

## OPEN ACCESS

## EDITED BY

Sandeep Kumar Mishra,  
Yale University, United States

## REVIEWED BY

Palash Ghosal,  
Sikkim Manipal University, India  
Prof. Surjeet Dalal,  
Amity University Gurgaon, India  
Fatma Taher,  
Zayed University, United Arab Emirates

## \*CORRESPONDENCE

Wenna Chen

✉ chenwenna0408@163.com

Ganqin Du

✉ dgq99@163.com

RECEIVED 19 October 2024

ACCEPTED 21 March 2025

PUBLISHED 11 April 2025

## CITATION

Chen W, Liu J, Tan X, Zhang J, Du G, Fu Q and Jiang H (2025) EnSLDe: an enhanced short-range and long-range dependent system for brain tumor classification. *Front. Oncol.* 15:1512739. doi: 10.3389/fonc.2025.1512739

## COPYRIGHT

© 2025 Chen, Liu, Tan, Zhang, Du, Fu and Jiang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# EnSLDe: an enhanced short-range and long-range dependent system for brain tumor classification

Wenna Chen<sup>1\*</sup>, Junqiang Liu<sup>2</sup>, Xinghua Tan<sup>2</sup>, Jincan Zhang<sup>2</sup>, Ganqin Du<sup>1\*</sup>, Qizhi Fu<sup>1</sup> and Hongwei Jiang<sup>1</sup>

<sup>1</sup>The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China, <sup>2</sup>College of Information Engineering, Henan University of Science and Technology, Luoyang, China

**Introduction:** Brain tumors pose significant harm to the functionality of the human nervous system. There are lots of models which can classify brain tumor type. However, the available methods did not pay special attention to long-range information, which limits model accuracy improvement.

**Methods:** To solve this problem, in this paper, an enhanced short-range and long-range dependent system for brain tumor classification, named as EnSLDe, is proposed. The EnSLDe model consists of three main modules: the Feature Extraction Module (FExM), the Feature Enhancement Module (FEnM), and the Classification Module. Firstly, the FExM is used to extract features and the multi-scale parallel subnetwork is constructed to fuse shallow and deep features. Then, the extracted features are enhanced by the FEnM. The FEnM can capture the important dependencies across a larger sequence range and retain critical information at a local scale. Finally, the fused and enhanced features are input to the classification module for brain tumor classification. The combination of these modules enables the efficient extraction of both local and global contextual information.

**Results:** In order to validate the model, two public data sets including glioma, meningioma, and pituitary tumor were validated, and good experimental results were obtained, demonstrating the potential of the model EnSLDe in brain tumor classification.

## KEYWORDS

brain tumor classification, feature extraction, feature enhancement, long-range dependencies, attention

## 1 Introduction

The brain is the control center of the body, in addition to maintaining the normal activities of our lives, it also controls our daily senses (hearing, sight, smell, etc.), cognition, memory, thinking, emotions, and many other aspects of our lives (1). Undoubtedly, the brain holds paramount importance in our lives. However, brain tumors stand as one of the

most prevalent afflictions of the nervous system, capable of significantly impairing its functionality. Timely detection of brain tumors is essential for enhancing and prolonging patient survival rates (2, 3). Tumors growing within the skull are generally known as brain tumors, which encompass primary brain tumors originating from brain tissue and secondary tumors that metastasize to the skull from elsewhere in the body (4). The common types of brain tumors include gliomas, meningiomas, and pituitary tumors (5).

Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) are two widely used imaging techniques in medicine that play an important role in labelling abnormalities in the shape, size or location of the brain (6). While CT is limited to cross-sectional imaging, MRI offers the flexibility to image in various orientations, including transverse, sagittal, coronal, and any desired section. Additionally, MRI excels in providing clearer differentiation of soft tissues in three dimensions compared to conventional imaging methods. These advantages have made MRI the most favored method among physicians and have led to increasing interest among researchers. However, the analysis of MRI images by medical professionals to discern the type of tumor is a complex and time-intensive process. The accuracy of their diagnosis can be influenced by the subjective expertise and skills of the physician (7, 8). It is well known that early detection and timely treatment are crucial for the recovery of brain tumor patients (9). If the type of brain tumor can be accurately and early identified, it will greatly increase the patient's valuable treatment time and thus significantly improve the likelihood of recovery.

Traditional Machine Learning (ML) has been widely used for classification problems in Computer-Aided Diagnostic (CAD) systems (10, 11). For example, Singh et al. (12) proposed a new classification method using generalized discriminant analysis and the 1-norm linear programming extreme learning machine. Shahid et al. (13) used a feature selection algorithm to find the effective feature subset, which was then used for classification by an Extreme Learning Machine (ELM) based on hybrid particle swarm optimization. Xie et al. (14) used the combination of Support Vector Machine (SVM) and ELM for feature selection, and the optimal features were used by the classifier to distinguish breast tumor types. Heidari et al. (15) applied stochastic projection algorithm to optimize the constructed SVM model embedded with multiple feature dimensionality reduction methods to improve the classification performance of the model.

Deep learning stands as a cutting-edge innovation in classification and prediction, showcasing outstanding performance in domains necessitating multi-level data processing such as classification, detection, and speech recognition (16). Deep learning has the capability to learn features from extensive image data and extract high-level features from images through layer-by-layer convolution and pooling operations, achieving automatic classification of brain tumors. Compared to traditional image processing methods, deep learning boasts superior feature extraction capability, higher classification accuracy, as well as automation and intelligence. In recent years, many studies have explored the application of deep learning in diagnosing various diseases. For example, Sarki et al. (17) classified mild and multiple

diabetic eye diseases by fine-tuning and optimizing the VGG16 model. Jeong et al. (18) used Inception V3 deep learning model to classify the presence or absence of cardiac enlargement, and the classification accuracy reached 96.0%. Chowdhury et al. (19) adopted the improved Xception model to diagnose hair and scalp diseases and achieved a high accuracy rate. Sharifrazi et al. (20) used Convolutional Neural Network (CNN) combined with k-means clustering method to automatically diagnose myocarditis, with an accuracy of 97.41%. The lesion area in brain tumor images constitutes only a small portion of the entire image. Furthermore, when distinguishing between types of brain tumors, both the tumor region and its surrounding area exert a significant impact on the classification results (21). In addition, multi-scale feature fusion has been widely applied to object detection, image segmentation, image classification, and other fields. Multi-scale networks are capable of simultaneously extracting features at different scales in images, thereby more comprehensively capturing the details and overall information of target objects. For example, in object detection tasks, small-scale features can be used to detect small objects, while large-scale features are helpful for detecting large objects. Features at different scales provide different contextual information, and multi-scale networks can effectively integrate this information, offering a more comprehensive and rich visual context. Multi-scale networks can handle input data at different scales, and this characteristic significantly enhances the algorithm's robustness and generalization performance in complex scenarios (22). A common method for multi-scale feature fusion is the pyramid structure. The pyramid structure extracts features at different scales and then fuses these features to obtain a more comprehensive feature representation. Specifically, improved methods based on the Feature Pyramid Network (FPN) architecture achieve deep integration of cross-scale features by constructing multi-level pyramid-like feature representations (23, 24).

However, most previous studies did not pay special attention to the surrounding areas of tumors, i.e., lacking the ability to capture long-range information, which would affect the performance of classification. To overcome the shortcoming, this study proposes a new multi-class brain tumor classification model with enhanced short-range and long-range dependence, named as EnSLDe. The model not only has the ability to capture short-range and long-range dependencies, but also retains local key information. It consists of three main modules: the Feature Extraction Module (FExM), the Feature Enhancement Module (FEnM), and the classification module. Within the FExM, convolutional layers are combined with residual connections to extract features, while incorporating an Effective Multi-scale Attention (EMA) mechanism that simultaneously focuses on channel-wise and spatial information. The FEnM further strengthens feature representation, enabling capture of crucial long-range dependencies while retaining key information within the local range. The classification module adopts a two-layer fully connected structure combined with dropout regularization for brain tumor classification. This approach enhances the model's generalization ability, reducing the risk of overfitting, and further improves the classification performance of the model. We utilized

two datasets to evaluate the model performance: a three-category dataset comprising gliomas, meningiomas, and pituitary tumors, and a four-category dataset including additional healthy categories.

The main contributions of this study are as follows:

- A new model with enhanced short-range and long-range dependence is proposed to classify brain tumor images from MRI.
- FExM is used to extract features from brain tumor images. The EMA module of FExM integrates channel attention and spatial attention to provide a more comprehensive feature representation.
- The FEnM is used to capture important dependencies across larger sequence scales. And it can also cooperate with the global adjustment network to fuse the retained local information with different levels of deep features.
- EnSLDe employs multi-scale parallel subnetworks that integrate shallow and deep features. This architecture enables the model to capture comprehensive contextual information across varying scales, which is critical for distinguishing between diverse tumor types.
- Based on experimental results using two public datasets, the proposed method exhibits excellent performance.

## 2 Related works

Classification of brain tumors is critical for evaluating tumors and determining treatment options for patients. There are already many CAD systems used in medical industries to help doctors make diagnoses. There have been many methods to classify brain tumors, which can be roughly divided into traditional ML methods, deep learning methods, and hybrid methods.

In the past, traditional ML has been used to classify brain tumors. For example, Bansal and Jindal (25) utilized a combination of grayscale co-occurrence matrix technology and shape-based feature technology to extract mixed features from the tumor area. Subsequently, a hybrid classifier consisting of Random Forest Classifier (RFC), K Nearest Neighbors (KNN) classifier, and Decision Tree (DT) classifier was used to classify brain tumors. 26 performed image segmentation through a marker-based watershed algorithm, then combined features with a sequence-based cascade method, and finally used SVM for classification.

In traditional ML, relevant domain knowledge is needed for feature extraction, while features can be automatically extracted by deep learning. The development of deep learning methods has had a significant impact on the field of medical image analysis applications, especially in disease diagnosis (27). Recently, deep learning has achieved remarkable results in brain tumor classification. For example, Raza et al. (28) proposed a hybrid deep learning model based on the GoogLeNet architecture. The last five layers of GoogLeNet were removed and 15 new layers were added to achieve high accuracy. Díaz-Pernas et al. (29) proposed a multi-scale processing based on CNN architecture design for brain

tumor classification. The elastic transformation data expansion method was used to increase the training dataset and prevent over-fitting. Finally, 97.3% classification accuracy was achieved. Ayadi et al. (30) proposed an innovative brain tumor classification model based on CNN architecture, automated processing and minimizing preprocessing requirements. To fully evaluate the accuracy of the model, it was tested on three different brain tumor datasets. Various performance indicators are analyzed in depth. Sreenivasa Reddy and Sathish (31) proposed a brain tumor classification and segmentation scheme based on deep structured architecture. Firstly, adaptive ResUNet3+ with multi-scale convolution was used to process the collected data. Then, the parameters of the deep learning method were optimized and adjusted through the arithmetic optimization algorithm accelerated by the improved mathematical optimizer. Finally, an attention-based ensemble convolutional network was introduced for brain tumor classification. The model demonstrated excellent performance in both segmentation and classification accuracy. P. Ghosal et al. (32) integrated the residual network architecture with the Squeeze and Excitation block to enhance feature extraction and refinement. Islam et al. (33) optimized the EfficientNet series for the purpose of brain tumor classification, with EfficientNetB3 demonstrating superior performance. Aurna et al. (34) utilized multiple MRI datasets and performed feature extraction by combining pre-trained models and newly designed CNN models. Among the extracted features, Principal Component Analysis (PCA) was used to select key features and input them into the classifier. Musallam et al. (35) proposed a three-step preprocessing to improve the quality of MRI images and a new Deep Convolutional Neural Network (DCNN) architecture with 10 convolutional layers. Kumar and Sasikala (36) fused the features extracted from the shallow and deep layers of the pre-trained Resnet18 network, and then adopted a hybrid classifier composed of SVM, KNN, and DT optimized by the Bayesian algorithm perform classification.

