classification cost. In order to obtain the optimal solution of Eq. 2, gradient descent method, back propagation method, or orthogonal matching pursuit (OMP) (Wang et al., 2012; Peng et al., 2020) can be used.

Convolution Dictionary Learning

Convolution dictionary learning has been proposed in recent years (Papyan et al., 2018; Sulam et al., 2018; Song et al., 2019).



The model is an extension of traditional dictionary learning. Its aim is to capture the deep structure of data and increase the discriminability of features. Convolution dictionary learning

follows the architecture of CNN and is used in a hierarchical way. The convolution of the filter in CNN corresponds to the sparse coding step in multi-layer convolution dictionary. Let

$\{D_m\}_{m=1}^M$ be M -layer convolution dictionary, where $D_m \in R^{d \times K_m}$ is the dictionary in the m -th layer dictionary and K_m is the size of D_m . The convolutional representation of X can be presented as $X \approx D_1 D_2 \dots D_M A_M$. In detail, with the decomposition constraint $D_{m-1} = D_m A_m$, this process can be described as

$$\begin{aligned} X &\approx D_1 A_1, \\ X &\approx D_1 D_2 A_2, \\ X &\approx D_1 D_2 D_3 A_3, \\ &\vdots \\ X &\approx D_1 D_2 \dots D_M A_M. \end{aligned} \tag{3}$$

where $A_m \in R^{K_m \times N}$ is the coding vector matrix in the m -th layer.

CONVOLUTIONAL DICTIONARY LEARNING WITH LOCAL CONSTRAINT

Objective Function

Consider multi-layer dictionary architecture with M layers, the coding A_m in the m -th layer can be written as

$$A_m \approx \phi(D_{m+1} A_{m+1}), \tag{4}$$

where ϕ is a nonlinear function, such as rectified linear unit (ReLU), Sigmoid, and TanHyperbolic activation function. In this case, coding using a dictionary is used as a projection into another feature space, and the coding vector is a new input for the next layer. In order to preserve the essential structure information of data, it is essential to reconstruct the original sample in the last layer. We consider the following equation:

$$\begin{aligned} J_1(D_1, D_2, \dots, D_M, A_M) = & \| X - D_1 \phi(D_2 \phi(\dots \phi(D_M A_M))) \|_F^2 \\ & + \lambda \Theta(A_M), \end{aligned} \tag{5}$$

where λ is the regularization parameter.

Following Cai et al. (2013) and Ding and Fu (2018), we use the rank operator for $\Theta(A_M)$. In this case, A_M can be approximated as $A_M \approx SH$, and the $\Theta(A_M)$ term can be written as

$$\| A_M - SH \|_F^2, \tag{6}$$

where $S \in R^{K_M \times C}$ and $H \in R^{C \times N}$. C is the class number of data samples.

To improve the classification performance, local information takes an important part in dictionary learning (Peng et al., 2020). Since the atoms of dictionary is more robust and stable than original samples, we use a graph Laplacian regularization term of atoms to trace the manifold structure of data. We build a supervised k -nearest neighbor graph W of atoms on dictionary D_M in the last layer. The element $w_{i,j}$ in graph W is denoted as follows:

$$w_{i,j} = \begin{cases} \exp\left(-\frac{\|d_{M,i} - d_{M,j}\|^2}{\sigma}\right), & \text{if } d_{M,i} \in KNN(d_{M,j}) \\ & \text{and } l_i = l_j \\ 0, & \text{else} \end{cases}, \tag{7}$$

where σ is an adjustable parameter. Different from the unsupervised graph Laplacian (Peng et al., 2020), we embed the supervised information in graph W , such that similar codes are enforced and more discriminative information can be learned.

Then, we construct the graph Laplacian regularization term in the last layer as follows:

$$J_2(A_M) = \sum_{i=1}^{K_M} \sum_{j=1}^{K_M} w_{i,j} \| a_{M,i} - a_{M,j} \|_2^2 = Tr(A_M^T L A_M), \tag{8}$$

where the Laplacian matrix $L = \text{diag}(w_1, w_2, \dots, w_{K_M}) - W$, and $w_i = \sum_{j=1}^{K_M} w_{i,j}$.

To further learn a discriminative dictionary by exploiting supervised information, we use the softmax classifier loss in the last layer, which is commonly used in CNN structure

$$J_3(A_M, \theta) = -\frac{1}{M} \sum_{i=1}^M \sum_{c=1}^C p_{c,i} \log \frac{e^{\theta_c^T a_{M,i}}}{\sum_{u=1}^C e^{\theta_u^T a_{M,i}}}, \tag{9}$$

where $p_{c,i}$ is the label probability of $a_{M,i}$ and is assigned to class c . $\theta = [\theta_1, \theta_2, \dots, \theta_C]$ is the adjustable parameter in softmax classifier.

For brain tumor image classification, we combine the three functions J_1, J_2 , and J_3 together, jointly optimizing convolutional dictionary learning and classifier. Therefore, we have the objective function of the CDLLC model

$$\begin{aligned} \arg \min J_1(D_1, D_2, \dots, D_M, A_M) + J_2(A_M) + J_3(A_M, \theta), \\ \text{s.t. } \| d_i \|_2^2 \leq 1, \forall i \end{aligned} \tag{10}$$

TABLE 1 | Confusion matrix of CDLLC on the Cheng dataset.

	Meningioma	Glioma	Pituitary
Meningioma	0.8875	0.0782	0.0343
Glioma	0.0444	0.9487	0.0069
Pituitary	0.0093	0.0070	0.9837

TABLE 2 | Classification performance of CDLLC on the Cheng dataset.

