Sensors and Materials, Vol. 38, No. 3 (2026) 1429–1445
MYU Tokyo

S & M 4386

Convolutional Autoencoder Network with Masked Contrast Enhancement for Brain Tumor Magnetic Resonance Sensor Image Recognition

Chih-Ta Yen[1*] and Hao-Siang Kao[2]

[1]Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei 10607, Taiwan
[2]Department of Electrical Engineering, National Taiwan Ocean University, Keelung City 202301, Taiwan

(Received September 24, 2025; accepted March 3, 2026)

Keywords: autoencoder network, brain tumor, deep learning, MRI, masking technology, unsupervised learning

Brain tumors vary in size and location in magnetic resonance imaging (MRI), and rising patient volumes at imaging centers delay radiologist feedback owing to increased diagnostic workload. To address this issue, we established an unsupervised learning model with a convolutional autoencoder for the extraction of features in images and explored the application of masking technology within this approach. The proposed method classifies tumors as gliomas, meningiomas, or pituitary tumors by analyzing brain magnetic resonance sensor images. First, a shallow autoencoder network was used for image reconstruction. It has excellent feature dimensionality reduction, robustness, and noise suppression capabilities, and thus reduces the likelihood of overfitting. Subsequently, the features extracted from the encoder were fed into a single-layer dense neural network, and finally, classification was tested on a softmax layer. The experimental results demonstrated that the incorporation of masking technology enabled the essential feature information to be precisely captured and resulted in highly satisfactory generalizability for unlabeled image test datasets. The developed model was trained and evaluated on the contrast-enhanced (CE)-MRI and Kaggle datasets, and achieved accuracies of 95.59 and 97.01%, respectively.

1. Introduction

Information technology and machine learning have found crucial applications in medicine. AI entails the development of machines capable of autonomous learning without human intervention, where this learning enables them to prepare for potential tasks. One critical application of AI is the development of measures for preventing brain tumors because tumor cells exhibit a highly complex and unpredictable behavior that cannot be controlled by traditional medicine. Death can result once tumor cells form in the human brain.(1) Brain tumors are tumors that initially form in the brain and subsequently spread to other parts of the body. AI systems are

*Corresponding author: e-mail: chihtayen@gmail.com
https://doi.org/10.18494/SAM5945

ISSN 0914-4935
https://myukk.org/

capable of learning from experience and applying their knowledge to solve new problems, such as those related to the accumulation of tumor cells in the brain.[2] One type of AI network, the convolutional neural network (CNN), has had a profound impact on the fields of medical imaging, AI, and digital image processing.[3]

According to the World Health Organization, gliomas are life-threatening primary tumors that are malignant, and they account for 50% of all incidences of tumors. Meningiomas and pituitary tumors are considered to be less severe than gliomas because they grow slowly and are benign. In patients with brain tumors, doctors must first identify the exact area of the brain that is affected before they can perform brain surgery or administer treatment. Tumor segmentation in brain images involves isolating the areas of these images that show tumor tissue rather than healthy tissue. This brain tumor segmentation process is one of the most challenging diagnostic techniques. Many of the techniques employed for brain tumor segmentation are not specifically designed for the brain tumor domain; they rely on edge-based, generic data, which may lead to inaccurate tumor type determination. Because of their ability to detect image features, deep learning algorithms have recently been employed for tumor segmentation.[4,5]

The accurate and timely diagnosis of brain tumors is crucial for improving treatment outcomes and patients' quality of life. According to Amin *et al*.,[6] the volume of a typical brain tumor can double in just 25 days, and without proper treatment, a patient is unlikely to survive longer than 12 months.[7] Given the severity of brain tumors, an automated brain tumor detection method is required. Currently, the manual analysis process that hospitals use to obtain numerous images is cumbersome, and the behavior of different tumors is not adequately interpreted. To better understand and predict the progression of complex brain lesions, a more precise, computer-based tumor detection and diagnosis technology is needed. Machine learning techniques have been employed to digitize detection processes in recent years. In particular, deep learning methods have emerged as the favored approach for achieving accurate and consistent detection of tumor cells.

Compared with X-rays and computed tomography, magnetic resonance imaging (MRI) provides more appropriate contrast characteristics in the imaging of various brain tissues. MRI is a suitable modality for diagnosis because it provides appropriate contrast for different brain tissues and employs controlled magnetic fields and radio waves, and different tissues relax at different rates when the RF pulse is turned off. Because of the complexity and diversity of tumors, MRI brain tumor detection can be a complex task. The processes of collecting, organizing, and analyzing medical images are now digital.[8] However, even when using cutting-edge technology, a thorough interpretation of medical images remains challenging because of time constraints and accuracy concerns. The challenge is particularly great when radiologists need to identify abnormal colors and shapes in specific regions.

Convolutional autoencoders (CAEs) have been shown to have great potential for brain tumor detection and diagnostic image analysis.[3] In this study, we employed a CAE model at the current knowledge level to address the complexity of imaging in brain tumor detection.

In one study,[9] approximately half the patients with brain tumors were found to experience headaches, and 15–95% of patients with brain tumors even experienced seizure. These symptoms can interfere with patients' functional abilities. In addition, brain tumors can cause

swelling in other parts of the body. There are 130 types of brain and central nervous system tumors, and they vary from benign to malignant and from extremely rare to relatively commozn. Tumors are also classified as either primary or secondary, which are ways of describing the origin of the tumor. Primary means that a condition or disease exists independently and occurs on its own; it is not caused by other factors. Secondary indicates that a condition or disease is caused by another underlying condition, situation, or factor.

In one study,[10] an automated brain tumor segmentation algorithm based on a 22-layer deep, three-dimensional CNN was proposed. This method was validated using the BRATS 2015 database, and the performance for the complete, core, and enhancing regions of the brain was evaluated using an online assessment platform. The Dice coefficients for the aforementioned regions were 0.84, 0.79, and 0.75, the positive predictive values were 0.88, 0.86, and 0.70, and the sensitivity values were 0.82, 0.75, and 0.86, respectively. In another study,[11] an approach based on CNNs and genetic algorithms was introduced. The model achieved an accuracy of 90.9% in classifying three glioma grades. Thus, deep learning is considered a favorable method in data science and AI for developing valuable decision-making capabilities by training models with data.