In addition, in order to further improve the accuracy and efficiency of brain tumor classification models, optimization algorithms could be used in deep learning. For example, Alshayegi et al. (37) attained a classification accuracy of 97.374% for automatic brain tumor classification by combining the layers of two CNN architectures and fine-tuning the hyperparameters through Bayesian optimization. Irmak (38) used CNN and grid search optimization algorithms to propose three different CNN models to complete three different classification tasks. Almost all hyperparameters in the model were tuned by grid search optimization algorithms. Rammurthy and Mahesh (39) used Whale Harris Hawks Optimization (WHHO), which was a combination of Whale Optimization Algorithm (WOA) and Harris Hawks Optimization (HHO) to optimize the deep convolutional network. Alyami et al. (40) used deep convolutional networks and the slap swarm algorithm to classify brain tumors from brain MRI. To enhance the accuracy of classification, an efficient feature selection technique—the slap swarm algorithm was introduced. This technique helps to identify key features that significantly influence the classification results while excluding

those with minor contributions, thereby ensuring that the classification model achieves optimal accuracy.

It is noteworthy that Transformer models have also been employed in brain tumor classification tasks. Sudhakar Tummala et al. (41) investigated the capability of pretrained and fine-tuned Vision Transformer (ViT) models for brain tumor classification using MRI images. GAZI JANNATUL FERDOUS et al. (42) proposed a novel Linear Complexity Data-efficient Image Transformer (LCDEiT). The LCDEiT adopts a teacher-student strategy, where the teacher model is a customized gated pooling convolutional neural network (CNN) responsible for transferring knowledge to the transformer-based student model. The student model achieves linear computational complexity through an external attention mechanism. Asiri et al. (43) employed Swin Transformer for multi-class brain tumor classification. Tapas Kumar Dutta et al. (44) developed GT-Net for brain tumor classification tasks. The core component of this model is the Global Transformation Module (GTM), which contains multiple Generalized Self-Attention Blocks (GSB) designed to explore long-range global feature relationships between lesion regions.

These studies, whether based on traditional ML methods, deep learning approaches, or hybrid methodologies, have achieved notable success in brain tumor classification. Many deep learning models (e.g., CNNs) automatically extract features but typically focus on local or global information rather than both. For instance, architectures like Inception-v3, ResNet, and DenseNet demonstrate strong performance yet generally emphasize localized details or global context without comprehensive integration. Hybrid approaches combining traditional machine learning and deep learning techniques may still fail to fully exploit multi-scale feature fusion or advanced attention mechanisms. While some models employ attention mechanisms, they often prioritize either channel-wise or spatial attention. This paper proposes a novel model named EnSLDe (Enhanced Short- and Long-range Dependency Extractor), designed to strengthen both short-term and long-range dependencies while preserving essential local information. EnSLDe uniquely integrates short- and long-range

dependencies through its FExM and FEnM. This dual processing proves critical for concurrently capturing localized tumor details and global contextual patterns in brain MRI images.

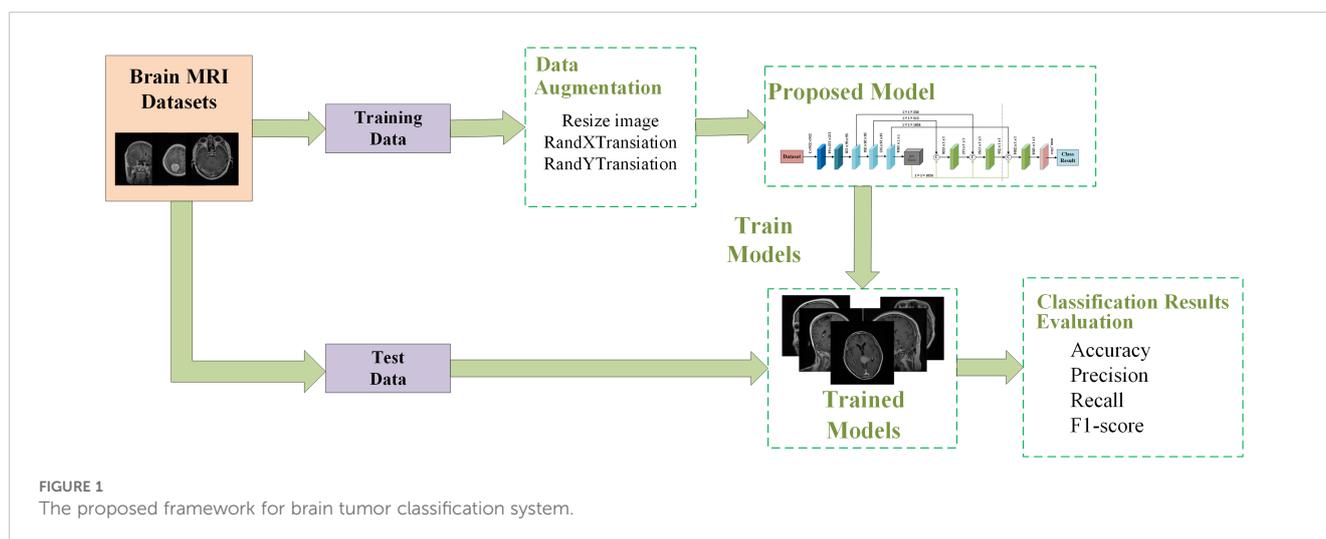
### 3 Proposed method

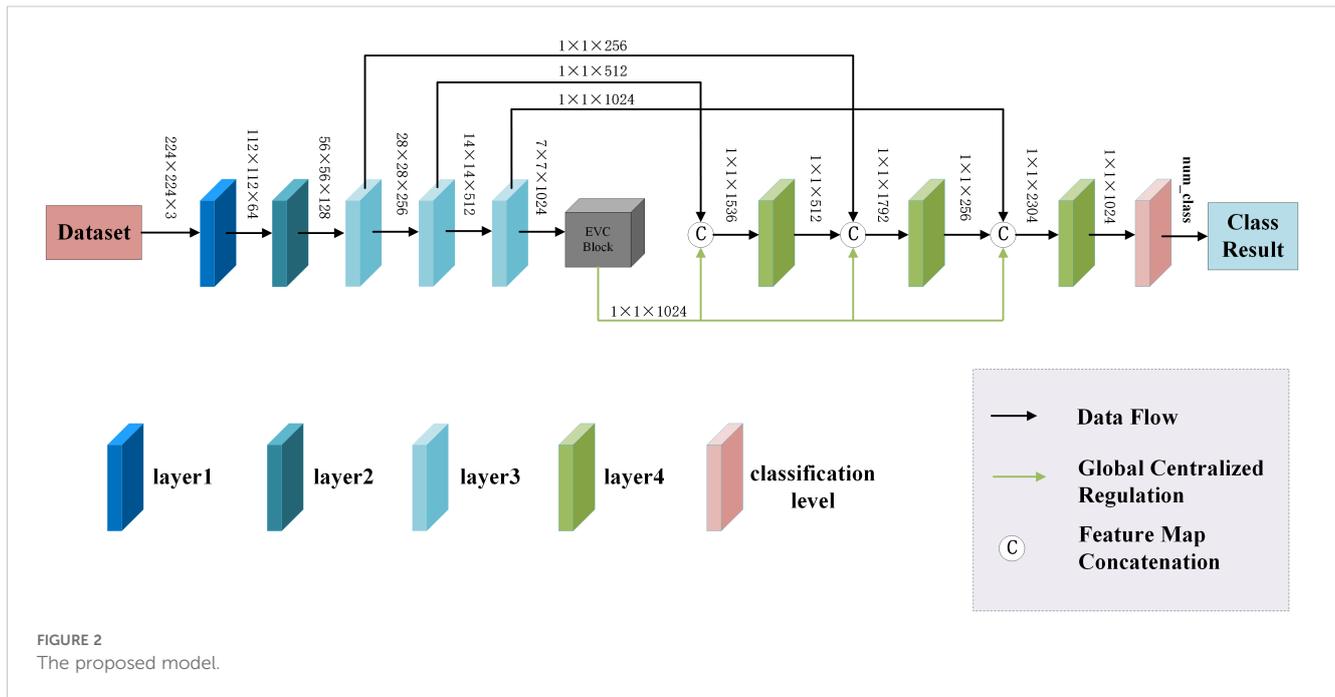
This section introduces our proposed brain tumor classification framework, which is shown in Figure 1. The training and testing phase of the proposed system works as follows:

1. The brain MRI dataset is divided into two disjoint sets: a training set and a test set.
2. Data augmentation techniques such as random rotation, random horizontal and vertical flipping are applied to the training dataset to mitigate overfitting issues.
3. The proposed network is trained by selecting appropriate hyperparameters and specifying the cross-entropy loss function.
4. Once training is completed, the trained model is saved.
5. The model is validated on a randomly partitioned test dataset, and the performance of the model is evaluated.

#### 3.1 Proposed brain tumor classification model

The EnSLDe consists of three main modules, namely feature extraction module, feature enhancement module and classification module, which is shown in Figure 2. Since both local and long-range dependent features play a crucial role in effectively classifying brain tumors from MRI images, the EnSLDe employs FExM and FEnM to extract and enhance these features. The classification module comprises two fully connected layers integrated with Dropout regularization, which enhances the model's generalization ability. Moreover, the stacked utilization of two fully connected layers can





amalgamate and transform features, thereby capturing more information and optimizing the representation capabilities of features to enhance model performance.

### 3.1.1 The feature extraction module

The feature extraction module consists of layer1, layer2, layer3-1, layer3-2, layer3-3, layer4-1, layer4-2, and layer4-3, and is used to extract multiple depth-level features from brain tumor images. The Feature Extraction Module (FExM) was designed to extract features from multiple intermediate layers to simultaneously capture short-range and long-range dependencies. This multi-scale parallel sub-network fuses shallow features (which retain fine-grained details) with deep features (encoding abstract, high-level contextual information). The selection of feature extraction layers was guided by empirical validation through ablation studies, which demonstrated that combining multiple layers achieved higher classification accuracy compared to those obtained using a single layer of features. Inspired by the C3 module in YOLOv5 and integrating the Effective Multi-scale Attention (EMA) proposed by (Ouyang et al. (45)), we have developed a novel Conv and Depthwise\_conv with EMA (CDE) module, as illustrated in Figure 3. The CDE module consists of a residual network and EMA. The structure of the residual network involves adding skip connections on top of the serial connection of two convolutional layers and a depthwise separable convolutional layer. This allows for the direct addition of input and output. Subsequently, the output features of the entire residual network are processed by EMA. Incorporating the residual network into the CDE module effectively alleviates the issues of gradient explosion or vanishing, making the model training process more stable and easier to optimize.

Additionally, depthwise separable convolution is used by CDE module, which significantly reduces computational costs while

maintaining powerful feature extraction capabilities, thus achieving a good balance between efficiency and performance. The inclusion of EMA allows the CDE module to form multi-scale parallel subnetwork while extracting features, which fuses shallow and deep features. This further enhances feature extraction and strengthens short-range and long-range dependencies. Moreover, it reshapes part of the channel dimensions into batch dimensions, effectively avoiding potential information loss caused by dimensionality reduction through conventional convolution. This improvement not only reduces computational overhead but also allows the model to focus more on extracting key features while retaining information from each channel. Layer1 consists of two convolutional layers and is mainly used to extract shallow image features. Layer2 consists of the residual network in the CDE module. layer3-1, layer3-2, and layer3-3 are all composed of CDE modules. Layer4-1, layer4-2, and layer4-3 are all composed of convolutional layers with a convolution kernel size of  $1 \times 1$ , which are used for channel dimensionality reduction after feature fusion.

The EMA divides the channel dimension of input feature maps into multiple sub-features and redistributes spatial-semantic features within each feature group. Specifically, EMA avoids traditional channel dimensionality reduction operations by reshaping the channel dimension into the batch dimension. This design enables EMA to model inter-channel dependencies through standard convolution operations without losing channel information. The EMA employs three parallel branches to extract attention weights:

1.  $1 \times 1$  Branch: Encodes channel attention along horizontal and vertical directions using two 1D global average pooling operations, thereby capturing long-range spatial dependencies while preserving precise positional information.