		Accuracy	Recall	Precision	F1-score	Balance loss
Fold-1	Training	96.94	94.97	95.17	94.95	96.56
	Test	96.36	94.59	94.78	94.70	96.03
Fold-2	Training	96.76	95.15	94.78	95.25	96.44
	Test	96.29	94.71	94.42	94.63	96.08
Fold-3	Training	96.77	95.02	95.21	95.30	96.41
	Test	96.32	94.62	94.51	94.65	96.11
Fold-4	Training	96.83	94.89	94.99	95.10	96.47
	Test	96.35	94.67	94.60	94.75	96.06
Fold-5	Training	97.12	95.09	95.28	95.17	96.96
	Test	96.39	94.64	94.61	94.70	96.22

To show all terms clearly, we expand the expression as follows:

$$\begin{aligned}
 & [D_1, D_2, \dots, D_M, A_M, \theta] \\
 & = \arg \min \|X - D_1 \phi(D_2 \phi(\dots \phi(D_M A_M)))\|_F^2 + \lambda_1 \|A_M - SH\|_F^2 \\
 & + \lambda_2 \text{Tr}(A_M^T L A_M) - \frac{\lambda_3}{M} \sum_{i=1}^M \sum_{c=1}^C p_{c,i} \log \frac{e^{\theta_c^T a_{M,i}}}{\sum_{u=1}^C e^{\theta_u^T a_{M,i}}}, \\
 & \text{s.t. } \|d_i\|_2 \leq 1, \forall i
 \end{aligned} \tag{11}$$

This jointly optimizing convolutional dictionary learning and classifier has benefits. During the procedure of optimizing, the convolutional dictionary learning gradually enhances the classification performance of the classifier; meanwhile, the learned classifier also improves the numerical stability and discriminative ability of dictionary coding.

Optimization of CDLLC

The optimization of Eq. 11 is not convex, and we solve dictionaries $\{D_1, D_2, \dots, D_M\}$, coding matrix A and classifier parameter θ by an alternative optimization approach. In each

step of iteration, we compute a certain parameter and fix the other parameters.

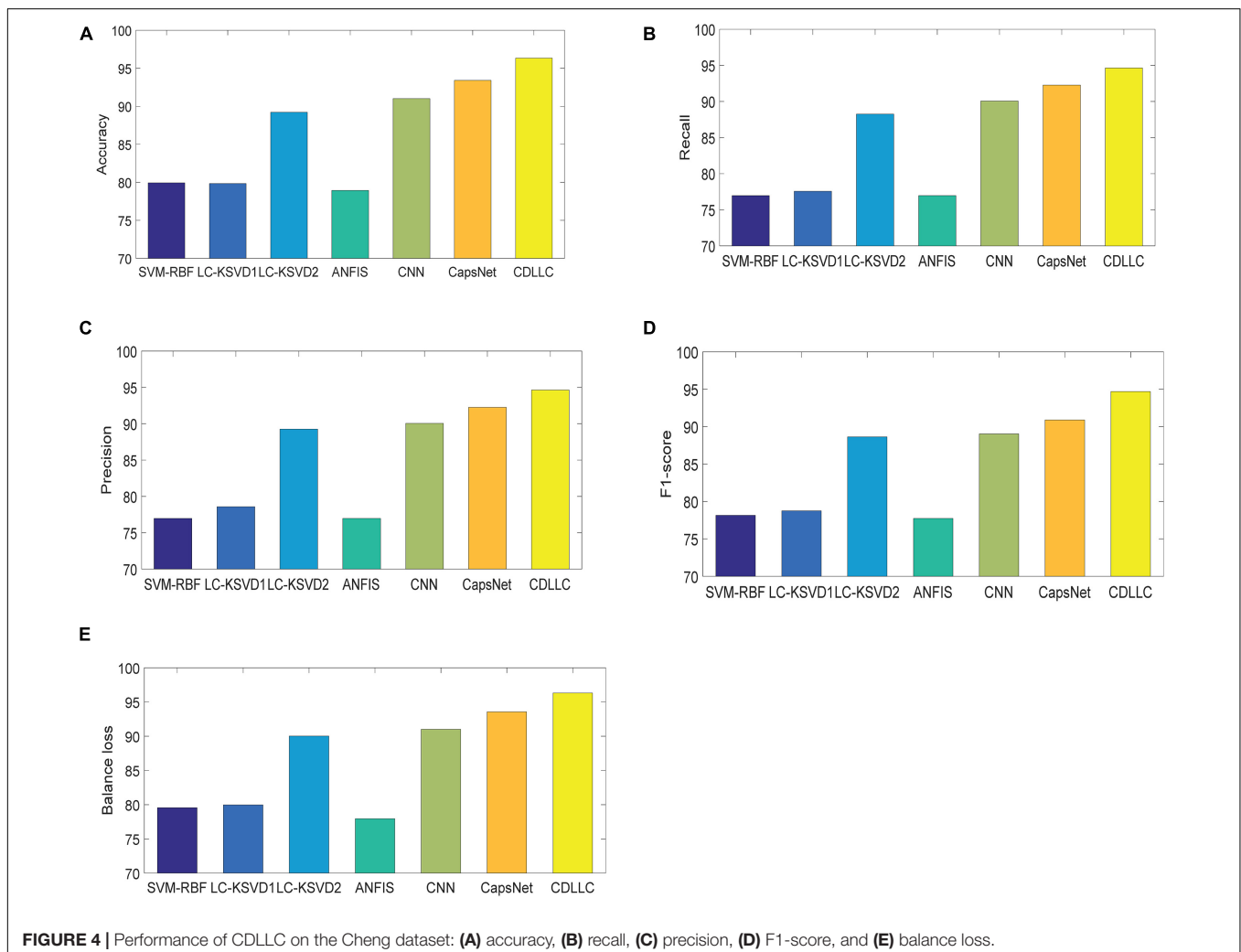
When fixing A and θ , we update dictionaries $\{D_1, D_2, \dots, D_M\}$. We use the chain rule to compute $D_m (1 \leq m \leq M)$ in each layer

$$\begin{aligned}
 \frac{\partial J}{\partial D_m} & = \frac{\partial J}{\partial (D_m A_m)} A_m^T \\
 & = \left[\frac{\partial J}{\partial \phi(D_m A_m)} \odot \phi'(D_m A_m) \right] A_m^T \\
 & = \left[\frac{\partial J}{\partial A_{m-1}} \odot \phi'(D_m A_m) \right] A_m^T,
 \end{aligned} \tag{12}$$

where \odot denotes the element-wise multiplication. Specifically, we can obtain $\frac{\partial J}{\partial D_1} = 2(D_1 A_1 - X) A_1^T$. After dictionary D_M is obtained, we use Eq. 7 to construct the graph Laplacian regularization.

