Baghdad Science Journal

Volume 23 | Issue 6 Article 27

6-26-2026

Weighted Average Ensemble Learning Model for Improving Brain Tumor Classification

Laith H. Alhasnawy
Presidency of University of Babylon, University of Babylon, Babylon, Iraq, laith.alhasnawi4@uobabylon.edu.iq

Hasanain Ali Al Essa
Presidency of University of Babylon, University of Babylon, Babylon, Iraq, hasanain@uobabylon.edu.iq

Tahseen A. Wotaifi
Presidency of University of Babylon, University of Babylon, Babylon, Iraq, tahseen.ali@uobabylon.edu.iq

Abbas F. Hommadi
Presidency of University of Babylon, University of Babylon, Babylon, Iraq, abbas.hommadi@uobabylon.edu.iq

Ameer K. AL-Mashanji
Presidency of University of Babylon, University of Babylon, Babylon, Iraq, amir.mashanji@uobabylon.edu.iq

How to Cite this Article

Alhasnawy, Laith H.; Essa, Hasanain Ali Al; Wotaifi, Tahseen A.; Hommadi, Abbas F.; and AL-Mashanji, Ameer K. (2026) "Weighted Average Ensemble Learning Model for Improving Brain Tumor Classification," *Baghdad Science Journal*: Vol. 23: Iss. 6, Article 27.
DOI: https://doi.org/10.21123/2411-7986.5341

BAGHDAD SCIENCE JOURNAL 2026;23(6):2323–2335

RESEARCH ARTICLE

Weighted Average Ensemble Learning Model for Improving Brain Tumor Classification

Laith H. Alhasnawy *, Hasanain Ali Al Essa, Tahseen A. Wotaifi, Abbas F. Hommadi, Ameer K. AL-Mashanji

Presidency of University of Babylon, University of Babylon, Babylon, Iraq

ABSTRACT

The detection of Brain Tumors (BTs) is of great benefit in identifying the growth of dangerous cells in the brain that pose a threat to the patient's life. Early detection of BTs remains a significant challenge due to subtle differences among tumor types and the complexity of brain anatomy. Manual diagnosis by radiologists is often time-consuming and subject to variations, which may delay treatment. Among available imaging techniques, Magnetic Resonance Imaging (MRI) is widely used because of its high-resolution capabilities. Deep Learning (DL) has achieved remarkable success in image classification tasks through Transfer Learning (TL), which utilizes pre-trained models. Therefore, this study proposes a new model, namely Weighted Average Ensemble Learning (WAEL), based on TL for BT classification. Convolutional Neural Network (CNN) architectures, including Visual Geometry Group (VGG16), Visual Geometry Group (VGG19), and Inception_v3, were employed as classifiers within the WAEL model for automatic BT prediction. The model was trained using the Figshare BT dataset, containing 3,064 images from three tumor classes (meningioma, glioma, and pituitary tumors). To enhance generalizability, benign (no-tumor) images were incorporated into the dataset. Experimental results demonstrated superior performance, with the WAEL model achieving an accuracy of 0.9908, compared to 0.9809, 0.9786, and 0.9832 for VGG16, VGG19, and Inception_v3, respectively. Furthermore, the proposed model achieved strong precision, recall, and F1-score values, confirming its robustness. The WAEL model also outperformed previous studies using the same dataset. Thus, the WAEL model can be a secondary tool for radiologists to discover tumors from brain MRI images.

Keywords: BTs, CNN, DL, Inception_v3, MRI, ML, TL, VGG16, VGG19, WAEL

Introduction

BTs are the spread of abnormal cells in or around the brain, and these cells affect brain functions and cause damage to healthy cells. BTs are classified into three types: meningioma, glioma, and pituitary tumors. These tumors are usually either benign or malignant. Early detection of these tumors is important to increase the chances of a patient being cured. There are several challenges that medical experts face in trying to identify BTs cells, as they occur in different locations and are of different dimensions and sizes. [1,2]

Medical image technology has contributed a lot to facing these challenges because many techniques were used to analyze the medical images, and the most important ones are MRI images, which play a vital role in BTs classification due to their high accuracy in providing comprehensive information about BTs characteristics. They are rather suitable for analysis using artificial intelligence techniques. The strengths of MRI images reflect their ability to provide not only a medical image but also a reliable, rich, and flexible feature set for building automated BTs classification models. These advantages make it ideal for AI applications that depend on DL and TL for improving the performance and accuracy of classification in distinguishing between BTs types. [3–5]