2. 3×3 Branch: Captures multi-scale feature representations through a 3×3 convolution kernel to expand the feature space.
3. Cross-Space Interaction: Fuses output feature maps from the two parallel branches via matrix dot product operations to capture pixel-level pairwise relationships and highlight global contextual information.

For an input feature  $X \in \mathbb{R}^{C \times H \times W}$ , it is first partitioned into  $G$  sub-features, each with a shape of  $(C/G) \times H \times W$ . In the 1×1 branch, two 1D feature vectors  $Z_H$  and  $Z_W$  are obtained by encoding channel attention through 1D global average pooling along horizontal and vertical directions, respectively.  $Z_H$  and  $Z_W$  can be calculated by Equation 1:

$$Z_H = \sum_{j=1}^H \chi_{c,j} \tag{1}$$

$$Z_W = \sum_{j=1}^H \chi_{c,j}$$

where,  $x_{c,i}$  and  $x_{c,j}$  denote the eigenvalues of the  $c$  channel in the horizontal and vertical directions, respectively. The vectors  $Z_H$  and  $Z_W$  are processed through 1×1 convolutions and the Sigmoid function to generate the channel attention maps  $A_H$  and  $A_W$ , can be calculated by Equation 2:

$$A_H = \sigma(\text{conv}(Z_H)) \tag{2}$$

$$A_W = \sigma(\text{conv}(Z_W))$$

Where,  $\sigma$  denotes the Sigmoid function. In the 3×3 branch, multi-scale feature representation  $F_{3 \times 3}$  is captured by the 3×3 convolution operation as shown in Equation 3:

$$F_{3 \times 3} = \text{Conv}_{3 \times 3}(X) \tag{3}$$

The final output feature map  $Y$  is obtained by fusing  $A_H$  and  $A_W$  matrix dot product is performed by  $F_{3 \times 3}$ , and the calculation formula is shown in Equation 4:

$$Y = \sigma(A_H \cdot A_W \cdot F_{3 \times 3}) \tag{4}$$

### 3.1.2 The feature enhancement module

The Explicit Visual Center (EVC) method (46) is used to enhance the features extracted by the model. The EVC can effectively extract global long-range dependencies from images while preserving crucial local information. The EVC combines a Multi-Layer Perceptron (MLP) based on top-level features with a Learnable Visual Center (LVC) mechanism, both of which operate in parallel to complement each other. The MLP is responsible for capturing the global long-range dependencies of the image, effectively addressing complex long-range dependency issues, and enhancing the model's perception of global information. Meanwhile, the LVC operates along the path of the MLP, focusing on preserving the crucial local information of the image to ensure that the model does not lose important local details while attending to the global context. For input  $F_{in}$ , the equation is calculated as follows (Equation 5):

$$F = \text{Cat}(\text{MLP}(F_{in}), \text{LVC}(F_{in})) \tag{5}$$

in the LVC model, the input ( $X$ ) is mapped to a set of ( $C$ )-dimensional features, ( $\{X_{in} = x_1, x_2, \dots, x_n\}$ , where ( $N=H \times W$ ) represents the total number of input features. Subsequently, LVC computes an intrinsic codebook ( $B = \{b_1, b_2, \dots, b_k\}$ ), which includes ( $K$ ) codewords (or visual centers) along with a set of smoothing factors ( $S = \{s_1, s_2, \dots, s_k\}$ ). The feature encoding is achieved through a series of convolutional layers. The encoded features are then matched against each codeword in the codebook. The discrepancies between the features and the codewords are computed, and learnable weights are derived from these

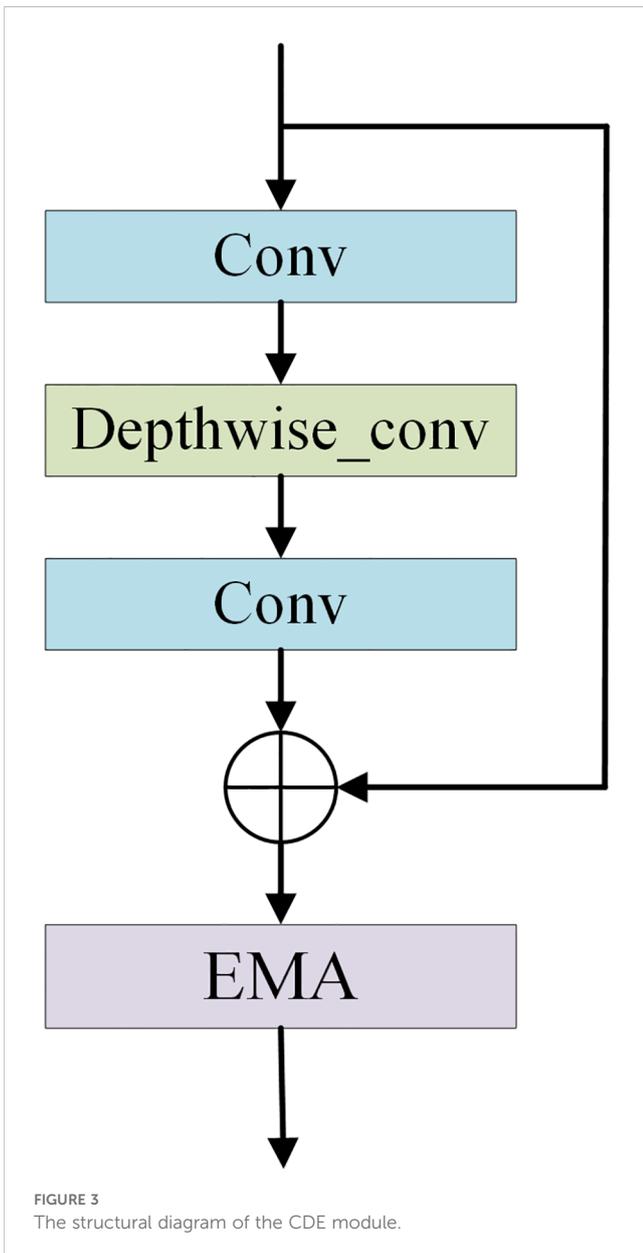


FIGURE 3 The structural diagram of the CDE module.

differences. The ultimate output is a (C)-dimensional vector (e) (Equation 6).

$$e_k = \sum_{i=1}^n \frac{e^{-S_k \|x_i - b_k\|^2}}{\sum_{j=1}^K e^{-S_k \|x_i - b_k\|^2}} (x_i - b_k) \tag{6}$$

The output of LVC is obtained by summing the features vector ( $X_{in}$ ) and the local features (Z) for each channel, as shown in Equation 7.

$$X_{out} = X_{in} \oplus Z \tag{7}$$

here, the local feature (Z) is derived by applying a Fully Connected (FC) layer that maps the feature (e) to an influence factor of dimensions  $C \times 1 \times 1$ . Subsequently, a channel-wise multiplication operation is conducted with ( $X_{in}$ ). The output following the Feature Enhancement Module is then obtained as follows (Equation 8):

$$F = Cat(X_{EVC}, X_d) \tag{8}$$

where, F represents the fusion feature,  $X_{EVC}$  denotes the feature output from the EVC, and  $X_d$  signifies the depth feature derived from various levels.

### 3.2 Loss function

The loss function we used during model training is the cross-entropy loss function (47). One can assume there are n classes, where the true label is represented by a K-dimensional vector  $y$  (with only one element being 1 and others being 0), and the model output probability is represented by a K-dimensional vector  $y'$  (with each element ranging from 0 to 1 and summing up to 1). The formula for multi-class cross-entropy loss function is defined as shown in Equation 9.

$$Loss = -\sum_{i=1}^n y_i \log y_i' \tag{9}$$

where, n is the number of categories,  $y_i$  is the i-th element of the true label vector  $y$ , and  $y_i'$  is the i-th element of the model output probability vector  $y_i$ .

The cross-entropy loss function is an efficient loss function in classification problems as it accurately measures the similarity between the true label distribution and the model's predicted

label distribution. Specifically, a smaller cross-entropy value indicates a closer resemblance between these two probability distributions, implying more accurate predictions by the model. When there is a significant disparity between the true and predicted distributions, the cross-entropy loss function yields a large loss value. This characteristic enables the model to update parameters more quickly during training, thus accelerating the learning process. The amplifying effect of the cross-entropy loss function makes the model more sensitive to prediction errors during training, facilitating more effective adjustment of model parameters and reducing the likelihood of erroneous predictions. Therefore, the cross-entropy loss function is well-suited as a loss function for classification models, particularly excelling in handling multi-class classification problems.

## 4 Results and discussion

This study was conducted on a computer equipped with RTX3080 graphics card of 10 GB video memory and 64 GB of RAM.

### 4.1 Brain tumor dataset and preprocessing

In this paper, two publicly available brain tumor MRI datasets are applied for the brain tumor multi-classification task. Details of these two datasets are provided in Table 1. Both Cheng dataset and BT-large-4c dataset contain different views of brain anatomy: axial, coronal and sagittal views. Additionally, both datasets contain different numbers of brain tumor categories obtained from different patients with differences in tumor grade, race, and age. The Cheng dataset contains 3 types of brain tumors, namely glioma, meningioma and pituitary tumor. Among them, there are 1426 glioma images, 708 meningioma images and 930 pituitary tumor images, for a total of 3064 grayscale brain Magnetic Resonance (MR) images (48). The BT-large-4c dataset consists of 3264 brain MR images, including 926 glioma, 940 meningioma and 901 pituitary tumor images, and the remaining 497 normal images (49). These two datasets are split into 80% for training and 20% for testing.

During the dataset preprocessing phase, we implemented an efficient and streamlined data preprocessing protocol. To ensure

TABLE 1 Details of the datasets used in this study.

| NO. | Dataset name | Classes    | Number of Each class | Total number of images |
|-----|--------------|------------|----------------------|------------------------|
| 1   | Cheng        | Glioma     | 1426                 | 3064                   |
|     |              | Meningioma | 708                  |                        |
|     |              | Pituitary  | 930                  |                        |
| 2   | BT-large-4c  | Glioma     | 926                  | 3264                   |
|     |              | Meningioma | 940                  |                        |
|     |              | Pituitary  | 901                  |                        |
|     |              | No tumor   | 497                  |                        |

image content integrity and feature stability in experimental settings, all images were uniformly resized to dimensions of 224×224×3 pixels. This standardized resizing not only preserves the spatial structure and informational completeness of images but also significantly reduces computational overhead during network training, thereby enhancing training efficiency. Additionally, a standardization procedure was applied—a conventional preprocessing technique in deep learning—to mitigate variations in illumination, contrast, and other attributes across images, enabling the model to focus on learning intrinsic features. Considering that deep neural networks typically require large-scale datasets for training while our study employed a relatively limited dataset, data augmentation strategies were systematically deployed to alleviate overfitting. Specifically, techniques including random rotation, cropping, and horizontal flipping were implemented. These operations effectively enhanced dataset diversity without introducing additional noise, thereby strengthening the model’s generalization capabilities.

### 4.2 System implementation and evaluation metrics

During the model training process, we will fine-tune hyperparameters such as batch size, optimizer type, learning rate, epochs, and loss function based on experience and actual requirements. The objective of this process is to identify the optimal combination of hyperparameters to enhance the model’s performance and achieve the desired training outcomes. In this model, we employ the Adam optimizer with an initial learning rate of 0.001, 150 epochs, and a mini-batch size of 16 samples.

In this study, the performance of the proposed method is given by accuracy, recall, precision, and F1 -score (Cohen’s) were used for evaluation Kappa( $\kappa$ ), Matthews Correlation Coefficient (MCC) are given by this is given by Equations 10–15 (50):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{F1 - score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{14}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{15}$$

where, True Positives (TP) are the number of actual and predicted positives. True Negatives (TN) are the number of negatives that are both actual and predicted. False Positives (FP) are the number of actual negatives that are predicted to be positive. False Negatives (FN) are the number of actual positives that are predicted to be negative.  $p_o$  is the proportion of inter-observers who actually agree.  $p_e$  is the proportion of agreement expected based on a random assignment.