When fixing $\{D_1, D_2, \dots, D_M\}$ and θ , we update coding matrix $A_m (1 < m \leq M-1)$ in each layer as

$$\begin{aligned}
 \frac{\partial J}{\partial A_m} & = (D_m)^T \frac{\partial J}{\partial (D_m A_m)} \\
 & = (D_m)^T \left[\frac{\partial J}{\partial \phi(D_m A_m)} \odot \phi'(D_m A_m) \right] \\
 & = (D_m)^T \left[\frac{\partial J}{\partial A_{m-1}} \odot \phi'(D_m A_m) \right].
 \end{aligned} \tag{13}$$



Specifically, we can obtain A_1 as

$$\frac{\partial J}{\partial A_1} = 2D_1^T(D_1A_1 - X). \tag{14}$$

The partial derivatives of the J regarding the A_M is given as

$$\begin{aligned} \frac{\partial J}{\partial A_M} = & 2(D_M)^T \left[\frac{\partial J}{\partial A_{M-1}} \odot \phi'(D_MA_M) \right] + 2\lambda_1(A_M - SH) \\ & + 2\lambda_2 L_M A_M - \frac{\lambda_3}{M} \sum_{c=1}^C \left(\theta_c - \frac{\sum_{u=1}^C \theta_u e^{\theta_u^T A_{M,i}}}{\sum_{u=1}^C e^{\theta_u^T A_{M,i}}} \right). \end{aligned} \tag{15}$$

Considering Eq. 6, we need to update matrixes S and H in turn. The optimal S can be computed as

$$\frac{\partial J}{\partial S} = A_M H^T - S H H^T. \tag{16}$$

Then, we can update S as

$$S = A_M H^T (H H^T)^\dagger, \tag{17}$$

where \dagger is the Moore–Penrose pseudoinverse.

Similarly, the optimal H can be computed as

$$\frac{\partial J}{\partial H} = S^T S H - S^T A_M. \tag{18}$$

Then, we can update S as

$$H = (S S^T)^\dagger S^T A_M. \tag{19}$$

In this study, we use the softmax classifier for brain tumor image classification. The optimal classifier parameter θ can be computed as

$$\frac{\partial J}{\partial \theta_c} = -\frac{1}{M} \sum_{i=1}^M \sum_{c=1}^C l_{c,i} \left(1 - \frac{e^{\theta_c^T a_{M,i}}}{\sum_{u=1}^C e^{\theta_u^T a_{M,i}}} \right) \alpha_{M,i}^T. \tag{20}$$

TABLE 3 | Confusion matrix of CDLLC on the REMBRANDT dataset.

	AST	OLI	GBM
AST	0.9686	0.0100	0.0214
OLI	0.0621	0.9127	0.0252
GBM	0.0583	0.0108	0.9309

TABLE 4 | Classification performance of CDLLC on the REMBRANDT dataset.

	Accuracy	Recall	Precision	F1-score	Balance loss	
Fold-1	Training	97.74	93.89	95.42	94.25	95.86
	Test	97.55	93.67	95.28	94.10	95.39
Fold-2	Training	97.80	93.92	95.53	94.34	95.91
	Test	97.74	93.90	95.38	94.21	95.39
Fold-3	Training	97.87	93.96	95.52	94.35	95.82
	Test	97.74	93.89	95.39	94.22	95.56
Fold-4	Training	97.85	93.97	95.58	94.47	95.99
	Test	97.80	93.92	95.52	94.34	95.71
Fold-5	Training	97.74	93.90	95.48	94.26	95.79
	Test	97.37	93.66	95.12	94.05	95.42

Testing Brain Tumor MR Images

Let x be the feature descriptor of a test image, based on the learned optimal S, H, θ , and $\{D_1, D_2, \dots, D_M\}$, we can compute the dictionary encoding of x by the following formulation:

$$\min_{a_M} \|x - D_1 \phi(D_2 \phi(\dots \phi(D_M a_M)))\|_F^2 + \lambda_1 \|a_M - S H\|_F^2. \tag{21}$$

Finally, the softmax classifier is used for classification, and the label probability of x assigned to class c can be computed as

$$p_c = \frac{e^{\theta_c^T a_M}}{\sum_{u=1}^C e^{\theta_u^T a_M}}. \tag{22}$$

EXPERIMENTS

Experimental Settings

In our experiment, we randomly select 1000 images in the REMBRANDT dataset and the whole Cheng dataset for simulation. Brain MR images are resized to 227×227 sizes to use for AlexNet. All convolution layers in CDLLC employ filters of size 3×3 . Stochastic gradient descent (SGD) is used as an optimizer and TensorFlow is implemented in model training. ReLU is used as the activation function. The initialization of dictionary is performed from D_1 and A_1 . We run K-SVD algorithm on each class of data and integrate subclass dictionary into the dictionary D_1 . Then, we obtain the D_m and A_m based on the learned dictionary and coding from the previous layer. The parameters λ_1, λ_2 , and λ_3 in CDLLC are selected in the search grid $\{0.001, 0.01, 0.1, 1, 10\}$. We use the fivefold cross-validation to complete with valid and comparable results. The 70% of the training fold are used for training and 30% are used as validation. We perform experiments for 10 times and then record their average values. We summarize the performance of all comparative methods in terms of accuracy, F1-score, precision, recall, and balance loss (Gu et al., 2018, 2020a).