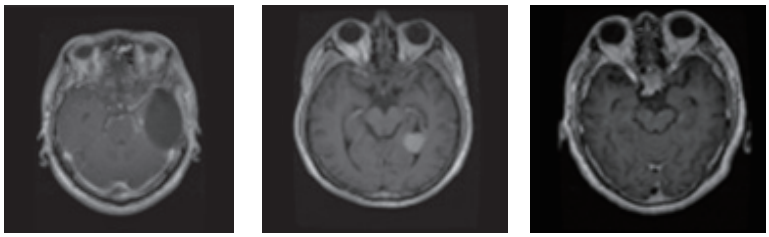
In research,[12] a deep wavelet autoencoder (DWA) was adopted for brain MRI disease classification. The encoded images obtained from input MRI data were processed through the discrete wavelet transform (DWT), and only approximate coefficients were calculated for the classification task. A deep neural network (DNN) with a learning rate of 0.8 and a batch size of 100 was employed for image classification. The developed DWA–DNN model achieved an average accuracy of 93%

Ali *et al.*[13] employed a deep convolutional generative adversarial network to generate high-quality synthetic MR images to expand a training dataset. A multimodal autoencoder was used to classify gliomas as low or high grade, with the results indicating a classification test accuracy of 92.04%. El-Dahshan *et al.*[14] introduced a hybrid method that classifies brain MR images as normal or abnormal. In this hybrid method, DWT and principal component analysis were conducted to reduce the number of features in images. A forward-propagation artificial neural network was then employed for image classification. However, the method was found to involve relatively long training and prediction times.

2. Research Objectives and Methods

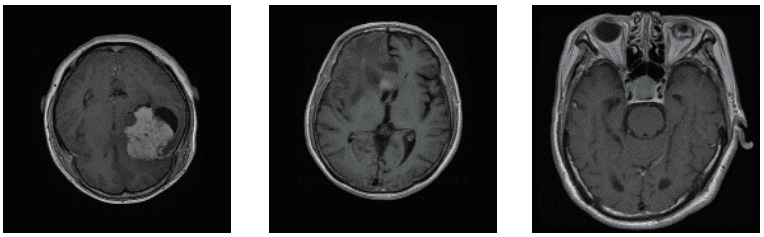
We employed brain MR images to develop a tumor classification technique. A deep learning network with CAEs was used, where the CAEs were constructed of neurons with learnable weights and biases. The primary objective of this study was to design a CAE model framework and determine the network's optimal hyperparameters to complete model training. Then, the model was employed to classify brain tumors to facilitate rapid, efficient, and accurate decision-making.

The publicly accessible Kaggle Brain Tumor dataset[15] and contrast-enhanced (CE)-MRI dataset[16] were employed; sample MR images for different brain tumors from the datasets are shown in Figs. 1 and 2, respectively. The Kaggle dataset is a brain tumor dataset with first- and



(a) (b) (c)

Fig. 1. Three categories of brain MR images in the Kaggle dataset: (a) glioma; (b) meningioma; and (c) pituitary tumor.



(a) (b) (c)

Fig. 2. Three categories of brain MR images in the CE-MRI dataset: (a) glioma; (b) meningioma; and (c) pituitary tumor.

second-order image features; the CE-MRI dataset contains 3064 2D MR slices collected from 233 individuals. The size of each image is 512 × 512 pixels for the axial, coronal, and sagittal views. In both datasets, the same three classification categories are employed: gliomas, meningiomas, and pituitary tumors. We standardized the images by collecting the broadest contour of each image and trimming any redundant proteins from the top, bottom, left, and right ends. Table 1 displays the numbers and categories of samples for the brain MRI datasets. The Kaggle dataset has a greater total number of samples than the CE-MRI dataset and exhibits a more balanced distribution of the three sample types.

We established a three-stage tumor classification system. The first stage is data preprocessing. Contrast-limited adaptive histogram equalization (CLAHE) was employed to improve the uniformity of the pixel distribution, thereby enhancing image contrast and suppressing noise. This process clarified image details and contours.[17] Subsequently, masking technology was applied to the processed images to emphasize the importance of specific image areas.

In the second stage, an autoencoder model[18] was used for high-level feature extraction and image reconstruction. We modified the architecture proposed in Ref. 19. The network's input image size was 128 × 128 × 3 pixels; the encoder contained three convolutional layers, filters, and maximum pooling layers to generate feature maps, and four convolutional layers were employed for image reconstruction, as illustrated in Fig. 3. The size of the kernel used to identify significant features was set to 3 × 3. We optimized the model by adjusting the number of output channels to 3 and the learning rate of the optimizer to 0.01. This adjustment enabled the model to learn feature representation across more channels and improved its feature extraction capability. Modifying the learning rate also contributed to the stability of the training process, facilitating superior model convergence.

Table 1
Numbers of images in the Kaggle and CE-MRI datasets.

Dataset	Kaggle			CE-MRI		
Category	Glioma	Meningioma	Pituitary tumor	Glioma	Meningioma	Pituitary tumor
Number	1621	1645	1757	1426	708	930
Total		5023			3064	

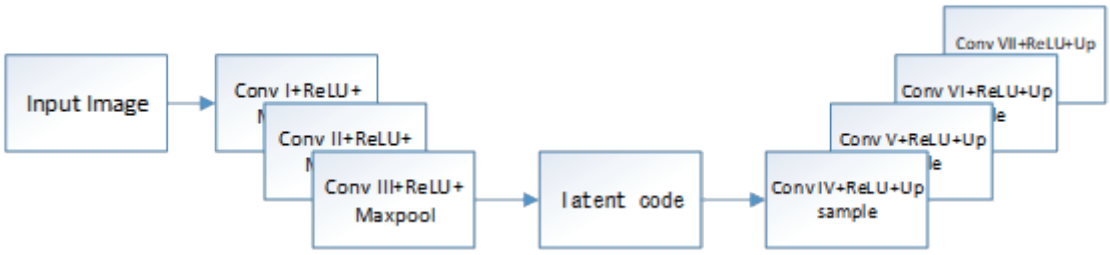


Fig. 3. (color online) Architecture of the autoencoder.