### 4.3 Experimental results

The proposed method is applied to the Cheng dataset and the BT-large-4c dataset for classification, and the corresponding confusion matrix is generated, as shown in Figures 4A, B. In these matrices, the label “G” represents glioma, “M” represents meningioma, “P” represents pituitary tumor, and “N” represents no tumor. The confusion matrices vividly illustrate the classification performance of the model for each category. Additionally, the detailed values of model metrics obtained on the Cheng and BT-large-4c datasets are shown in Table 2. These metrics offer a quantitative basis for comparison, facilitating the evaluation of the model’s performance and comparison with other methods. It is noteworthy that on the Cheng dataset, our model demonstrated

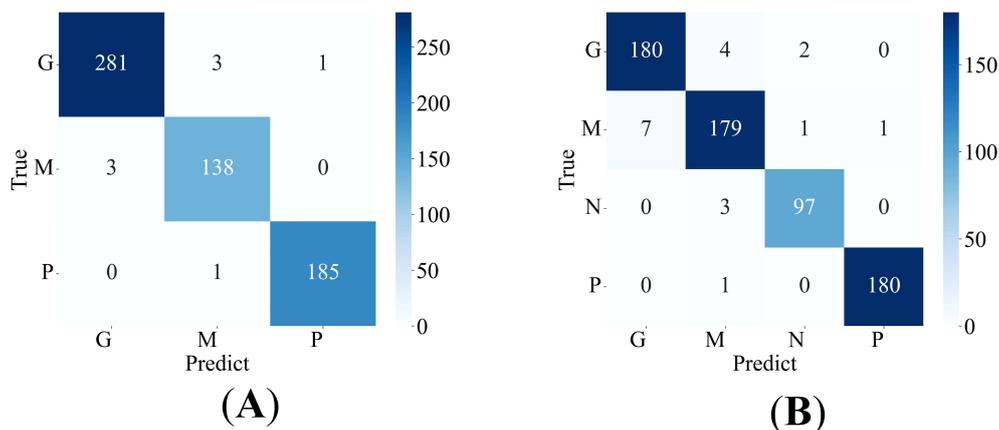


FIGURE 4 Confusion matrix of the proposed model (A) on the Cheng dataset, (B) on the BT-large-4c dataset.

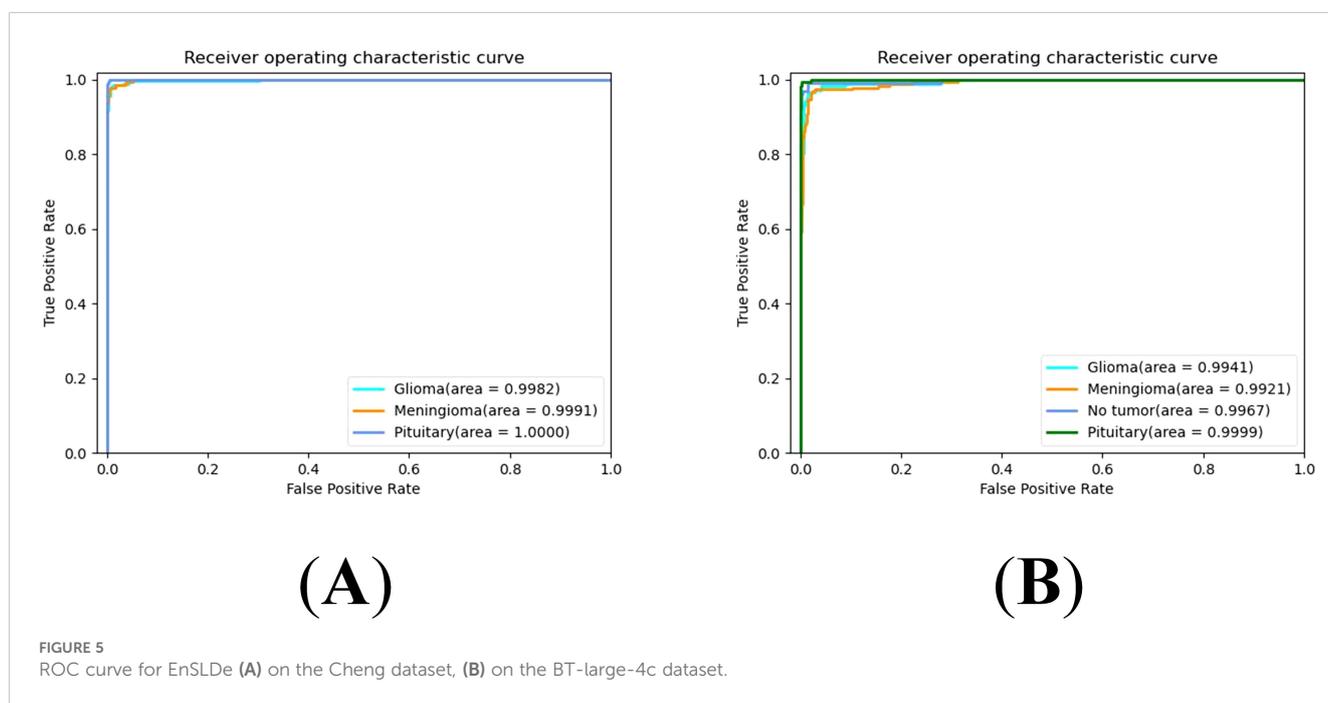
TABLE 2 Detailed metric values of the proposed model on Cheng and BT-large-4c datasets.

| Dataset     | Tumor type     | Precision     | Recall        | F1-score      | Accuracy | $\kappa$ | Mcc    |
|-------------|----------------|---------------|---------------|---------------|----------|----------|--------|
| Cheng       | Glioma         | 0.9894        | 0.9860        | 0.9877        | 0.9869   | 0.9795   | 0.9795 |
|             | Meningioma     | 0.9718        | 0.9787        | 0.9753        |          |          |        |
|             | Pituitary      | 0.9946        | 0.9946        | 0.9946        |          |          |        |
|             | <b>Average</b> | <b>0.9853</b> | <b>0.9864</b> | <b>0.9859</b> |          |          |        |
| BT-large-4c | Glioma         | 0.9626        | 0.9677        | 0.9651        | 0.9710   | 0.9607   | 0.9607 |
|             | Meningioma     | 0.9572        | 0.9521        | 0.9547        |          |          |        |
|             | No tumor       | 0.9700        | 0.9700        | 0.9700        |          |          |        |
|             | Pituitary      | 0.9945        | 0.9945        | 0.9945        |          |          |        |
|             | <b>Average</b> | <b>0.9711</b> | <b>0.9711</b> | <b>0.9711</b> |          |          |        |

exceptionally high classification performance, achieving an accuracy of 98.69%. Similarly, on the BT-large-4c dataset, the model achieved a classification accuracy of 97.10%. The total number of parameters in the EnSLDe model is 87 million (87M). The total memory size required for the model during operation (including training and inference) is 2792.73MB. The memory size required for one forward and backward propagation process in the model is 2459.25MB.

The Receiver Operating Characteristic (ROC) curve is a graphical tool used to represent the performance of a classification model. It effectively evaluates the performance of the model under different classification thresholds by taking the False Positive Rate (FPR) and True Positive Rate (TPR) as the horizontal and vertical coordinates.

The Area Under the Curve (AUC) quantitatively assesses the quality of the classification model. Higher AUC values indicate better model performance, with values closer to 1 indicating more ideal classification performance. Specifically, the ROC curves of our proposed model on the Cheng dataset and BT-large-4c dataset are depicted in Figures 5A, B, respectively. On the Cheng dataset, the AUC values for glioma, meningioma, and pituitary tumor in our proposed model are 0.9982, 0.9991, and 1.0000, respectively. On the BT-large-4c dataset, the AUC values for glioma, meningioma, pituitary tumor, and no tumor in our proposed model are 0.9941, 0.9921, 0.9999, and 0.9967, respectively. These results indicate that our proposed model exhibits excellent classification performance on both the Cheng dataset and BT-large-4c dataset.



### 4.4 Ablation experiment

This ablation experiment aims to comprehensively evaluate the impact of attention module, FEnM and data enhancement on model performance. The following three subsections will demonstrate in detail the contribution and importance of these three key components to model performance.

#### 4.4.1 The impact of the attention module on the model

In this section, the influence of various attention modules on our proposed model is investigated. The new models reconstructed from these attention modules and our proposed model include: Squeeze-and-Excitation(SE) (51) instead of EMA in EnSLDe named as EnSLDe-SE, Coordinate Attention (CA) (52) instead of EMA in EnSLDe named as EnSLDe-CA, Convolutional Block Attention Module (CBAM) (53) instead of EMA in EnSLDe named as EnSLDe-CBAM and the one removing EMA from EnSLDe named as EnSLDe-NoEMA. These models are used for classification prediction on the Cheng dataset, and the results are shown in Figure 6.

From Figure 6, it is evident that the EnSLDe-SE does not perform well in these models, with an accuracy of only 96.41%. Conversely, the EnSLDe exhibits exceptional performance in these models, achieving an accuracy of 98.69% and demonstrating excellent performance across other evaluation metrics.

Specifically, the EnSLDe attains 98.53%, 98.64%, and 98.59% in precision, recall, and F1-score parameters, respectively. Moreover, when the EMA module is removed, the model’s accuracy significantly drops to 97.06%. This comparison underscores the crucial role of the EMA module in enhancing the performance of the proposed model. The inclusion of the EMA module not only boosts the classification accuracy of the model but also achieves balanced optimization across multiple evaluation metrics, thereby enabling the model to maintain high performance levels.

#### 4.4.2 The impact of the FExM on the model

FExM is the cornerstone of the EnSLDe architecture, designed to hierarchically extract multi-scale contextual features through the combination of convolutional layers, residual connections, and the EMA mechanism. To rigorously evaluate its contribution, we conducted a comparative analysis of the model’s performance with and without the FExM module. When the FExM was not used, the model’s performance metrics—Precision, Recall, F1-score, and Accuracy—were 0.9656, 0.9722, 0.9683, and 0.9706, respectively, which were consistently lower than those of the model with FExM. It is worth noting that the precision dropped by 1.63%, highlighting the crucial importance of FExM to the overall model performance. Furthermore, in the ablation study, the p-value for the paired t-test of accuracy was 0.0013 (below the significance level,  $\alpha = 0.05$ ), with a confidence interval ranging from [0.0442, 0.1815].