There are many categories of methods used for brain tumor MR image classification. We compare several traditional machine learning and deep learning methods in the experiments. The traditional machine learning methods include the support vector machine with RBF kernel (SVM-RBF) (Nandpuru et al., 2014), the dictionary learning method label consistent K-SVD (called LC-KSVD1 in the experiment) (Jiang et al., 2013), and the neural network fuzzy inference system (ANFIS) (Thirumurugan and Shanthakumar, 2016). The deep learning classification methods include CNN (Krizhevsky et al., 2012) and CapsNet (Afshar et al., 2018). For these traditional classification methods, feature selection is a key step, where first- and second-order statistical texture features are widely used in brain tumor image classification. In our experiment, we follow (Fujima et al., 2019; Zhang et al., 2019) and use the statistical texture features: mean, variance, standard deviation, skewness, kurtosis, contrast, energy, entropy, correlation, and homogeneity. In addition, to compare with LC-KSVD using deep features, we exploit features from CNN architecture AlexNet and apply these features into LC-KSVD (called LC-KSVD2 in the experiment). We conduct the experiments on a computer with Intel Xeon Processor E5-2620

v4 and 64 GB RAM. All methods are implemented on Python 2.7, using Keras library and Tensor Flow.

Experimental Results on Cheng Dataset

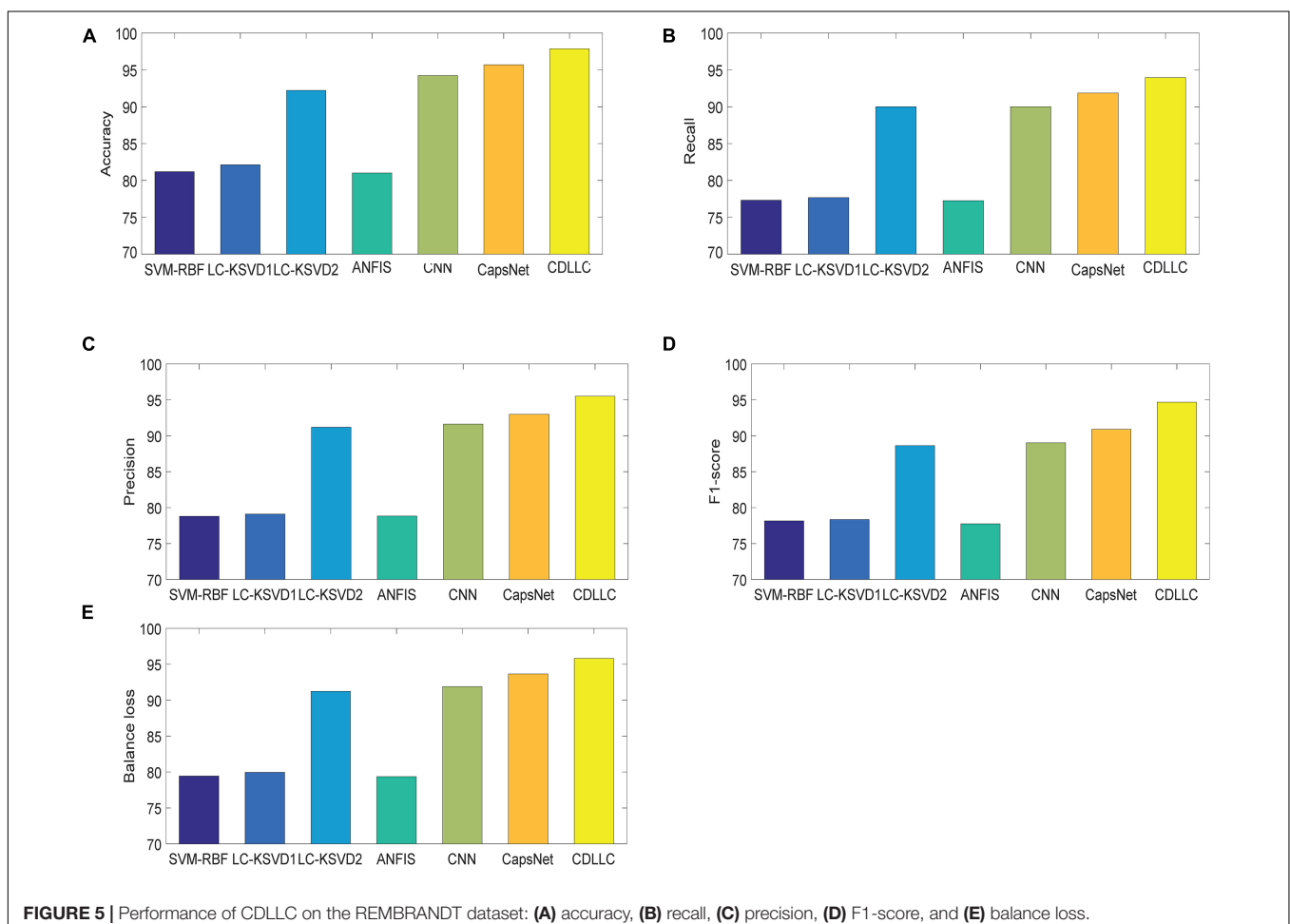
In this subsection, we observe the classification performance of CDLLC on the Cheng dataset. The images of the Cheng dataset used in our experiment contain three types of brain tumors: meningioma, glioma, and pituitary. First, we show the confusion matrix for three classes obtained by CDLLC in **Table 1**. The confusion matrix provides the valuable information about the predicted labels. From **Table 1**, we can see that the pituitary tumor is classified with the highest accuracy, glioma tumors are classified with the second accuracy, and meningioma tumors are classified with the lowest accuracy. Generally, the overall classification accuracy is satisfactory.

Second, we use accuracy, F1-score, precision, recall, and balance loss as the evaluation indexes. The performance of these five indexes of CDLLC in five folds is shown in **Table 2** in detail. **Table 2** shows that CDLLC obtains high average values and small standard deviation on accuracy, F1-score, precision, recall, and balance loss. Then, we compare our method with RBF-SVM, LC-KSVD1, LC-KSVD2, ANFIS, CNN, and CapsNet. The average experimental results on the Cheng

dataset are shown in **Figure 4**. It can be observed from these results that (1) CDLLC achieves the best results in comparison with other methods. This indicates that more discriminative information can be exploited by the proposed method. It suggests that the multi-layer dictionary learning, which addresses both feature representation and encoding in the nonlinear space, can exploit discriminative features from deep learning structure. In addition, graph Laplacian regularization can preserve the locality structure information of sparse codes, which can largely improve the model discriminative ability. (2) Among all methods, deep learning methods gain better classification performance than traditional machine learning (SVM-RBF, LC-KSVD1, and ANFIS) with statistical texture features. Using the deep features, the classification performance of LC-KSVD2 is obviously improved than LC-KSVD1. It indicates that deep features are more adapted to brain tumor image classification.