The third stage is focused on the classification task. The features trained by the encoder were extracted and then passed through a flatten layer and two fully connected layers. Finally, training and testing were conducted on the softmax layer for classification. In the following sections, the methods proposed in this study are described.

2.1. Data preprocessing

Data extraction was performed on the original data. The size of the original images was 512 × 512 pixels in the axial, coronal, and sagittal views. To enhance computational speed, the images' size was adjusted to 128 × 128 pixels. The data were divided into training and validation sets at a ratio of 8:2.

2.1.1 CLAHE

CLAHE is an image enhancement technique[17] designed to improve image contrast and detail. It is an improvement upon histogram equalization (HE) and is capable of enhancing contrast within local areas, avoiding the problems of overenhancement and noise amplification that can occur with traditional HE during global enhancement.

The fundamental concept of CLAHE is to divide the image into numerous small tiles and then equalize the histogram within each tile. This localized approach to processing helps prevent discontinuities between tiles within the image and retains local details. In CLAHE, HE within each tile is achieved by stretching the local histogram. Initially, a histogram for each tile is determined, and the pixel values in the histogram are equalized. For pixels in the histogram exceeding a predetermined contrast limit, contrast limiting is applied to prevent the overenhancement of contrast.

Although traditional HE can improve the contrast of an image, it is a global image processing technique. When applied to images with uneven lighting, it tends to darken the dark areas and brighten the light areas, leading to major information loss. This outcome is unacceptable in

medical imaging. Hence, to preserve image details while improving image quality, we employed CLAHE. Figure 4 presents nonprocessed images (upper) and the corresponding CLAHE-processed images (lower). The processed images have clearer details and contours. To obtain data suitable for training, we processed MR images through CLAHE. The contrast threshold was set to 2, and every 8 × 8 pixel tile was processed. After processing, the original image (128 × 128 × 1 pixels) and the CLAHE-processed image (128 × 128 × 1 pixels) were concatenated, resulting in data with fused features (of size 128 × 128 × 2 pixels). This process clarified image features and improved model accuracy.

2.1.2 Masking technology

Masking technology[20] is used to modify specific parts of an image or data to emphasize or conceal certain information. In image processing, masking can be used to set specific areas of an image to certain values or pixels, thereby directing a model's attention to these parts and improving the accuracy and robustness of feature extraction. In this study, we employed masking technology to highlight the central part of the image, directing the model's attention to the central features. A rectangular mask was used to mark the central part of the image as the area of interest, and the pixel values of the central part were set to 0.2. The fixed rectangular mask with a value of 0.2 was introduced to encourage the autoencoder to focus on the central region of the image, where tumors are frequently located, while simultaneously preserving weak low-level intensity information to prevent gradient signal vanishing. This masking technique was employed during the training of the autoencoder model and was an effective approach to emphasizing features in unsupervised learning.[21]

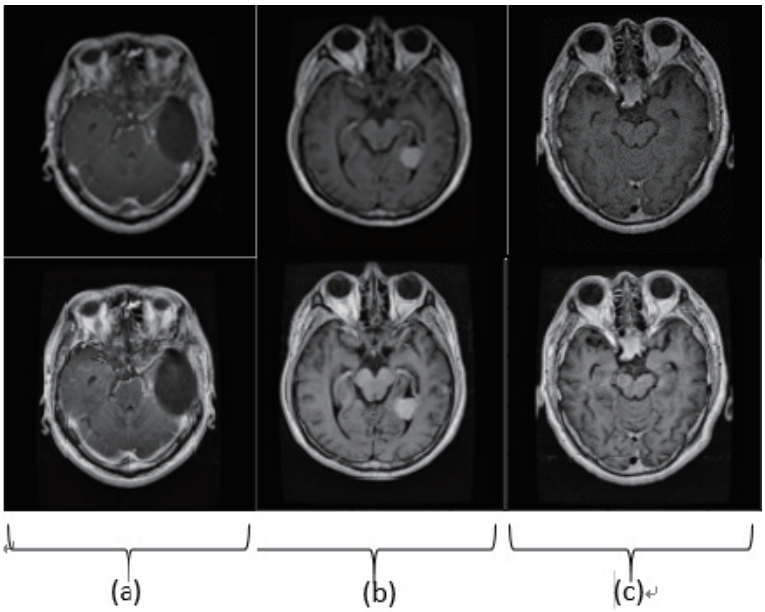


Fig. 4. Nonprocessed (upper) and CLAHE-processed brain MR images: (a) glioma; (b) meningioma; and (c) pituitary tumor.

2.2 Convolutional autoencoder

CNNs have been demonstrated to have potential for brain imaging analysis in various domains and have been proven effective in classifying brain diseases from MR images. Brain tumor classification is a relatively complex and challenging task among classification problems. The primary challenges involve the following:

1. Brain tumors vary greatly in their severity, shape, and size,[9] and tumors of different pathological types may have similar appearance.
2. Despite the need for the labeling of extensive training data, the existing medical imaging datasets remain relatively limited.