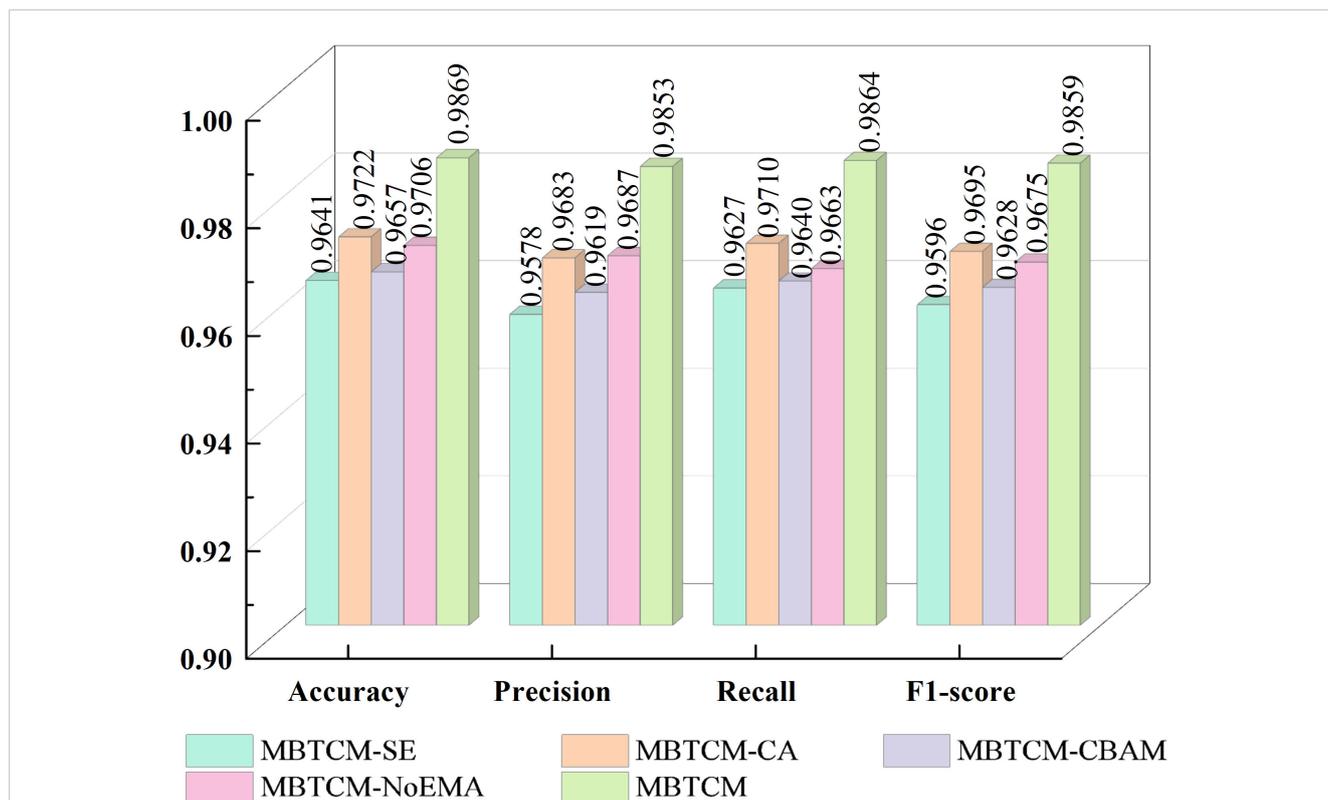
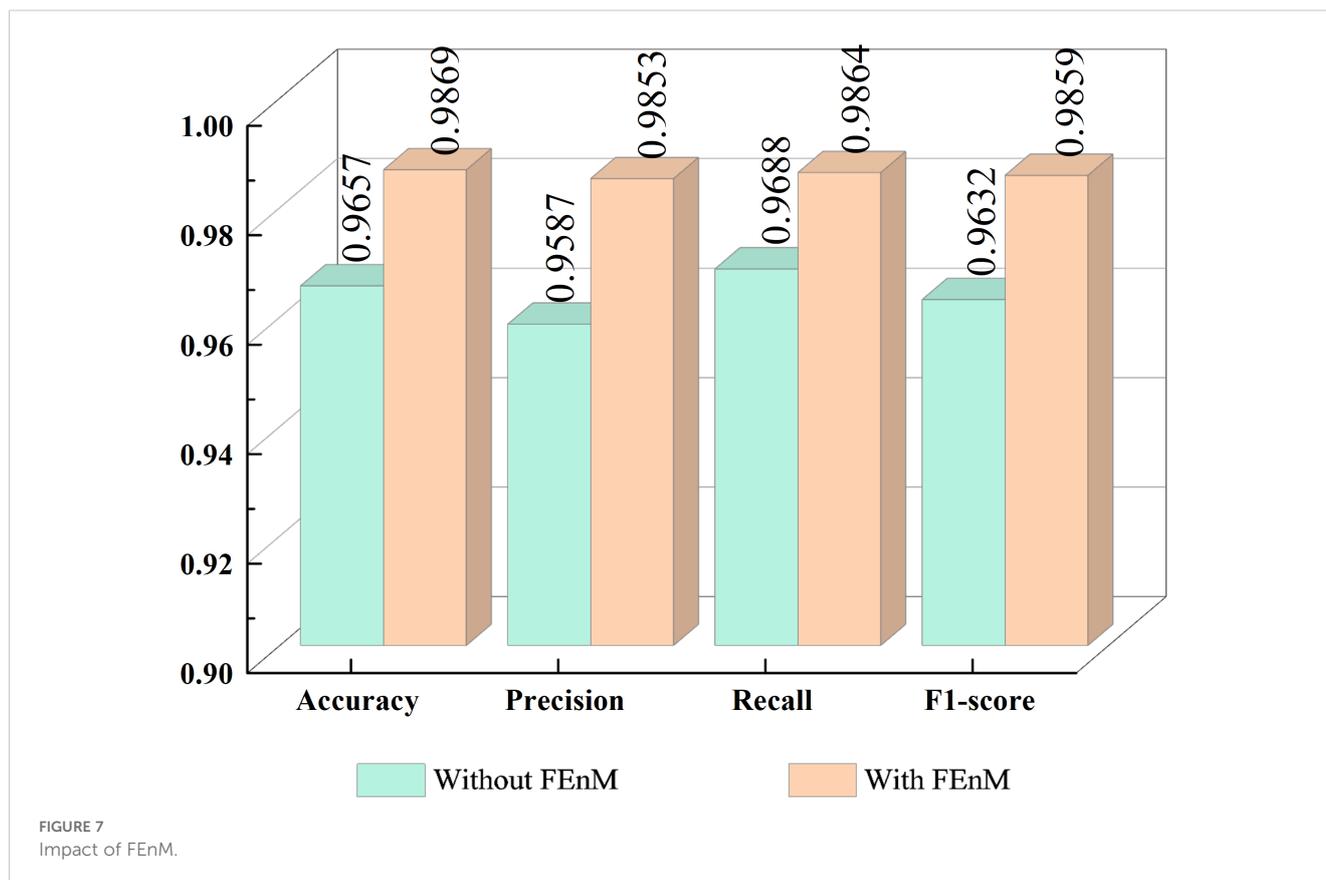


FIGURE 6 Impact of each attention module.



#### 4.4.3 The impact of the FEnM on the model

This section primarily examines the impact of the FEnM on the proposed model, with specific results depicted in Figure 7. The figure clearly illustrates that introducing FEnM significantly enhances the classification performance of the model on the Cheng dataset. Specifically, the accuracy, precision, recall, and F1-score of the model have increased by 2.12%, 2.66%, 1.76%, and 2.27%, respectively. The p-value of the paired t-test for accuracy with and without FEnM was 0.0094 (which is below the significance level,  $\alpha = 0.05$ ), and the confidence interval range was [0.0228, 0.1595]. The notable performance improvement can be attributed to the effective role of the FEnM. The FEnM not only substantially enhances the extracted features but also excels in capturing important long-range dependencies. Moreover, the FEnM can integrate the retained local key information with different levels of deep features, thereby enriching the expressive capabilities of features. Through this feature enhancement method, the model can more accurately identify brain tumors in classification tasks.

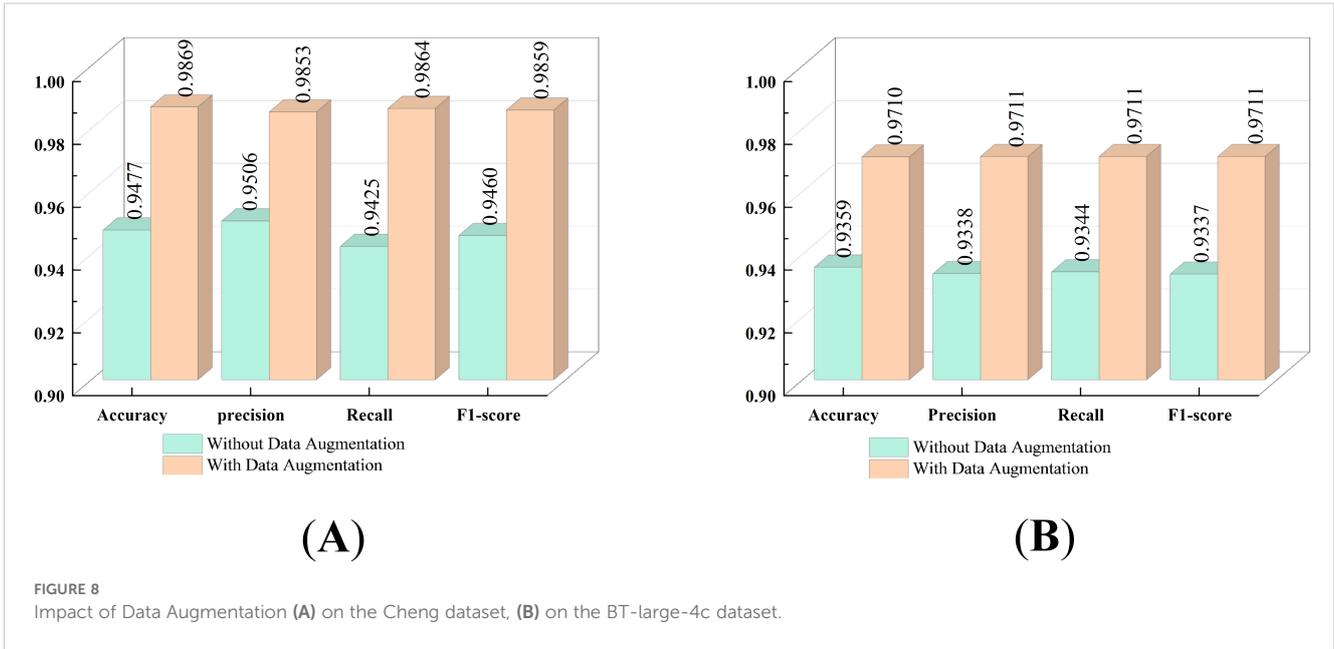
#### 4.4.4 The impact of data augmentation on models

This experiment utilizes two datasets: the Cheng dataset and the BT-large-4c dataset. Through the application of data augmentation techniques, the classification performance of the proposed model on these datasets is significantly enhanced. The impact of data augmentation on the model is illustrated in Figure 8. Specifically, for the Cheng dataset, the accuracy is improved by 3.92%, and for the

BT-large-4c dataset, the accuracy is improved by 3.51%. These results highlight the crucial role of data augmentation techniques in enhancing model performance. In particular, by incorporating data augmentation with random horizontal or vertical flipping of images, the model becomes adept at learning tumor characteristics from various orientations and locations. This implies that the model can effectively identify and classify tumors even when their orientation or location varies in real-world applications.

#### 4.4.5 Ablation studies on layer selection

To further validate the selection of feature extraction layers, we conducted an ablation study, the results of which are summarized in Table 3. When features were extracted from a single layer (shallow or deep), classification accuracy was consistently lower than that achieved via a multilayer fusion approach. To assess whether the observed differences in performance were statistically significant, paired t-tests were conducted. The tests compared classification accuracies of deep layers (which demonstrated superior performance to shallow layers) and multilayer fusion, positing the null hypothesis that there was no significant difference in performance. The paired t-test produced a p-value of 0.03 (below the significance level,  $\alpha = 0.05$ ), indicating a statistically significant difference in performance. By combining features from shallow and deep layers, the model captured a more holistic representation of the input data. The confidence interval for the difference in accuracy (which ranged from [-0.013, -0.0019]) excluded zero, confirming



that the multilayer fusion approach surpassed single-layer extraction. The shallow layer provided detailed local information, whereas the deep layer captured global contextual features. This combination enhanced the model’s ability to discern complex patterns in brain tumor images.

#### 4.4.6 Impact of hyperparameter selection on model performance

Hyperparameters are an important aspect that affects model performance, and different hyperparameters can lead to different experimental results. In this section, the impact of the hyperparameters batch size, lr, and optimizer on model performance will be verified. Table 4 presents the experimental results. By comparing Tables 2, 4, it can be found that the hyperparameter values selected in this paper are quite good.