Experimental Results on REMBRANDT Dataset

In this subsection, we observe the classification performance of CDLLC on the REMBRANDT dataset. The images of the REMBRANDT dataset used in our experiment contain three types of brain tumors: AST, OLI, and GBM.

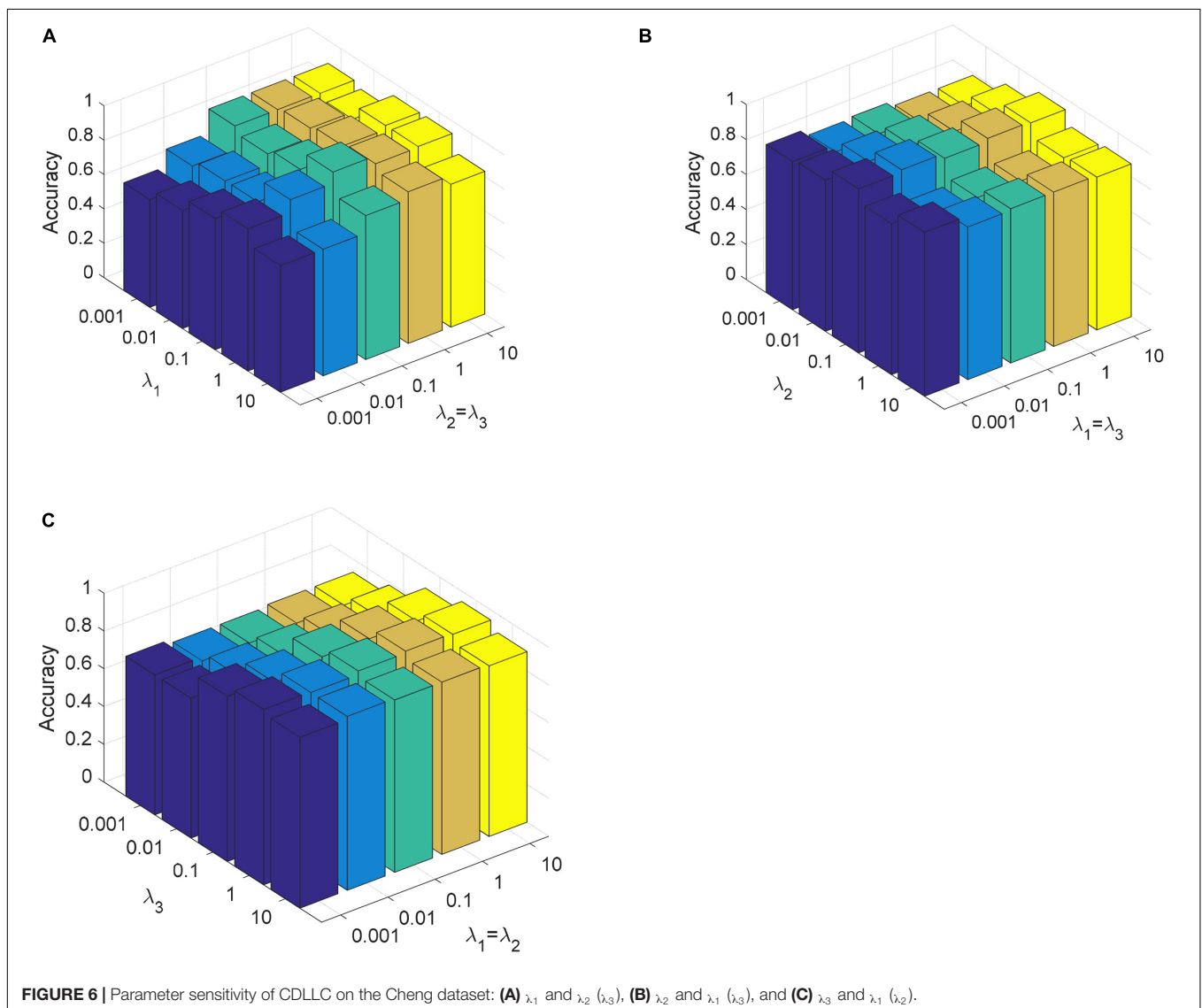


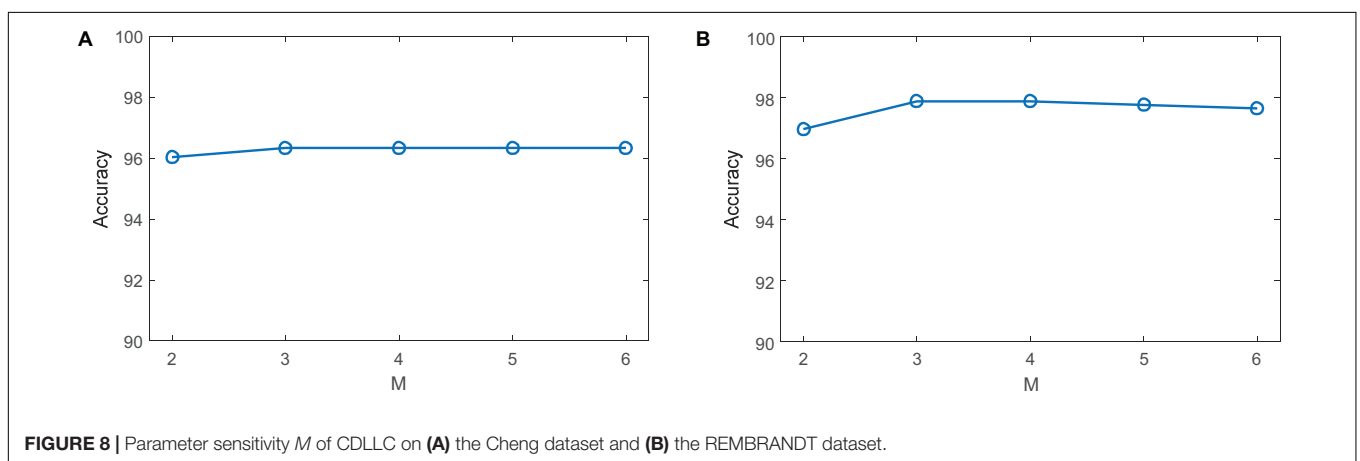
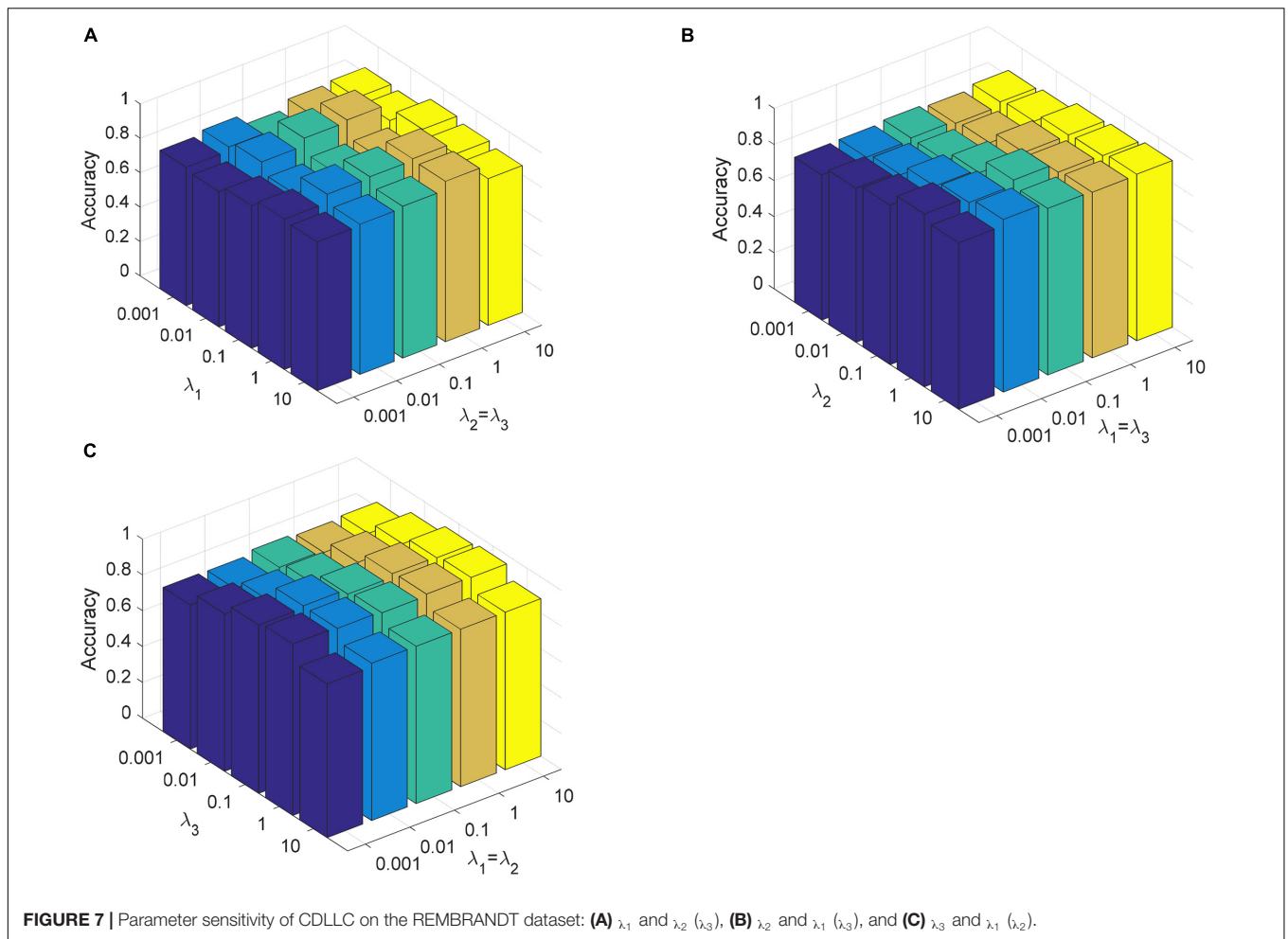
We first show the confusion matrix for three classes obtained by CDLLC in **Table 3**. From **Table 3**, we can see that the classification performance of CDLLC for three types of brain tumor images is comparable. The classification rates of AST, OLI, and GBM are 0.9686, 0.9127, and 0.9309, respectively. Second, we summarize the performance of CDLLC in terms of accuracy, F1-score, precision, recall, and balance loss in **Table 4**. From the fivefold results in **Table 4**, we can see that CDLLC gains the satisfactory results on the REMBRANDT dataset. Our multi-layer dictionary structure is not only convolutional but also sparse, and all parameters are updated within joint optimization learning. Then, we compare our method with RBF-SVM, LC-KSVD, ANFIS, CNN, and CapsNet. The average experimental results on the REMBRANDT dataset are shown in **Figure 5**. Similar to the results in **Figure 4**, CDLLC gains the best performance in five evaluation indexes. LC-KSVD1 is as a baseline method of our method. Whether using statistical texture features or deep features, the performance of LC-KSVD1 is much lower than that

of the proposed CDLLC. The reason is that LC-KSVD1 learns the dictionary in the original space and such dictionary cannot well exploit the discriminative information.

Parameter Analysis

In this section, we analyze the parameter sensitivity of CDLLC on the Cheng and REMBRANDT datasets. First, we discuss the parameters λ_1 , λ_2 , and λ_3 in CDLLC. These three parameters are selected in the search grid $\{0.001, 0.01, \dots, 10\}$. We set $\lambda_1 = \lambda_2$ and visualize the change of classification accuracy of CDLLC with different values of λ_2 and λ_3 . Similarly, we set $\lambda_2 = \lambda_3$ ($\lambda_1 = \lambda_3$) and visualize the change of classification accuracy of CDLLC with different values of λ_1 and λ_3 (λ_1 and λ_2). The results of classification accuracy are shown in **Figures 6, 7**. We can see that different values of the parameters λ_1 , λ_2 , and λ_3 have a significant impact on the classification accuracy of CDLLC. It indicates that the grid search strategy is appropriate for λ_1 , λ_2 , and λ_3 .