To address these challenges, we employed a CAE, which has been shown to help physicians diagnose and detect abnormalities, and to facilitate accurate and rapid decision-making. An autoencoder is an unsupervised deterministic feedforward artificial neural network and a technique for reducing the dimensionality of data.[22] The initial concept is straightforward: an image is input and processed through a neural network, and an output that is as close to the input image as possible is then produced. Typically, an autoencoder can be decomposed into an encoder and a decoder, two neural networks that perform completely opposite operations. The encoder processes the input image through a neural network to produce a vector $\boldsymbol{z}$ with fewer dimensions. This vector $\boldsymbol{z}$ is then input to the decoder, which restores it to the original size of the image, as shown in Fig. 3. The vector $\boldsymbol{z}$ contains the main information describing the distribution of the input data. Mathematically, the encoder is represented as

$$\boldsymbol{z} = f_{enc}(\boldsymbol{w}_{enc\boldsymbol{x}} + \boldsymbol{b}_{enc}), \tag{1}$$

where $\boldsymbol{w}_{enc\boldsymbol{x}}$ is the weight matrix, $\boldsymbol{b}_{enc}$ is the bias vector, $\boldsymbol{x}$ is the input vector, and f_{enc} is the activation function, which performs a nonlinear transformation on the linear mapping $\boldsymbol{w}_{enc\boldsymbol{x}} + \boldsymbol{b}_{enc}$. Typically, f_{enc} is the sigmoid, tanh, rectified linear unit (ReLU) or linear function. The encoder processes the original input image into a low-dimensional vector $\boldsymbol{z}$, which represents the hidden features within the image, known as the latent code. Obtaining the latent code is an efficient method for representing data features for classification or generation purposes. However, because data contain various interconnected and coupled features, models have difficulty discerning their relationships, leading to inefficient learning. The deep relationships hidden beneath the superficial features should be explored, and these relationships should then be decoupled to extract hidden features. Therefore, how the CAE network should be structured and the parameters set on the basis of input data is a critical issue.

The decoder converts the latent code into the input image as much as possible, remapping the feature vector from the low-dimensional space back to the high-dimensional space. All operations within the decoder are precisely opposite those within the encoder. The model reconstructs an estimate of the input data vector $\boldsymbol{x}$ by taking the latent vector $\boldsymbol{z}$ as an input. This model can be represented by

$$\hat{\boldsymbol{x}} = f_{enc}(w_{dec\boldsymbol{x}} + b_{dec}), \tag{2}$$

where $\boldsymbol{w}_{dec\boldsymbol{x}}$ and $\boldsymbol{b}_{dec}$ are the decoder's weight matrix and bias vector, respectively, and $\hat{\boldsymbol{x}}$ is the reconstruction vector. The purpose of training an autoencoder is to learn and update the weight matrix and bias vector. For a given input vector $\boldsymbol{x}$, calculating the estimated vector $\hat{\boldsymbol{x}}$ and comparing it with the original target enable the assessment of the autoencoder's performance. Therefore, the loss function can be simply defined as the error between these two images. Learning is accomplished by minimizing the loss function. The loss function most commonly used for this architecture is the mean square error:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}\left\|\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n\right\|, \tag{3}$$

where N is the number of observed values and their corresponding predicted values, $\boldsymbol{x}_n$ is the actual value of the n-th observation, and $\hat{\boldsymbol{x}}_n$ is the model's predicted value for the n-th observation. In the proposed autoencoder, the ReLU, expressed below, was used as the activation function. Table 2 presents the suggested architecture for the image reconstruction autoencoder.

$$ReLU(x)=\begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{4}$$

The signal extraction process strongly affects the intensity of MR images; therefore, preprocessing must be conducted to normalize images. In this study, min–max scaling was adopted, whereby data features are scaled to within a fixed range and features are normalized to the range [0, 1], expressed as

$$\boldsymbol{X}_{normalized} = \frac{\boldsymbol{X} - \boldsymbol{X}_{min}}{\boldsymbol{X}_{max} - \boldsymbol{X}_{min}}, \tag{5}$$

where $\boldsymbol{X}$ is the original data feature value and $\boldsymbol{X}_{min}$ and $\boldsymbol{X}_{max}$ are the minimum and maximum values of feature $\boldsymbol{X}$, respectively.

Table 2
Suggested architecture for the image reconstruction autoencoder.

Layers	Filters	Output shape
ConvI + ReLU + Maxpool	3 × 3 × 32	64 × 64 × 32
ConvII + ReLU + Maxpool	3 × 3 × 64	32 × 32 × 64
ConvIII + ReLU + Maxpool	3 × 3 × 128	16 × 16 × 128
Decoder layers		
ConvIV + ReLU + Up sample	3 × 3 × 128	32 × 32 × 128
ConvV + ReLU + Up sample	3 × 3 × 64	64 × 64 × 64
ConvVI + ReLU + Up sample	3 × 3 × 32	128 × 128 × 32
ConvVII + ReLU + Up sample	3 × 3 × 1	128 × 128 × 3

2.3 Classifier

We used the features extracted by the autoencoder to train the classifier. Table 3 presents the architecture of the classifier and Table 4 displays the architecture of the classifier used.[19] The flatten layer spreads the feature maps into a one-dimensional vector, which is then input to the fully connected layer. The dense layer is a fully connected layer in which each neuron is connected to all neurons in the previous layer, meaning that each output from the previous layer becomes an input for each neuron in the dense layer. The neurons in the dense layer typically have weights and biases, which are learned during the training process. The fully connected layer combines the features from the previous layer to form higher-level feature representations that are ultimately used for the model's output or further processing. Finally, classification is performed in the softmax layer. Compared with using another autoencoder for classification,[19] we employed a smaller model that exhibits faster training convergence and has fewer parameters and simpler classification layers, resulting in the network having higher generalizability.

2.4 Deep learning development tools

Some researchers have used development tools such as MATLAB for brain tumor detection. However, we used Python programming to achieve our goals. The reasons for selecting Python were as follows.

1. Python is free and available through open source and offers a greater array of datasets and graphic packages than does MATLAB.[23]
2. Python code is more concise and accurate than is the MATLAB code. Python provides greater control over implementations, allowing for more effective implementations and the naming of visuospatial skills.

Table 3
Architecture of the classifier.