#### 4.5 Cross-dataset validation

To comprehensively validate the model, cross-validation was employed. The BT-large-4c dataset, comprising glioma, pituitary tumor, and meningioma data, was used to evaluate the model trained on the Cheng dataset. The cross-validation results for accuracy, precision, recall, and F1-score were 92.98%, 93.2%, 93.02%, and 93.01%, respectively. These outcomes indicate that the proposed model exhibits significant robustness.

#### 4.6 Discussion

To further quantify the performance of the proposed model. The classification results obtained by our proposed model are compared with those obtained by previous state-of-the-art models using the same dataset, as shown in Table 5. Noreen et al. (54) proposed a method integrating deep learning with machine learning models, employing deep learning for feature extraction, including the Inception-v3 and Xception models. Additionally, the classification of brain tumors through deep learning and machine learning algorithms such as softmax, RF, SVM, KNN, and ensemble techniques were explored. Bodapati et al. (55) developed a dual-channel deep neural network architecture for brain tumor classification using pre-trained InceptionResNetV2 and Xception models, incorporating attention mechanisms to enhance accuracy and generalization capabilities in brain tumor recognition. Shaik and Cherukuri (56) designed and implemented a multi-level attention network (MANet). The proposed MANet includes spatial and channel-wise attention mechanisms, prioritizing tumor regions while maintaining the inter-channel temporal dependencies in the semantic feature sequences obtained from the abnormal areas. Öksüz et al. (57) utilized pre-trained AlexNet, ResNet-18, GoogLeNet, and ShuffleNet networks to extract deep features from images, and designed a shallow network for extracting shallow features, fusing these features and classifying them with SVM and KNN. Jaspin and Selvan (58) proposed a multi-class

TABLE 3 Layer selection of experimental results in dataset Chen.

| Method                   | Precision | Recall | F1-score | Accuracy |
|--------------------------|-----------|--------|----------|----------|
| Shallow layer            | 0.9554    | 0.9539 | 0.9546   | 0.9592   |
| Deep layer               | 0.9683    | 0.9645 | 0.9664   | 0.9690   |
| Multilayer fusion (ours) | 0.9853    | 0.9864 | 0.9859   | 0.9869   |

TABLE 4 Experimental results for different hyper-paramete.

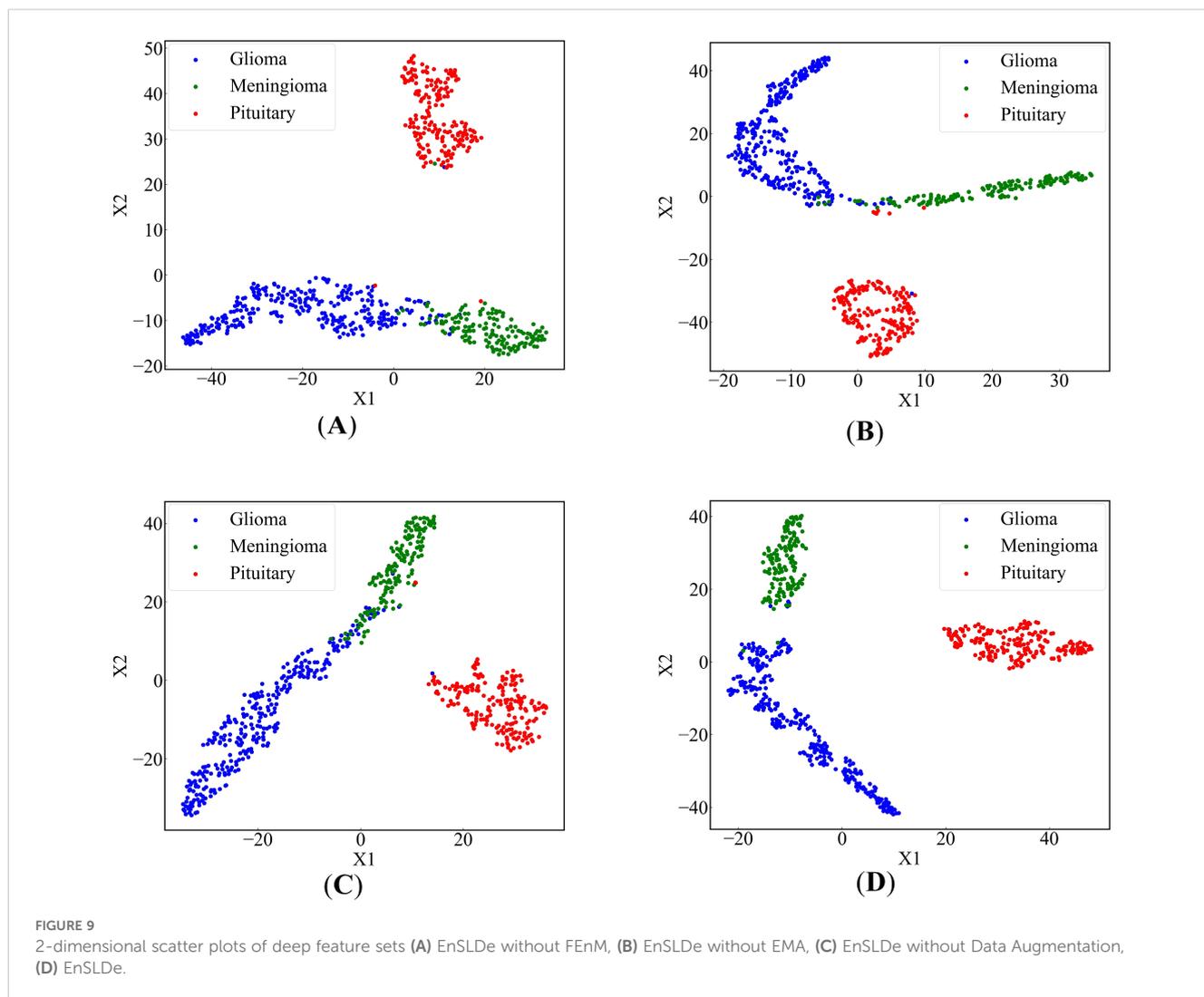
| Hyper-parameter | Value  | Precision | Recall | F1-score | Accuracy |
|-----------------|--------|-----------|--------|----------|----------|
| Batch           | 8      | 0.9739    | 0.9770 | 0.9754   | 0.9771   |
| Lr              | 0.0001 | 0.9835    | 0.9811 | 0.9823   | 0.9837   |
| Optimizer       | SGD    | 0.9800    | 0.9757 | 0.9778   | 0.9804   |

convolutional neural network (MCCNN) model for identifying tumors in brain MRI images. This network, consisting of an 11-layer structure including three convolutional layers, three max-pooling layers, one flattening layer followed by three dense layers, and an output layer, achieved classification performance on par with pre-trained models. Md. S. I. Khan et al. (59) designed a 23-layer convolutional neural network for brain tumor classification. Satyanarayana et al. (60) introduced a density convolutional neural network model based on mass correlation mapping (DCNN-MCM) for brain tumor classification. This model leverages the average mass elimination algorithm (AMEA) and mass correlation analysis (MCA) for the extraction and training of significant features of brain tumors, using a CNN model for efficient classification. Kibriya et al. (61) developed a 13-layer CNN specifically for brain tumor classification. Dutta et al. (62) introduced an attention-based residual multi-scale CNN, termed ARM-Net. This model includes a lightweight residual multi-scale CNN architecture known as RM-Net and introduces a lightweight global attention module (LGAM) to selectively learn more discriminative features. S. U. R. Khan et al. (63) employed the DenseNet169 model for feature extraction and fed the extracted features into three multi-class machine learning classifiers: RF, SVM, and gradient-boosting decision trees

(XGBoost). Brain tumor classification was performed through the integration of these classifiers using a majority voting strategy. Demir and Akbulut (64) used a new multi-level feature selection algorithm to select the 100 deep features with the highest significance and adopted the SVM algorithm with Gaussian kernel for classification and achieved better performance. Senan et al. (65) employed both AlexNet and ResNet18 in conjunction with SVM for brain tumor classification and diagnosis. Initially, deep learning techniques were used to extract robust and significant deep features through deep convolutional layers, followed by classification using SVM. Ravinder et al. (66) proposed a graph convolutional neural network (GCN) model. This model integrates graph neural networks (GNN) with traditional CNNs. Our EnSLDe achieves superior performance compared to other methods. This depends on its ability to enhance short-range and long-range dependencies. EnSLDe yields experimental results for the Chen dataset. On the BT-large-4c dataset, EnSLDe underperforms AlexNet+SVM by a margin of 0.0139 in terms of precision. Nonetheless, it excels in other performance indicators. The EnSLDe model demonstrates exceptional performance on the Cheng and BT-large-4c datasets, achieving high accuracy rates of 98.69% and 97.10%, respectively. These results highlight the

TABLE 5 Comparison of our proposed model with previous models.

| Reference                  | Dataset     | Method                  | Precision | Recall | F1-score | Accuracy |
|----------------------------|-------------|-------------------------|-----------|--------|----------|----------|
| Noreen et al. (54)         | Cheng       | Inception-v3+Ensemble   | –         | –      | –        | 0.9434   |
| Bodapati et al. (55)       |             | Two-Channel DNN         | –         | –      | 0.9779   | 0.9523   |
| Shaik and Cherukuri (56)   |             | MANet                   | 0.9614    | 0.9599 | 0.9603   | 0.9651   |
| Öksüz et al. (57)          |             | ResNet18+ShallowNet+SVM | 0.9525    | 0.9527 | 0.9526   | 0.9725   |
| Jaspin and Selvan (58)     |             | MCCNN                   | 0.95      | 0.95   | 0.96     | 0.9517   |
| Md. S. I. Khan et al. (59) |             | 23-layer CNN            | 0.965     | 0.964  | 0.964    | 0.978    |
| Satyanarayana et al. (60)  |             | DCNN-MCN                | –         | –      | –        | 0.94     |
| Kibriya et al. (61)        |             | 13-layer CNN            | 0.97      | 0.96   | 0.965    | 0.972    |
| Dutta et al. (62)          |             | ARM-Net                 | 0.9646    | 0.9609 | 0.9620   | 0.9664   |
| S. U. R. Khan et al. (63)  |             | Hybrid-NET              | 0.95      | 0.94   | 0.94     | 0.951    |
| Dutta et al. (44)          |             | GT-Net                  | –         | –      | 96.39    | 97.11    |
| The Proposed Method        |             | EnSLDe                  | 0.9853    | 0.9864 | 0.9859   | 0.9869   |
| Demir and Akbulut (64)     |             | BT-large-4c             | R-CNN+SVM | 0.964  | 0.9645   | 0.964    |
| Senan et al. (65)          | AlexNet+SVM |                         | 0.985     | –      | –        | 0.951    |
| Ravinder et al. (66)       | GCNN        |                         | 0.9525    | 0.965  | 0.9587   | 0.9501   |
| The Proposed Method        | EnSLDe      |                         | 0.9711    | 0.9711 | 0.9711   | 0.971    |



model's ability to effectively capture both short-range and long-range dependencies in brain tumor images, leading to improved classification accuracy. And multi-scale parallel subnetworks fuse shallow and deep features to capture comprehensive information. However, it is important to note that the performance of any model, including EnSLDe, can vary depending on the specific characteristics of the data it is applied to. While EnSLDe outperforms several state-of-the-art models on these datasets, its generalizability to real-world applications requires further validation.

In order to more intuitively display the effect of our proposed method, we used the t-SNE (67) algorithm to reduce the dimensionality of high-dimensional feature data and drew a scatter plot on a 2-dimensional plane. Figures 9A–C depict scatter plots obtained by removing FEnM, EMA, and Data Augmentation, respectively. There are instances where the glioma class and the meningioma class are interconnected and nested. However, in Figure 9D, obtained by EnSLDe, the sample points of each class are closely clustered together, with clear separation between different categories. This intuitively underscores the significance

of FEnM, EMA, and Data Augmentation for the model. The ability of the model to distinguish features effectively is enhanced by them.