Next, we discuss the number of layers M in CDLLC on the Cheng and REMBRANDT datasets. We set M in grid $\{2, 3, \dots, 6\}$ to evaluate its effect on classification accuracy. The classification result is shown in **Figure 8**. We see that the classification accuracy of CDLLC improves when M increases from 1 to 3. When M is greater than 3, the classification accuracy of CDLLC is reliable on the Cheng and REMBRANDT datasets.

CONCLUSION

In this study, we propose CDLLC method for brain tumor MR image classification. The CNN structure is utilized to seek sparse representation in the nonlinear space, so that the resulting coding vectors of different classes can give the discriminative approximation. Meanwhile, the proposed method CDLLC uses

the locality constraint of atoms to preserve the manifold structure of the codes. Different from the traditional dictionary learning that uses manual feature extraction, CDLLC extracts the useful CNN features automatically in the architecture of deep learning. Classification of types of meningiomas, gliomas, and pituitary tumors on the Cheng dataset and types of AST, OLI, and GBM on the REMBRANDT dataset is carried out with high performance in accuracy, recall, precision, F1-score, and balance loss. The shortcoming of CDLLC is that the selection of parameters becomes complicated as the number of layers increases. In our next work, we will try to design a more reasonable program to select these parameters. Besides, we will compare various network architectures on CDLLC, such as VGG and GoogLeNet. Also, we will adjust our method so that it could be applied to other medical MR images.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The Cheng dataset analyzed for this study can

be found at: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. The REMBRANDT dataset analyzed for this study can be found at: <https://wiki.cancerimagingarchive.net/display/Public/REMBRANDT>.

AUTHOR CONTRIBUTIONS

TN and XG conceived and developed the theoretical framework of the manuscript. All authors carried out experiment and data process, and drafted the manuscript.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under grant 61806026 and by the Natural Science Foundation of Jiangsu Province under grant BK20180956.

REFERENCES

- Adebileje, S. A., Ghasemi, K., Aiyelabegan, H. T., and Saligheh Rad, H. (2017). Accurate classification of brain gliomas by discriminate dictionary learning based on projective dictionary pair learning of proton magnetic resonance spectra. *Magn. Reson. Chem.* 55, 318–322. doi: 10.1002/mrc.4532
- Afshar, P., Mohammadi, A., and Plataniotis, K. N. (2018). “Brain tumor type classification via capsule networks,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, (New Jersey: IEEE), 3129–3133. doi: 10.1109/ICIP.2018.8451379
- Al-Shaikhli, S. D. S., Yang, M. Y., and Rosenhahn, B. (2014). “Coupled dictionary learning for automatic multi-label brain tumor segmentation in flair MRI images,” in *In Proceedings of the International Symposium on Visual Computing*, (Cham: Springer), 489–500. doi: 10.1007/978-3-319-1424-9_4_46
- Al-Shaikhli, S. D. S., Yang, M. Y., and Rosenhahn, B. (2016). Brain tumor classification and segmentation using sparse coding and dictionary learning. *Biomed. Tech.* 61, 413–429. doi: 10.1515/bmt-2015-0071
- Anaraki, A. K., Ayati, M., and Kazemi, F. (2019). Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybern. Biomed. Eng.* 39, 63–74. doi: 10.1016/j.bbe.2018.10.004
- Bunevicius, A., Schregel, K., Sinkus, R., Golby, A., and Patz, S. (2020). Review: MR elastography of brain tumors. *NeuroImage Clin.* 25:102109. doi: 10.1016/j.nicl.2019.102109
- Cai, X., Ding, C., Nie, F., and Huang, H. (2013). “On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions,” in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York: ACM), 1124–1132. doi: 10.1145/2487575.2487701
- Chahal, P. K., Pandey, S., and Goel, S. (2020). A survey on brain tumor detection techniques for MR images. *Multimed. Tools Appl.* 79, 21771–21814. doi: 10.1007/s11042-020-08898-3
- Chen, X., Nguyen, B. P., Chui, C. K., and Ong, S. H. (2017). An automatic framework for multi-label brain tumor segmentation based on kernel sparse representation. *Acta Polytechnica Hungarica* 14, 25–43. doi: 10.12700/APH.14.1.2017.1.3
- Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., et al. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS One* 10:e0140381. doi: 10.1371/journal.pone.0140381
- Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., et al. (2016). Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PLoS One* 11:e0157112. doi: 10.1371/journal.pone.0157112
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* 26, 1045–1057. doi: 10.1007/s10278-013-9622-7
- Deepak, S., and Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Comput. Biol. Med.* 111:103345. doi: 10.1016/j.compbimed.2019.103345
- Ding, Z., and Fu, Y. (2018). Deep transfer low-rank coding for cross-domain learning. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 1768–1779. doi: 10.1109/TNNLS.2018.2874567
- Fujima, N., Homma, A., Harada, T., Shimizu, Y., Tha, K. K., Kano, S., et al. (2019). The utility of MRI histogram and texture analysis for the prediction of histological diagnosis in head and neck malignancies. *Cancer Imaging* 19:5. doi: 10.1186/s40644-019-0193-9
- Ge, C., Gu, I. Y. H., Jakola, A. S., and Yang, J. (2020). Deep semi-supervised learning for brain tumor classification. *BMC Med. Imaging* 20:1–11. doi: 10.1186/s12880-020-00485-0
- Ghasemi, M., Kelarestaghi, M., Eshghi, F., and Sharifi, A. (2020). T2-FDL: a robust sparse representation method using adaptive type-2 fuzzy dictionary learning for medical image classification. *Expert Syst. Appl.* 158:113500. doi: 10.1016/j.eswa.2020.113500
- Ghasemi, M., Kelarestaghi, M., Eshghi, F., and Sharifi, A. (2021). A FDL: a new adaptive fuzzy dictionary learning for medical image classification. *Pattern Anal. Appl.* 24, 145–164. doi: 10.1007/s10044-020-00909-1
- Gu, X., Chung, F. L., and Wang, S. (2018). Fast convex-hull vector machine for training on large-scale ncRNA data classification tasks. *Knowl. Based Syst.* 151, 149–164. doi: 10.1016/j.knsys.2018.03.029
- Gu, X., Chung, F. L., and Wang, S. (2020a). Extreme vector machine for fast training on large data. *Int. J. Mach. Learn. Cybern.* 11, 33–53. doi: 10.1007/s13042-019-00936-3
- Gu, X., Zhang, C., and Ni, T. (2020b). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2020.3006699 [Epub Online ahead of print].
- Gumaei, A., Hassan, M. M., Hassan, M. R., Alelaiwi, A., and Fortino, G. (2019). A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. *IEEE Access* 7, 36266–36273. doi: 10.1109/ACCESS.2019.2904145