Received 30 April 2025; revised 9 October 2025; accepted 20 December 2025.
Available online 26 June 2026

* Corresponding author.
E-mail addresses: laith.alhasnawi4@uobabylon.edu.iq (L. H. Alhasnawy), hasanain@uobabylon.edu.iq (H. A. Al Essa), tahseen.ali@uobabylon.edu.iq (T. A. Wotaifi), abbas.hommadi@uobabylon.edu.iq (A. F. Hommadi), amir.mashanji@uobabylon.edu.iq (A. K. AL-Mashanji).

https://doi.org/10.21123/2411-7986.5341

Over the past several years, CNNs have emerged as a powerful approach in analyzing medical images. These networks are typified by their capabilities for extracting salient features from the images using ML without human intervention. These networks have realized exceptional performances in classification, pattern recognition, and other tasks.[5–7]

DL is a branch of Machine Learning (ML) that is capable of learning through self-learning without human intervention, based on unlabeled and unstructured data, by extracting the hierarchical features from the data. DL techniques were used in various applications such as speech recognition, object detection, pattern classification, and performing other decision tasks. DL techniques have been used to solve a lot of complex problems and give meaningful results.[2,6] In the healthcare field, when training DL models, there is difficulty in obtaining medical data that can be shared publicly due to the security and privacy of the data. To address the lack of that data, TL has been mainly used in the medical field.[3,4,8]

In TL, a DL model that has been previously trained and developed for a particular task is used as a starting point for a learning procedure for another task.[6,7] TL is characterized by the speed of learning, as the training time is reduced when using the model that has been previously trained, and it gives accurate results using little data because it uses the model that was previously trained with much data, and it also provides better performance through adding new fully connected layers to the model that was previously trained[1,9,10]

In this study, a TL-based model, WAEL, has been proposed for BTs classification. In this proposal, CNNs represented by VGG16, VGG19, and Inception_v3 were used to classify BTs from MRI images into meningioma, glioma, and pituitary tumors. This model was trained and tested on a publicly available Kaggle dataset. The proposed model achieved better performance in terms of precision, accuracy, F1-score, and recall, reaching 99%. The main contributions of this study are as follows:

- A new WAEL model is proposed, making its decision based on three pre-trained models (VGG16, VGG19, and Inception_v3) to improve the prediction accuracy of BTs types from MRI images.
- The study conducts a comparative performance evaluation with pre-trained CNN models for each model individually, and the proposed model yields better results in key metrics such as accuracy, precision, recall, and F1-score.
- The WAEL model attains state-of-the-art results on a widely used MRI-based BTs dataset, emphasizing its potential as a dependable computer-aided diagnostic tool to support radiologists in clinical decision-making.

Related works

Several models for BTs detection and classification have been presented in related works.

Srinivas et al.[1] presented a performance analysis comparison between CNN architectures that leverage pre-trained TL models. These architectures are VGG16, ResNet_50, and Inception_v3. The upper layers of these architectures were fine-tuned using MRI images to improve their performance in classifying BTs. However, this work relied exclusively on independent model evaluation without any fusion strategy. Consequently, their framework lacks robustness across heterogeneous MRI datasets. In contrast, the proposed WAEL model in this research mitigates such instability by aggregating multiple models' outputs to enhance consistency.

Ahmmed et al.[2] presented an approach using two TL models, Inception_v3 and ResNet_50, to classify BTs images using three optimization tools: Stochastic Gradient Descent, Nesterov Accelerated Gradient, and Root Mean Squared Propagation. This approach shows that the combination of ResNet_50 together with the Adam optimizer provides superior performance in regard to the measures followed in this approach. However, this approach focused on combining the improvement level rather than aggregating the prediction level, making it less adaptive to changing data distributions. Also, this approach used internal integration at the optimization level rather than external integration at the output level, which made it less flexible compared to models that use weighted ensemble techniques for prediction, as in the proposed model.

Malakouti et al.[3] presented a study based on ML techniques and TL using a modified Google Net model. In this study, MRI images and digital data were used to classify individuals as healthy or diseased. These results indicate the potential of combining ML techniques with TL for the classification of BTs. The proposed TL model outperformed the other traditional ML techniques in terms of accuracy. This work illustrates the advantage of combining features from ML and DL based on a single underlying network. In contrast, the proposed WAEL model combines the outputs of multiple TL networks through a weighted combination strategy that statistically tends to reduce overfitting and generalizes better across MRI images.