As shown in Figure 4 and Table 2, the EnSLDe model achieves superior classification performance for pituitary tumors (precision: 0.9946, recall: 0.9946) compared to gliomas (precision: 0.9894, recall: 0.9946) and meningiomas (precision: 0.9718, recall: 0.9787), the latter of which exhibits the lowest performance metrics. A comparison of Figures 9A–D illustrates that EnSLDe employs effective strategies to differentiate gliomas from meningiomas. However, persistent feature overlap hinders the model's ability to achieve optimal classification accuracy.

The EnSLDe model is designed to capture both short- and long-range dependencies within images, demonstrating considerable potential for generalization beyond the classification of brain tumors. Its architecture, which incorporates a multi-scale parallel subnetwork and feature enhancement modules, is well-suited for a wide range of medical imaging tasks. Additionally, the model is adaptable to the classification of tumors in various organs, such as lung, breast, and liver tumors. The model's ability to effectively capture contextual information makes it suitable for the

identification of different lesion types and the detection of abnormalities across a diverse array of medical conditions.

Adapting the EnSLDe model to a new task necessitates several adjustments. First, the model requires retraining on a task-specific dataset, including modifying the number of output categories and fine-tuning the classification module. Furthermore, the feature extraction module may require modification to account for variations in imaging characteristics, such as resolution and contrast. Despite its design efficiency, the EnSLDe model exhibits limited scalability, particularly in resource-constrained environments. Training the model demands substantial computational resources, particularly for large-scale datasets. However, incorporating efficient convolutional layers and depthwise separable convolutions mitigates these computational demands. To address scalability challenges, several strategies may be implemented. For instance, model compression techniques (e.g., pruning and quantization) can substantially reduce computational complexity while maintaining competitive performance.

To further understand the decision-making process of the proposed EnSLDe model and validate its ability to focus on relevant regions in brain tumor classification, we visualized the feature maps using the Grad-CAM++ method. The results are shown in Figure 10. Grad-CAM++ is a widely used technique for visualizing the regions of interest in image classification tasks, providing insights into the model's attention mechanism. As shown in Figure 10, the feature maps generated by the EnSLDe model effectively highlight brain tumor regions, demonstrating the model's ability to distinguish between brain tumor and non-tumor regions. This visualization confirms that the model focuses on tumor regions, which is critical for accurate classification. However, it is also clear that the model focused on other non-tumor regions. This observation suggests that the model effectively captures key brain tumor features while incorporating additional contextual information from surrounding brain regions, which may

contribute to its high classification accuracy. While the EnSLDe model demonstrated strong performance in focusing on relevant regions, the visualization results also highlighted areas for potential improvement. Specifically, the model's focus on non-tumor regions suggests that there may be opportunities to refine the feature extraction and enhancement modules to emphasize the most critical features further. Future work could explore advanced attention mechanisms or additional regularization techniques to ensure that the model focuses more precisely on tumor regions, potentially leading to higher classification accuracy.

## 5 Conclusion

A new multi-class brain tumor classification model, named EnSLDe, has been proposed. This model is primarily composed of three modules: FExM (Feature Extraction Module), FEnM (Feature Enhancement Module), and the classification module. FExM efficiently extracts features using convolutional layers and residual networks and combines EMA (Efficient Multi-Attention) to simultaneously focus on both channel and spatial information of the features. This effectively preserves the information of each channel, preventing the loss of important features during the compression of the channel dimension. The design of FEnM aims to deeply integrate shallow and deep features, facilitating a more comprehensive understanding of the features and the extraction of advanced and important features. Additionally, the model's ability to capture short-range and long-range dependencies has been enhanced. The feature enhancement module further strengthens the features by effectively capturing important dependencies over a large sequence range while preserving local key information. The double-layer fully connected structure is adopted as the core of the classification module and combined with dropout regularization technology, which further improves the model classification

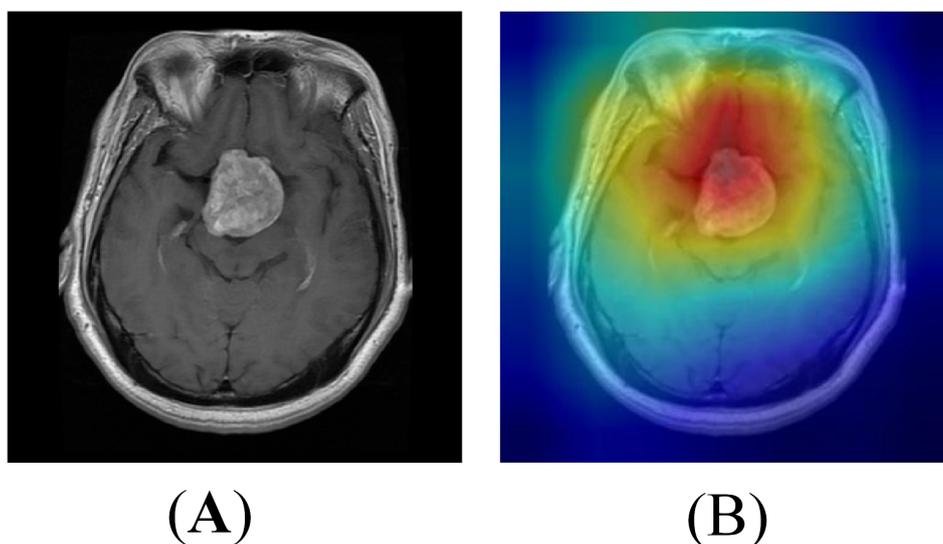


FIGURE 10  
Heat map visualization of the model (A) Original image (B) Heat map.

performance. Experimental evaluations conducted on the challenging Cheng dataset and BT-large-4c dataset demonstrate the excellent performance of our model in brain tumor classification tasks. On the Cheng dataset, the model achieves accuracy, recall, precision, and F1-score of 98.69%, 98.53%, 98.64%, and 98.59%, respectively. Similarly, on the BT-large-4c dataset, the model attains accuracy, recall, precision, and F1-score of 97.10%, 97.11%, 97.11%, and 97.11%, respectively. Indeed, the differentiation between glioma and meningioma remains suboptimal. Further refinement is required to enhance the model's ability to distinguish accurately between these two tumor types. Future studies should augment the dataset to include a broader range of brain disorders, thereby enriching the model's training corpus and enhancing its capacity to differentiate among diverse neurological pathologies. Additionally, strategic modifications to the model's architecture, training protocols, and loss functions could be implemented to optimize its discriminative performance in distinguishing gliomas from meningiomas. And the model was deployed, and the clinical capabilities of the model were verified by combining the doctors commanded by experience.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: [https://figshare.com/articles/dataset/brain\\_tumor\\_dataset/1512427](https://figshare.com/articles/dataset/brain_tumor_dataset/1512427) <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>.

## Author contributions

WC: Conceptualization, Investigation, Project administration, Writing – original draft. JL: Formal Analysis, Software, Validation, Visualization, Writing – review & editing. XT: Formal Analysis, Software, Writing – original draft. JZ: Conceptualization, Project administration, Software, Writing – review & editing.

## References

- Asif S, Yi W, Ain QU, Hou J, Yi T, Si J. Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from MR images. *IEEE Access*. (2022) 10:34716–30. doi: 10.1109/ACCESS.2022.3153306
- Bouhafra S, El Bahi H. Deep learning approaches for brain tumor detection and classification using MRI images, (2020 to 2024): A systematic review. *J Of Imaging Inf In Med*. (2024). doi: 10.1007/s10278-024-01283-8
- Ghosal P, Reddy S, Sai C, Pandey V, Chakraborty J, Nandi D. A deep adaptive convolutional network for brain tumor segmentation from multimodal MR images. *TENCON 2019 - 2019 IEEE Region 10 Conf (TENCON)*. (2019), 1065–70. doi: 10.1109/TENCON.2019.8929402
- Yu Z, Li X, Li J, Chen W, Tang Z, Geng D. HSA-net with a novel CAD pipeline boosts both clinical brain tumor MR image classification and segmentation. *Comput Biol Med*. (2024) 170:108039. doi: 10.1016/j.compbiomed.2024.108039
- Almalki YE, Ali MU, Kallu KD, Masud M, Zafar A, Alduraibi SK, et al. Isolated convolutional-neural-network-based deep-feature extraction for brain tumor classification using shallow classifier. *Diagnostics*. (2022) 12:1793. doi: 10.3390/diagnostics12081793
- Mehnatkesh H, Jalali SMJ, Khosravi A, Nahavandi S. An intelligent driven deep residual learning framework for brain tumor classification using MRI images. *Expert Syst Appl*. (2023) 213:119087. doi: 10.1016/j.eswa.2022.119087
- Mondal A, Shrivastava VK. A novel Parametric Flatten-p Mish activation function based deep CNN model for brain tumor classification. *Comput Biol Med*. (2022) 150:106183. doi: 10.1016/j.compbiomed.2022.106183
- ZainEldin H, Gamel SA, El-Kenawy E-SM, Alharbi AH, Khafaga DS, Ibrahim A, et al. Brain tumor detection and classification using deep learning and sine-cosine fitness grey wolf optimization. *Bioengineering*. (2022) 10:18. doi: 10.3390/bioengineering10010018
- Louis DN, Perry A, Reifenberger G, Von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathologica*. (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1
- Amin J, Sharif M, Raza M, Saba T, Rehman A. Brain tumor classification: feature fusion. In: *2019 International Conference on Computer and Information Sciences*

GD: Project administration, Supervision, Writing – review & editing. QF: Validation, Writing – review & editing. HJ: Project administration, Validation, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Foundation of He'nan Educational Committee (Grant No. 24A320004), Major Science and Technology Projects of Henan Province (Grant No. 221100210500), the Medical and Health Research Project in Luoyang (Grant No. 2001027A), and the Construction Project of Improving Medical Service Capacity of Provincial Medical Institutions in Henan Province (Grant No. 2017-51).