Layers	Output shape	Param #
Input layer	16 × 16 × 128	0
Flatten	32768	0
Dense (Dense)	128	4194432
Dense_1 (Dense)	3	387
Total parameters	4194819	
Trainable parameters	4194819	

Table 4
Architecture of the classifier used by Jyothi *et al*.[19]

Layers	Filters	Output shape
Conv I + ReLU + BN	3 × 3 × 32	128 × 128 × 32
ConvII + ReLU + BN + Maxpool	32768	0
ConvIII + ReLU + BN	128	4194432
ConvIV + ReLU + BN + Maxpool	3	387
Flatten layer	4194819	
Dense I	128	
Dense II	3	
Total parameters	16938051	

3. Python generally utilizes resources more efficiently on the same hardware configuration than does MATLAB. For example, the image processing and analysis required for brain tumor detection may consume less memory and computational power when performed in Python, enabling a smoother operation in limited hardware environments.
4. Certain libraries and tools in Python have higher performance and efficiency when handling large datasets, which could be crucial in image processing for brain tumor detection.

2.5 Scoring criteria

The following indices were used to evaluate our model's performance. They are based on the four fundamental elements of a confusion matrix: the numbers of true positives (*TP*), true negatives (*TN*), false positives (*FP*), and false negatives (*FN*). The four indices are as follows.

$$Accuracy\,(\%) = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$Precision\ (PPV) = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1\ score = \frac{2(Percision * Recall)}{Percision + Recall} \tag{9}$$

The F1 score is the mean of precision and sensitivity, offering a balance between them. It is particularly suitable for situations involving imbalanced categories.

3. Results

The fundamental goal of this study was to develop an improved brain MRI recognition model that has a high accuracy of brain tumor detection. We made improvements to the model architecture from that previously reported.[19] We adjusted the number of output channels to enhance the model's ability to recognize specific features. Additionally, we optimized the learning rate for the training process to increase the model's convergence speed and accuracy. CLAHE and masking technology were incorporated for image preprocessing, enhancing image contrast, and reducing unnecessary background noise, thereby providing the model with clearer, more informative inputs. Figure 5 illustrates the workflow of our proposed brain tumor classification method. To assess the performance of the improved model, we used a confusion matrix and related performance indices. A confusion matrix is a $N \times N$ matrix, where N represents the number of categories. Tables 5 and 6 present the confusion matrices for the two

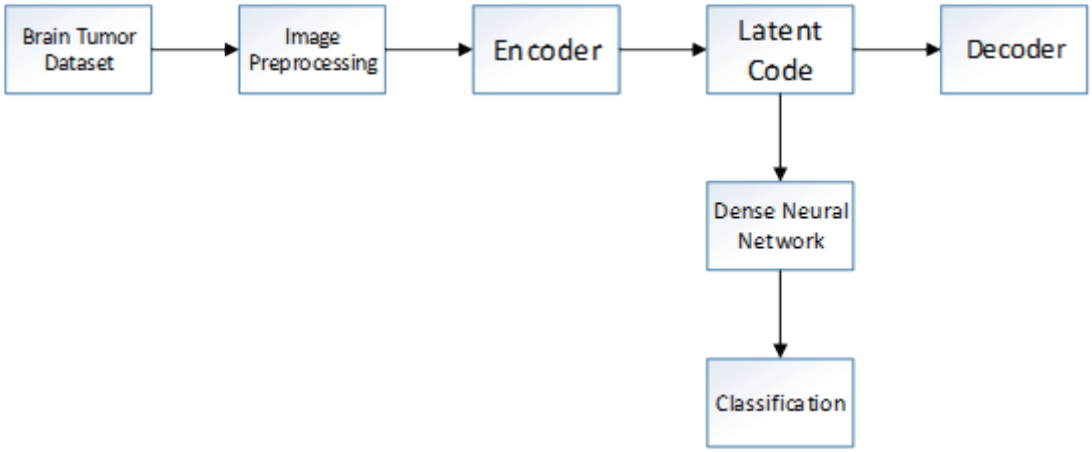


Fig. 5. (Color online) Workflow of our proposed brain tumor classification method.

Table 5
Confusion matrix for the Kaggle dataset after classification.

Kaggle	Glioma	Pituitary	Meningioma
Glioma	332	17	1
Pituitary	10	301	1
Meningioma	0	1	342

Table 6
Confusion matrix for the CE-MRI dataset after classification.

CE-MRI	Glioma	Pituitary	Meningioma
Glioma	283	10	0
Pituitary	12	128	3
Meningioma	0	2	175

datasets after the Adam classifier was applied. The numbers on the main diagonal represent the number of samples correctly predicted by the model, whereas the off-diagonal elements indicate errors in the model's predictions.

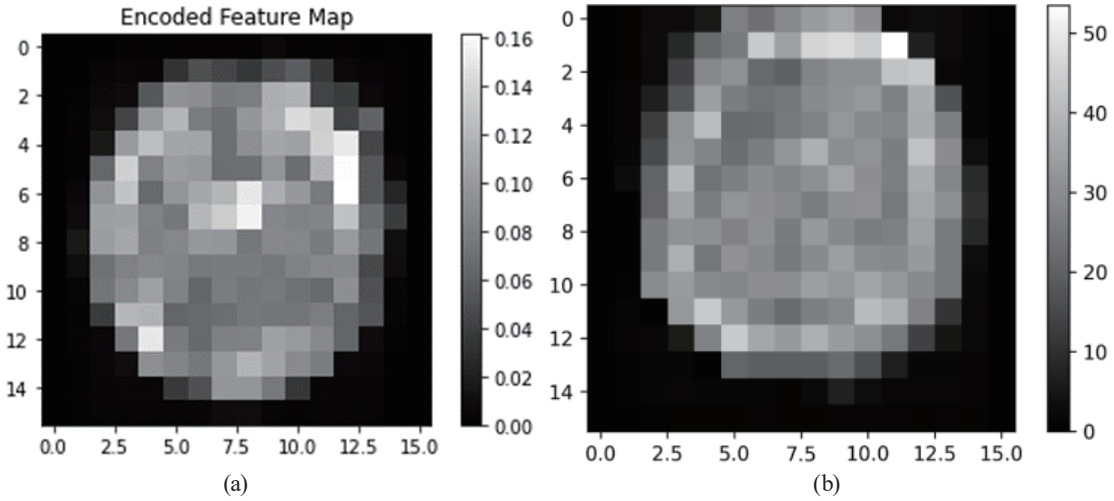
Table 7 displays the average performance index values for the Adam classifier in classifying the two datasets, namely, the accuracy, recall, precision, and F1 score. These indices were used to evaluate the model's predictive ability for three categories. Accuracy indicates the proportion of correct predictions made by the model among all predictions. Recall indicates the model's ability to find all positive class samples. Precision represents the proportion of samples predicted as positive by the model that were actually positive. Finally, the F1 score is the harmonic mean of the precision and recall and thus offers a balance between them.