Amarnath et al.[4] presented a method for the classification of BTs by TL techniques using MRI images. The techniques used consisted of CNNs: X_ception, VGG16, ResNet_50, EfficientNet_V2-S, and ResNet_152V2. Among all the networks used in this methodology, X_ception achieved the highest F1-score performance. A single technique was chosen in this methodology because it performed best, and a combination of techniques was not employed. Therefore, it did not have an ensemble mechanism in a structured way to take advantage of different techniques' complementary strengths, as achieved through the WAEL model.

Albalawi et al. [5] proposed a DL-based model for the BTS discrimination by MRI. In this work, a multi-layer CNN was designed to model intricate image patterns. This is that the MRI image features are abstracted by their images, then learned automatically to aid classification. This network resulted in accurate classification and generalization for several kinds of BTs. However, this model was based on one architecture only, while the WAEL model uses a systematic, complete set of various DL techniques that benefit more to improve not only accuracy but also generalization for the MRI images.

Alam et al. [6] proposed a study based on Swin_TransformerV2 to classify BTs from MRI images. Two main architectures were adopted in the study: the first is a Dual_Branch_Down_Sampling unit for image feature extraction and reducing their dimensions, and the second is an improved attention mechanism that enables the model to access the most important locations in MRI images. The results of this study show the improvement in performance metrics achieved by the proposed model. While this model has shown better feature localization, it involves intensive computational resources. In contrast, the WAEL model, which is more computationally balanced, integrates both CNN and TL models for comparable accuracy with less complexity.

Khoramipour et al. [7] introduced DL with Winter-tight multi-path CNN and SVM for better BTs classification in MRI images. This model is based on two CNN paths with three convolutional layers, which use 4, 8, and 16 filters to detect different features and forward them to an SVM classifier to classify the BTs. Even though the fusion of CNN and SVM introduces diversity, they lack probabilistic weighting techniques that can better composite multiple deep models, like the WAEL model, which provides a mathematical combination of outputs from multiple deep models with higher accuracy and stability.

Nassar et al. [8] proposed a study on BTs classification in two approaches. The first approach used three TL models: SERes_Net, Conv_NeXtBase, and Res_Net101V2. In the second approach, a Vision Transformer was utilized based on the AdamW optimizer. SERes_Net achieved higher accuracy compared to the other models used in this study. However, this study's analysis did not explore hybrid or ensemble combinations among these approaches. Therefore, while they achieved strong individual performance, these approaches lack the multi-model synergy provided by the proposed WAEL model.

Ullah et al. [9] used nine pre-trained TL classifiers to classify BTs: Inceptionresnet_v2, Xception, Inception_v3, Resnet_18, Shufflenet, Resnet_50, Densenet_201, Resnet_101, and Mobilenet_v2. The experimental outcomes showed that the Inceptionresnet_v2 algorithm is superior to other TL algorithms in the measures used in this method. Despite the breadth of comparison, their work still treated each model in isolation. The absence of a mechanism for model output integration or weighting leaves the potential of ensemble improvement unexplored. Although the comparison is extensive, this work treats each model separately due to the absence of a mechanism for merging model outputs, which is avoided in the proposed WAEL model.

Disci et al. [10] presented an approach to classify BTs in MRI images. TL was performed with the CNNs: Xception, Inception_v3, MobileNet_V2, VGG16, ResNet_50andDenseNet121. But in the approach, Xception was the best performing among the networks applied. The observations from this line of investigation support that CNNs differ over the dataset statistics, indicating the necessity for an ensemble integration, which is exactly what the proposed WAEL model improves.

Most related work focuses on the detection of BTs and DL and TL approaches. Many have been made to accomplish this task. There are some limitations in these works; they primarily depend on a single CNN architecture for classification. These architectures are strong on their own but weak at generalizing across heterogeneous MRI datasets. The majority of them do not have a systemic ensemble strategy utilizing the complementary properties among different models, which will decrease robustness and predictive performance. In this study, we fill the gap by proposing the WAEL model that combines VGG16, VGG19 and Inception_v3 according to performance-related weighting. Experimental results show that the WAEL model outperforms both individual models and previous methods and offers a trustworthy decision-making supplement for radiologists to assist in the diagnosis of BTs. It can be seen from Table 1 that the proposed model is the only one using the WAEL model for multiple pre-trained TL networks with better performance, and this difference sets it apart from other related works.