- Jiang, Z., Lin, Z., and Davis, L. S. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105. doi: 10.1145/3065386
- Li, H., He, X., Tao, D., Tang, Y., and Wang, R. (2018). Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recognit.* 79, 130–146. doi: 10.1016/j.patcog.2018.02.005
- Mittal, M., Goyal, L. M., Kaur, S., Kaur, I., Verma, A., and Hemanth, D. J. (2019). Deep learning based enhanced tumor segmentation approach for MR brain images. *Appl. Soft Comput.* 78, 346–354. doi: 10.1016/j.asoc.2019.02.036
- Mohan, G., and Subashini, M. M. (2018). MRI based medical image analysis: survey on brain tumor grade classification. *Biomed. Signal Process. Control* 39, 139–161. doi: 10.1016/j.bspc.2017.07.007
- Nandpuru, H. B., Salankar, S. S., and Bora, V. R. (2014). “MRI brain cancer classification using support vector machine,” in *Proceedings of the 2014 IEEE Students’ Conference on Electrical, Electronics and Computer Science*, (New Jersey: IEEE), 1–6. doi: 10.1109/SCECS.2014.6804439
- Ni, T., Gu, X., and Jiang, Y. (2020). *Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy*. Netherland: Springer.
- Papayan, V., Romano, Y., Sulam, J., and Elad, M. (2018). Theoretical foundations of deep learning via sparse representations: a multilayer sparse model and its connection to convolutional neural networks. *IEEE Signal Process. Mag.* 35, 72–89. doi: 10.1109/MSP.2018.2820224
- Peng, Y., Liu, S., Wang, X., and Wu, X. (2020). Joint local constraint and fisher discrimination based dictionary learning for image classification. *Neurocomputing* 398, 505–519. doi: 10.1016/j.neucom.2019.05.103
- Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., and Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J. Comput. Sci.* 30, 174–182. doi: 10.1016/j.jocs.2018.12.003
- Song, J., Xie, X., Shi, G., and Dong, W. (2019). Multi-layer discriminative dictionary learning with locality constraint for image classification. *Pattern Recognit.* 91, 135–146. doi: 10.1016/j.patcog.2019.02.018
- Sulam, J., Papayan, V., Romano, Y., and Elad, M. (2018). Multilayer convolutional sparse modeling: Pursuit and dictionary learning. *IEEE Trans. Signal Process.* 66, 4090–4104. doi: 10.1109/TSP.2018.2846226
- Sun, R., Wang, K., Guo, L., Yang, C., Chen, J., Ti, Y., et al. (2019). A potential field segmentation based method for tumor segmentation on multi-parametric MRI of glioma cancer patients. *BMC Med. Imaging* 19:48. doi: 10.1186/s12880-019-0348-y
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: gLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 1–41. doi: 10.3322/caac.21660
- Thirumurugan, P., and Shanthakumar, P. (2016). Brain tumor detection and diagnosis using ANFIS classifier. *Int. J. Imaging Syst. Technol.* 26, 157–162. doi: 10.1007/978-3-642-31552-7_35
- Tong, J., Zhao, Y., Zhang, P., Chen, L., and Jiang, L. (2019). MRI brain tumor segmentation based on texture features and kernel sparse coding. *Biomed. Signal Process. Control* 47, 387–392. doi: 10.1016/j.bspc.2018.06.001
- Vu, T. H., Mousavi, H. S., Monga, V., Rao, G., and Rao, U. A. (2015). Histopathological image classification using discriminative feature-oriented dictionary learning. *IEEE Trans. Med. Imaging* 35, 738–751. doi: 10.1109/TMI.2015.2493530
- Wang, J., Kwon, S., and Shim, B. (2012). Generalized orthogonal matching pursuit. *IEEE Trans. Signal Process.* 60, 6202–6216. doi: 10.1109/TSP.2012.2218810
- Wu, G., Chen, Y., Wang, Y., Yu, J., Lv, X., Ju, X., et al. (2017). Sparse representation-based radiomics for the diagnosis of brain tumors. *IEEE Trans. Med. Imaging* 37, 893–905. doi: 10.1109/TMI.2017.2776967
- Yang, Y., Yan, L. F., Zhang, X., Han, Y., Nan, H. Y., Hu, Y. C., et al. (2018). Glioma grading on conventional MR images: a deep learning study with transfer learning. *Front. Neurosci.* 12:804. doi: 10.3389/fnins.2018.00804
- Zeng, K., Zheng, H., Cai, C., Yang, Y., Zhang, K., and Chen, Z. (2018). Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Comput. Biol. Med.* 99, 133–141. doi: 10.1016/j.compbiomed.2018.06.010
- Zhang, Y., Chen, C., Tian, Z., Feng, R., Cheng, Y., and Xu, J. (2019). The Diagnostic Value of MRI-Based Texture Analysis in Discrimination of Tumors Located in Posterior Fossa: a Preliminary Study. *Front. Neurosci.* 13:1113. doi: 10.3389/fnins.2019.01113

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Gu, Shen, Xue, Fan and Ni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.