Figure 6 presents the feature maps generated by the filter. When an image or another type of data is passed through the filter, the filter responds to different features and produces corresponding feature maps. These feature maps provide information about the input data and enable feature extraction at various levels of the model.

Figures 7 and 8 present the sample of images reconstructed by the autoencoder and the corresponding original images for the two datasets. The reconstructed images were blurrier than

Table 7
Average performance index values for the classifier and two datasets.

Dataset	Accuracy (%)	Recall (%)	Precision (%)	F1 score (%)
CE-MRI	95.59	95.22	94.98	95.10
Kaggle	97.01	96.95	97.01	96.97

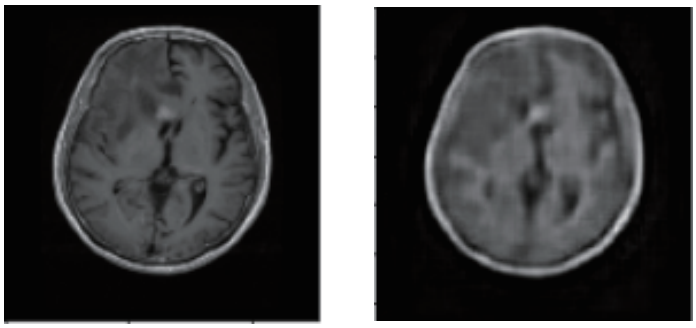


(a) (b)

Fig. 6. Feature maps generated by the filter: (a) Kaggle and (b) CE-MRI dataset images.

(a) (b)

Fig. 7. (a) MR image from the Kaggle dataset and (b) corresponding image reconstructed using the autoencoder.



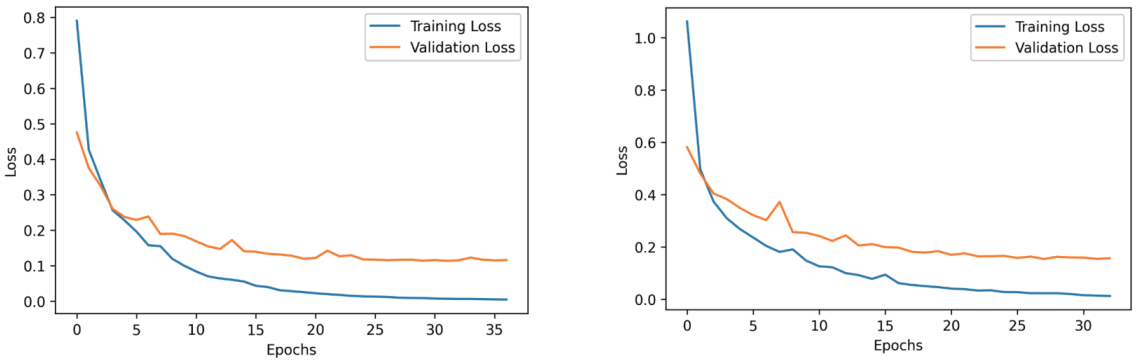
(a) (b)

Fig. 8. (a) MR image from the CE-MRI dataset and (b) corresponding image reconstructed using the autoencoder.

the original images; this was because the model compressed the original information into a lower-dimensional representation and then attempted to reconstruct the original image. However, the primary task of our model was to extract meaningful and useful features rather than to perfectly restore the original image. The blurriness may serve to express higher-level features through less information. This compromise might lead to slight distortions in image details but the retention of the most crucial features.

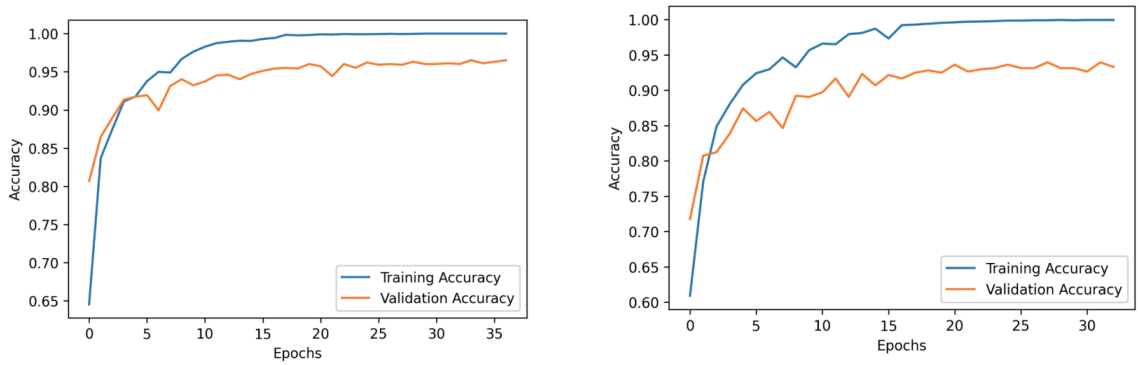
3.1 Training and validation

The training and validation accuracies and losses for the two datasets using the Adam classifier are illustrated in Figs. 9 and 10. The training and validation accuracies both improved over time, and meaningful features for the classification task were obtained. Figure 11 displays the receiver operating characteristic curves obtained after classification.