Table 1. Comparison of the proposed model with other related works.

Study (References)	TL Models Used	Optimization Strategy	Approach
Proposed Model	VGG16, VGG19, Inception_v3	Adam	WAEL Model
Srinivas et al. [1]	VGG16, ResNet_50, and Inception_v3	SGD	Single Model Evaluation
Ahmmed et al. [2]	Inception_v3, ResNet_50	SGD, RMSprop, NAG, Adam	Comparative Optimizer Evaluation
Malakouti et al. [3]	Modified GoogLe_Net	Not specified	Hybrid ML + TL
Amarnath et al. [4]	X_ception, VGG16, ResNet_50, etc.	Not specified	Single Model Evaluation
Albalawi et al. [5]	Multi-layer CNN	Not specified	Single Deep CNN Architecture
Alam et al. [6]	Swin_TransformerV2	Custom attention mechanism	Not applicable
Khoramipour et al. [7]	Multi-path CNN + SVM	Not specified	Hybrid CNN + SVM
Nassar et al. [8]	SERes_Net, Conv_NeXtBase, Res_Net101V2	AdamW	Comparison with Vision Transformer
Ullah et al. [9]	Inceptionresnet_v2, Xception, Inception_v3, etc.	Adam	Single Model Evaluation
Disci et al. [10]	Xception, Inception_v3, MobileNet_V2, etc.	Not specified	Single Model Evaluation

Materials and methods

This section describes the main ingredients that were used to design and test the approach. It starts with the MRI image dataset used, followed by details on how we augment the images to better serve the trained models. Next, we describe the use of TL to leverage pre-trained DL models. Finally, we present the proposed methodology, detailing the architectural design and implementation strategy.

Dataset

The Figshare BTs Dataset is a publicly available dataset containing a collection of preprocessed (MRI) scans. The dataset includes a total of 3,064 MRIs classified into three tumor types: meningioma, glioma, and pituitary tumor.[11] Images in the given dataset have dimensions of 512×512, and the dataset contains images from multiple datasets, which can be useful for multi-class classification. More details are shown in Fig. 1.

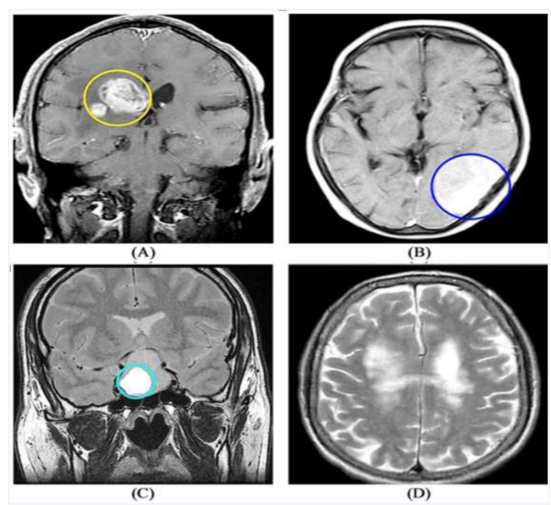

Fig. 1. Classification of the Dataset: (A) Glioma, (B) Meningioma, (C) Pituitary, and (D) No Tumor.

Image augmentation

Generally, image augmentation is a critical process in the preparation of image data to enhance its usability in tasks such as ML. This concerns a number of methods that are used on the raw images in order to enhance them, eliminate noise, and gain essential features from the input images to optimize the required model.[12] In the practice of categorizing BTs utilizing the MRI images, some of the filtrations or pre-processing techniques are employed as follows:

Resizing: For all the images in the dataset to be of the same dimension, resizing of these images is essential. The benefit of this technique is to approach and reduce computational operations, and to ensure that a model accepts inputs of equal dimensions.[13]

Normalization: This is done to place the pixel intensities in the range between 0 and 1 so as to boost the length, which is taken when training the model, as well as boost the efficiency of the model. It also has a positive effect insofar as it helps to get rid of the effects that result from within-screen pixel density differences.[14]

Noise Reduction: This technique is utilized on images to enhance the features extracted from the images. Some of the common denoising methods include median filtering, Gaussian smoothing, and denoising methods such as the bilateral filter.[15]