## 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.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- (ICCCIS) (2019) (Sakaka, Saudi Arabia: Institute of Electrical and Electronics Engineers (IEEE)), p. 1–6. doi: 10.1109/ICCCIS.2019.8716449
11. Njeh I, Sallemi L, Ben Slima M, Ben Hamida A, Lehericy S, Galanaud D. A Computer Aided Diagnosis 'CAD' for Brain Glioma Exploration. In: *2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (2014) (Soussse: Institute of Electrical and Electronics Engineers (IEEE)). p. 243–8. doi: 10.1109/ATSIP.2014.6834615
  12. Singh RS, Saini BS, Sunkaria RK. Classification of cardiac heart disease using reduced chaos features and 1-norm linear programming extreme learning machine. *Int J For Multiscale Comput Eng.* (2018) 16:465–86. doi: 10.1615/IntJMultCompEng.2018026587
  13. Shahid AH, Singh MP, Roy B, Aadarsh A. Coronary artery disease diagnosis using feature selection based hybrid extreme learning machine. In: *2020 3rd International Conference On Information And Computer Technologies (Icict 2020)* (2020) (San Jose, CA, USA: Institute of Electrical and Electronics Engineers (IEEE)). p. 341–6. doi: 10.1109/ICICT50521.2020.00060
  14. Xie W, Li Y, Ma Y. Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing.* (2016) 173:930–41. doi: 10.1016/j.neucom.2015.08.048
  15. Heidari M, Lakshmiarahan S, Mirniaharikandehei S, Danala G, Maryada SKR, Liu H, et al. Applying a random projection algorithm to optimize machine learning model for breast lesion classification. *IEEE Trans On Biomed Eng.* (2021) 68:2764–75. doi: 10.1109/TBME.2021.3054248
  16. Pyrkov TV, Slipensky K, Barg M, Kondrashin A, Zhurov B, Zenin A, et al. Extracting biological age from biomedical data via deep learning: Too much of a good thing? *Sci Rep.* (2018) 8:5210. doi: 10.1038/s41598-018-23534-9
  17. Sarki R, Ahmed K, Wang H, Zhang Y. Automated detection of mild and multi-class diabetic eye diseases using deep learning. *Health Inf Sci And Syst.* (2020) 8:32. doi: 10.1007/s13755-020-00125-5
  18. Jeong W-Y, Kim J-H, Park J-E, Kim M-J, Jong-Min L. Evaluation of classification performance of inception V3 algorithm for chest X-ray images of patients with cardiomegaly. *한국방사선학회논문지.* (2021) 15:455–61. doi: 10.7742/jksr.2021.15.4.455
  19. Chowdhury MS, Sultan T, Jahan N, Mridha MF, Safran M, Alfarhood S, et al. Leveraging deep neural networks to uncover unprecedented levels of precision in the diagnosis of hair and scalp disorders. *Skin Res And Technol.* (2024) 30:e13660. doi: 10.1111/srt.13660
  20. Sharifrazi D, Alizadehsani R, Joloudari JH, Band SS, Hussain S, Sani ZA, et al. CNN-KCL: Automatic myocarditis diagnosis using convolutional neural network combined with k-means clustering. *Math Biosci And Eng.* (2022) 19:2381–402. doi: 10.3934/mbe.2022110
  21. Shahin AI, Aly W, Aly S. MBTFNCN: A novel modular fully convolutional network for MRI brain tumor multi-classification. *Expert Syst Appl.* (2023) 212:118776. doi: 10.1016/j.eswa.2022.118776
  22. Xue Z, Chen W, Li J. Enhancement and fusion of multi-scale feature maps for small object detection. In: *2020 39th Chinese Control Conference (CCC)* (2020) (Shenyang, China: Institute of Electrical and Electronics Engineers (IEEE)). p. 7212–7. doi: 10.23919/CCC50068.2020.9189352
  23. Chen Y, Zhang C, Chen B, Huang Y, Sun Y, Wang C, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases. *Comput Biol Med.* (2024) 170:107917. doi: 10.1016/j.combiomed.2024.107917
  24. Lin T-Y, Dollar P, Girshick R, He K, Hariharan B, Belongie S. (2017). Feature pyramid networks for object detection, in: *The conference title is Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2117–25.
  25. Bansal T, Jindal N. An improved hybrid classification of brain tumor MRI images based on conglomeration feature extraction techniques. *Neural Computing Appl.* (2022) 34:9069–86. doi: 10.1007/s00521-022-06929-8
  26. Khan MA, Lali IU, Rehman A, Ishaq M, Sharif M, Saba T, et al. Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. *Microscopy Res Technique.* (2019) 82:909–22. doi: 10.1002/jemt.23238
  27. Mehmood A, Yang S, Feng Z, Wang M, Ahmad AS, Khan R, et al. A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. *Neuroscience.* (2021) 460:43–52. doi: 10.1016/j.neuroscience.2021.01.002
  28. Raza A, Ayub H, Khan JA, Ahmad I, S. Salama A, Daradkeh YI, et al. A hybrid deep learning-based approach for brain tumor classification. *Electronics.* (2022) 11:1146. doi: 10.3390/electronics11071146
  29. Diaz-Pernas FJ, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D. A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. *Healthcare.* (2021) 9:153. doi: 10.3390/healthcare9020153
  30. Ayadi W, Elhamzi W, Charfi I, Atri M. Deep CNN for brain tumor classification. *Neural Process Lett.* (2021) 53:671–700. doi: 10.1007/s11063-020-10398-2
  31. Sreenivasa Reddy B, Sathish A. A Multiscale Atrous Convolution-based Adaptive ResUNet3+ with Attention-based ensemble convolution networks for brain tumour segmentation and classification using heuristic improvement. *Biomed Signal Process Control.* (2024) 91:105900. doi: 10.1016/j.bspc.2023.105900
  32. Ghosal P, Nandanwar L, Kanchan S, Bhadra A, Chakraborty J, Nandi D. Brain tumor classification using resNet-101 based squeeze and excitation deep neural network. In: *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)* (2019) (Gangtok, India: Institute of Electrical and Electronics Engineers (IEEE)). p. 1–6. doi: 10.1109/ICACCP.2019.8882973
  33. Islam M, Talukder M, Uddin M, Akhter A, Khalid M. BrainNet: precision brain tumor classification with optimized efficientNet architecture. *Int J Of Intelligent Syst.* (2024) 2024. doi: 10.1155/2024/3583612
  34. Aurna NF, Abu Yousuf M, Abu Taher K, Azad AKM, Moni MA. A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Comput In Biol And Med.* (2022) 146. doi: 10.1016/j.compbiomed.2022.105539
  35. Musallam AS, Sherif AS, Hussein MK. A new convolutional neural network architecture for automatic detection of brain tumors in magnetic resonance imaging images. *IEEE Access.* (2022) 10:2775–82. doi: 10.1109/ACCESS.2022.3140289
  36. Kumar SA, Sasikala S. Automated brain tumour detection and classification using deep features and Bayesian optimised classifiers. *Curr Med Imaging.* (2023) 20. doi: 10.2174/1573405620666230328092218
  37. Alshayehi M, Al-Buloushi J, Ashkanani A, Abed S. Enhanced brain tumor classification using an optimized multi-layered convolutional neural network architecture. *Multimedia Tools Appl.* (2021) 80:28897–917. doi: 10.1007/s11042-021-10927-8
  38. Irmak E. Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework. *Iranian J Sci Technology Trans Electrical Eng.* (2021) 45:1015–36. doi: 10.1007/s40998-021-00426-9
  39. Rammurthy D, Mahesh PK. Whale Harris hawks optimization based deep learning classifier for brain tumor detection using MRI images. *J Of King Saud University-Computer And Inf Sci.* (2022) 34:3259–72. doi: 10.1016/j.jksuci.2020.08.006
  40. Alyami J, Rehman A, Almutairi F, Fayyaz AM, Roy S, Saba T, et al. Tumor localization and classification from MRI of brain using deep convolution neural network and salp swarm algorithm. *Cogn Comput.* (2023) 16:2036–46. doi: 10.1007/s12559-022-10096-2
  41. Tummala S, Kadry S, Bukhari SAC, Rauf HT. Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling. *Curr Oncol.* (2022) 29:7498–511. doi: 10.3390/currenconcol29100590
  42. Ferdous GJ, Sathi KA, Md. A, Hoque MM, Dewan MAA. LCDEiT: A linear complexity data-efficient image transformer for MRI brain tumor classification. *IEEE Access.* (2023) 11:20337–50. doi: 10.1109/ACCESS.2023.3244228
  43. Asiri AA, Shaf A, Ali T, Pasha MA, Khan A, Irfan M, et al. Advancing brain tumor detection: Harnessing the Swin Transformer's power for accurate classification and performance analysis. *Peerj Comput Sci.* (2024) 10. doi: 10.7717/peerj-cs.1867
  44. Dutta TK, Nayak DR, Pachori RB. GT-Net: Global transformer network for multi-class brain tumor classification using MR images. *Biomed Eng Lett.* (2024) 14:1069–77. doi: 10.1007/s13534-024-00393-0
  45. Ouyang D, He S, Zhang G, Luo M, Guo H, Zhan J, et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* (2023), 1–5. doi: 10.1109/ICASSP49357.2023.10096516
  46. Quan Y, Zhang D, Zhang L, Tang J. Centralized feature pyramid for object detection. *IEEE Trans Image Process.* (2023) 32:4341–54. doi: 10.1109/TIP.2023.3297408
  47. Zhang Z, Sabuncu M. Generalized cross entropy loss for training deep neural networks with noisy labels. *Adv Neural Inf Process Syst.* (2018) 31.
  48. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, et al. Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS One.* (2015) 10:e0140381. doi: 10.1371/journal.pone.0140381
  49. Bhuvaji S, Kadam A, Bhumkar P, Dedge S, Kanchan S. Brain tumor classification (MRI). *Kaggle.* (2020) 10.
  50. Alsagoff W, Cömert Z, Nour M, Polat K, Brdesee H, Toğaçar M. Predicting fetal hypoxia using common spatial pattern and machine learning from cardiocytography signals. *Appl Acoustics.* (2020) 167:107429. doi: 10.1016/j.apacoust.2020.107429
  51. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* (2018) (Salt Lake City, USA: Institute of Electrical and Electronics Engineers (IEEE)). pp. 7132–41. pp. 7132–41.
  52. Hou Q, Zhou D, Feng J. Coordinate Attention for Efficient Mobile Network Design. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* (2021) (Nashville, TN, USA: Institute of Electrical and Electronics Engineers (IEEE)). pp. 13713–22. pp. 13713–22.
  53. Woo S, Park J, Lee J-Y, Kweon IS. CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV).* (2018) (Munich, Germany: Springer). pp. 3–19.
  54. Noreen N, Palaniappan S, Qayyum A, Ahmad I, O. Alassafi M. Brain tumor classification based on fine-tuned models and the ensemble method. *Computers Materials Continua.* (2021) 67:3967–82. doi: 10.32604/cmc.2021.014158
  55. Bodapati JD, Shaik NS, Naralasetti V, Mundukur NB. Joint training of two-channel deep neural network for brain tumor classification. *Signal Image Video Process.* (2021) 15:753–60. doi: 10.1007/s11760-020-01793-2

56. Shaik NS, Cherukuri TK. Multi-level attention network: Application to brain tumor classification. *Signal Image Video Process.* (2022) 16:817–24. doi: 10.1007/s11760-021-02022-0
57. Öksüz C, Urhan O, Güllü MK. Brain tumor classification using the fused features extracted from expanded tumor region. *Biomed Signal Process Control.* (2022) 72:103356. doi: 10.1016/j.bspc.2021.103356
58. Jaspin K, Selvan S. Multiclass convolutional neural network based classification for the diagnosis of brain MRI images. *Biomed Signal Process Control.* (2023) 82:104542. doi: 10.1016/j.bspc.2022.104542
59. Khan Md.SI, Rahman A, Debnath T, Karim M, Nasir MK, Band SS, et al. Accurate brain tumor detection using deep convolutional neural network. *Comput Struct Biotechnol J.* (2022) 20:4733–45. doi: 10.1016/j.csbj.2022.08.039
60. Satyanarayana G, Appala Naidu P, Subbaiah Desanamukula V, Satish kumar K, Chinna Rao B. A mass correlation based deep learning approach using deep Convolutional neural network to classify the brain tumor. *Biomed Signal Process Control.* (2023) 81:104395. doi: 10.1016/j.bspc.2022.104395
61. Kibriya H, Masood M, Nawaz M, Nazir T. Multiclass classification of brain tumors using a novel CNN architecture. *Multimedia Tools Appl.* (2022) 81:29847–63. doi: 10.1007/s11042-022-12977-y
62. Dutta TK, Nayak DR, Zhang Y-D. ARM-Net: Attention-guided residual multiscale CNN for multiclass brain tumor classification using MR images. *Biomed Signal Process Control.* (2024) 87:105421. doi: 10.1016/j.bspc.2023.105421
63. Khan SUR, Zhao M, Asif S, Chen X. Hybrid-NET: A fusion of DenseNet169 and advanced machine learning classifiers for enhanced brain tumor diagnosis. *Int J Imaging Syst Technol.* (2024) 34:e22975. doi: 10.1002/ima.22975
64. Demir F, Akbulut Y. A new deep technique using R-CNN model and L1NSR feature selection for brain MRI classification. *Biomed Signal Process Control.* (2022) 75:103625. doi: 10.1016/j.bspc.2022.103625
65. Senan EM, Jadhav ME, Rassem TH, Aljaloud AS, Mohammed BA, Al-Mekhlafi ZG. Early diagnosis of brain tumour MRI images using hybrid techniques between deep and machine learning. *Comput Math Methods Med.* (2022) 2022:1–17. doi: 10.1155/2022/8330833
66. Ravinder M, Saluja G, Allabun S, Alqahtani MS, Abbas M, Othman M, et al. Enhanced brain tumor classification using graph convolutional neural network architecture. *Sci Rep.* (2023) 13:14938. doi: 10.1038/s41598-023-41407-8
67. Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res.* (2008) 9:2579–605.