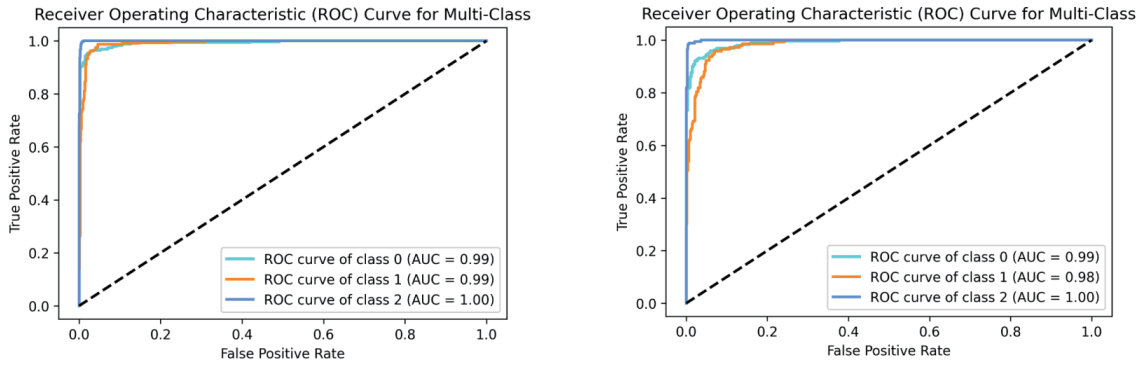
(a) (b)

Fig. 9. (Color online) Training and validation losses: (a) Kaggle and (b) CE-MRI datasets.



(a) (b)

Fig. 10. (Color online) Training and validation accuracies: (a) Kaggle and (b) CE-MRI datasets.



(a) (b)

Fig. 11. (Color online) Receiver operating characteristic curves for the (a) Kaggle and (b) CE-MRI datasets, where class 0: glioma; class 1: meningioma; and class 2: pituitary tumor.

Table 8
Performance indices for proposed method and in previous study for the same dataset.

Method	Accuracy (%)	Recall (%)	Precision (%)	F1 score (%)
Shallow autoencoder[19]	95.26	91.87	93.51	92.68
Ours	95.59	95.22	94.98	95.10

Table 9
Results of our model and models presented in literature.

Method	Dataset	Accuracy (%)
Multistream CAE and 2D generative adversarial network for glioma classification[13]	BraTS17	92.04
Multiscale CNN for brain tumor segmentation and diagnosis[24]	BraTS17	81
Deep learning and multisensor fusion for glioma classification by using Multistream 2D convolutional networks[25]	BraTS17	90.87
Deep learning for brain tumor classification[26]	CE-MRI	90.26
Transfer learning and fine-tuning on MR images for brain tumor classification[27]	CE-MRI	94.82
CNN for brain tumor classification[28]	CE-MRI	84.18
Shallow-autoencoder-based neural network for multiclass classification of brain tumors using MR images[19]	CE-MRI	95.26
Our proposed model	CE-MRI	95.59

3.2 Result analysis and discussion

The confusion matrices presented in Tables 5 and 6 reveal that for both datasets, the model made more errors in classifying gliomas and pituitary tumors than in classifying meningiomas. This was attributable to certain similarities between gliomas and pituitary tumors. The results presented in Tables 6 and 7 indicate that the classifier exhibited a higher performance for the Kaggle dataset than for the CE-MRI dataset. This was possibly because the Kaggle dataset is larger and has a better balance among the different categories. Table 8 shows our model's performance index values and those from Ref. 19 for the same CE-MRI dataset. Our model exhibited superior performance on the test set, achieving higher accuracy, recall, precision, and

F1 score. This implies that the proposed CAE model performs effectively on different datasets for the targeted classification tasks, confirming its greater generalizability than that of the model in Ref. 19.

4. Comparative Analysis

Table 9 shows the results of our proposed model and those of the models introduced in other studies for the CE-MRI or BraTS17 dataset.[4] In addition to the CE-MRI dataset, we used the Kaggle dataset to validate our model and obtained an accuracy of 97.01%. According to the indices in Table 8, our model exhibited the highest accuracy and generalizability among all models on the same dataset.

5. Conclusions

We proposed an autoencoder architecture that is based on a previously reported architecture[19] and evaluated the number of output channels to enhance the model's ability to identify specific features. In addition, we optimized the learning rate during the training process to improve the model's convergence speed and accuracy. CLAHE and masking technology were introduced for image preprocessing to enhance image contrast and eliminate unnecessary background noise, thus providing clearer and more informative inputs for the model, increasing its accuracy. The proposed method was found to achieve a higher accuracy for brain MRI classification on the Kaggle and CE-MRI datasets than that achieved by classical machine learning techniques; the model's accuracy reached 95.59 and 97.01% for these datasets, respectively. Furthermore, the proposed method was compared with those introduced in previous studies (Table 4) in terms of performance and discovered to exhibit superior accuracy and stability. Additionally, it is lighter in weight, which could reduce its computational burden and enable its use on portable computing devices in the future, indicating superior scalability. Beyond accuracy, our proposed method offers several advantages compared with CNN or transfer learning approaches. Our model is computationally efficient, with only ~4.19M parameters, significantly fewer than the 10–20M parameters reported in some prior studies, and its lower FLOPs enables deployment on CPUs or mid-range GPUs. Furthermore, the lightweight design makes the model suitable for low-resource devices and clinical edge deployment, as it requires minimal memory, can be readily converted to ONNX or TFLite formats, and can be further optimized through 8-bit quantization to reduce latency and power consumption. Collectively, these characteristics make our model highly practical for clinical applications requiring lightweight deployment. However, the current model still has potential for improvement; for example, attention mechanisms could be incorporated, which may be a key direction for future research. The inclusion of attention mechanisms (e.g., spatial attention or channel attention mechanisms) could enhance the model's attention to essential features, thereby reducing error rates.

Acknowledgments

We are thankful for the support provided by the National Science and Technology Council NSTC (grant nos. NSTC 113-2221-E-011-163, NSTC 114-2221-E-011-146) and the National Taiwan University of Science and Technology. We also thank the editor for his kind coordination. Moreover, we are grateful to the editor and reviewers for their constructive suggestions.

References

1 V. Anitha and S. Murugavalli: IET Comput. Vis. **10** (2016) 9. https://doi.org/10.1049/iet-cvi.2014.0193
2 S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang: Clin. Pharmacol. Ther. **107** (2020) 871. https://doi.org/10.1002/cpt.1796
3 F. Bashir-Gonbadi, H. Khotanlou, I. S. Jacobs, and C. P. Bean: Multimed. Tools Appl. **80** (2021) 19909. https://doi.org/10.1007/s11042-021-10637-1
4 B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput: IEEE Trans. Med. Imaging **34** (2015) 1993. https://doi.org/10.1109/TMI.2014.2377694
5 M. Yaqub, J. Feng, M. S. Zia, K. Arshid, K. Jia, Z. U. Rehman, and A. Mehmood: Brain Sci. **10** (2020) 427. https://doi.org/10.3390/brainsci10070427
6 J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes: Pattern Recognit. Lett. **139** (2020) 118. https://doi.org/10.1016/j.patrec.2017.10.036
7 S. V. Ellor, T. A. Pagano-Young, and N. G. Avgeropoulos: J. Law Med. Ethics **42** (2014) 171. https://doi.org/10.1111/jlme.12133
8 L. M. De Angelis: N. Engl. J. Med. **344** (2001) 114. https://doi.org/10.1056/NEJM200101113440207.
9 T. A. Dolecek, J. M. Propp, N. E. Stroup, and C. Kruchko: Neuro-Oncol. **14** (2012) v1. https://doi.org/10.1093/neuonc/nos218
10 R. S. Abu Al-Araj, S. K. Abed, A. N. Al-Ghoul, and S. S. Abu-Naser: Int. J. Acad. Eng. Res. (IJAER) **4** (2020) 23.
11 H. H. Cho and H. Park: 39th Ann. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC) 3081. https://doi.org/10.1109/EMBC.2017.8037508
12 P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari: IEEE Access **7** (2019) 46278. https://doi.org/10.1109/ACCESS.2019.2902252
13 M. B. Ali, I. Y.-H. Gu, and A. S. Jakola: Proc. Int. Conf. Comput. Anal. Images Patterns (2019) 234–245. https://doi.org/10.1007/978-3-030-29888-3_19.
14 E. A. El-Dahshan, T. Hosny, and A.-B. M. Salem: Digit. Signal Process. **20** (2010) 433. https://doi.org/10.1016/j.dsp.2009.07.002
15 Brain Tumor Dataset: https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor/data (accessed 2020).
16 Brain Tumor Dataset (version 5): https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri (accessed 2017).
17 K. Zuiderveld: Graphics Gems, P. S. Heckbert, Ed. (Academic Press Professional, Inc. San Diego, 1994) pp. 474–485. https://doi.org/10.1016/B978-0-12-336156-1.50061-6
18 C. Baur, B. Wiestler, S. Albarqouni, and N. Navab: Proc. Int. MICCAI Brainlesion Workshop, Granada, Spain, (2018) 161. https://doi.org/10.1007/978-3-030-11723-8_16
19 P. Jyothi, S. Dhanasekaran, and R. A. Singh: IEEE 2nd Int. Conf. Mobile Networks and Wireless Communications (ICMNWC) (2022) 1. https://doi.org/10.1109/ICMNWC56175.2022.10031727
20 P. Viola and M. J. Jones: Int. J. Comput. Vis. **57** (2004) 137. https://doi.org/10.1023/B:VISI.0000013087.49260.fb
21 M. D. Zeiler and R. Fergus: 2014 Proc. Eur. Conf. Comput. Vis. (ECCV) 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

22 S. Takaki and J. Yamagishi: 2016 Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP) 5535–5539. https://doi.org/10.1109/ICASSP.2016.7472692
23 R. L. Devi: Proc. 5th Int. Conf. I-SMAC (IoT Soc., Mobile, Analytics Cloud) (2021) 1280–1284. https://doi.org/10.1109/I-SMAC52330.2021.9640670
24 P. M. S. Raja and A. V. Rani: Biocybern. Biomed. Eng. **40** (2020) 440. https://doi.org/10.1016/j.bbe.2020.01.006
25 C. Ge, I. Y.-H. Gu, A. S. Jakola, and J. Yang: Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC) (2018) 5894–5897. https://doi.org/10.1109/EMBC.2018.8513556
26 S. Paul, J. Joseph, P. Andrew, A. Plassard, B. A. Landman, and D. Fabbri: Biomedical Applications in Molecular, Structural, and Functional Imaging **10137** (2017) 1013710. https://doi.org/10.1117/12.2254195
27 Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu: Comput. Med. Imaging Graphics **75** (2019) 34. https://doi.org/10.1016/j.compmedimag.2019.05.001
28 N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko: World Congr. Medical Physics and Biomedical Engineering 2018, IFMBE Proc. **68** (2019) 183. https://doi.org/10.1007/978-981-10-9035-6_33

About the Authors

Chih-Ta Yen received his PhD from the Department of Electrical Engineering at National Cheng Kung University, Taiwan, in 2008. He is currently a professor in the areas of AI applications, multiple access communications and optical design technologies at the Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, Taiwan. His major interests are in the areas of multiuser optical communications, wireless communication systems, machine learning, deep learning, and optical design. (chihtayen@gmail.com)

Hao-Siang Kao was born in Taipei, Taiwan, in February 2000. He is studying for his MS degree in the Department of Electrical Engineering, National Taiwan Ocean University, Taiwan. His major interests are in the areas of biomedical deep learning.
(jeremy26954646@gmail.com)