Contrast Enhancement: By using this technique, the visibility of the structures and the features of the image important in classification is enhanced. Histogram equalization and adaptive histogram equalization are the techniques employed for contrast enhancement.[16]

***Intensity Normalization*:** This technique helps standardize brightness and contrast levels to facilitate the process of identifying and classifying images with high accuracy for the model.[17]

***Data Augmentation*:** This technique helps to generate more training data by applying rotation, flip, and crop augmentations in terms of strength and dimensions. This technique also helps in improving the training set diversity to avoid overfitting in the model.[18]

***Registration*:** This technique is instrumental in enhancing the degree of homogeneity in extracted features, whereby images of different patients are aligned to a common reference frame, which is beneficial to the classification model.[19]

Rotation: This technique is used to change the orientation of an image by rotating it by a specific angle around a specific point to generate new images from the original image. The purpose of this technique is to enable the learning model to recognize a specific part of the image, even if it is at an unfamiliar angle, increasing the model's generalization power.[20]

Zoom Range: This technique is used to determine the amount of zoom in or zoom out of the image during the data training process, thus enabling the learning model to recognize patterns in the image even if they are larger or smaller than usual, thus increasing the model's ability to generalize to new images.[20]

Transfer learning

TL is one of the approaches for DL based on CNN networks, and it has been used recently in many fields of study due to its high accuracy. This approach is beneficial in reusing a model trained for a specific task on a new dataset, hence requiring less data and computations.[21] It has been repeatedly used in analyzing and classifying images using the CNN network, where the features required for image classification are extracted in order to enhance the accuracy and efficiency of classification without the need to specify additional features.[22] The CNN network consists of five main layers: the convolutional layer, the nonlinear layer, the pooling layer, the flattening layer, and the fully connected layer. CNN can recognize objects through certain features, such as corners, color spots, edges, etc. CNN uses filters to detect those edges and corners to interpret the images. In the convolutional layer, features such as corners and edges are extracted from the input image. The images are combined with filters to find feature maps. These feature maps are then transferred to the next convolutional layer to extract other features of the image. Eq. (1) below shows the structure of the two-dimensional convolution.[8]

$$S(x, y) = (I * K)(x, y) = \sum (m) \sum (n)\, I(x, y)\, K(x - m, y - n) \tag{1}$$

The second layer in this network is the non-linear layer, also called the activation layer. This layer is used to convert the network into a non-linear structure so that the network can learn faster. Then a pooling layer is used, and the main goal of this layer is to shrink the input size to preserve the most important features in the feature maps. Then a Flattening layer converts the data from the convolution and pooling layers into one-dimensional data to prepare that data for the fully connected layer. In the last fully connected layer, the input image is recognized and classified.

Various CNN architectures have achieved great success in improving image classification performance and have outperformed traditional ML methods in the past. In the proposed approach, three powerful CNN architectures are used to classify and identify the type of BTs using MRI images, namely VGG16, VGG19, and Inception_v3.

VGG16

VGG16 is the abbreviation for (Visual Geometry Group) and is widely applied in many fields due to its relatively simple structure and generalizability. This network achieves 92.7% top-5 test accuracy on the ImageNet dataset, which includes more than 14 million images belonging to 1000 categories. This network is named VGG16 because it contains 16 layers, 13 of which are convolutional layers, and 3 of which are fully connected layers. This network filters 224×224 RGB input images. This network uses 3×3 filters in the convolution layers in step 1. In step 2, a maximum of five pooling operations are performed with a window size of 2×2. In the last portion of this network, there are three connected layers containing 4096, 4096, and 1000 neurons, respectively, and the last layer in this network consists of a Softmax layer to determine the class to which the network outputs belong by producing a probability value between 0 and 1, and in all hidden layers in this network, the rectified linear unit (ReLU) is used as the activation function. The number of trainable parameters in the VGG16 network is approximately 138 million.[4,23]

VGG19

VGG19 is another model of a CNN. This network is similar to VGG16 in terms of architecture, except that it contains 19 layers. Thus, there are three additional convolutional layers in this network, of which sixteen are convolutional layers and three are fully connected layers. This network uses 3×3 filters in the convolution layers in step 1, as in the VGG16 network, and in step 2, pooling operations are performed with a maximum window size of 2×2. The three fully connected layers in this network consist of 4096, 4096, and 1000 neurons, respectively. The last layer in this network consists of a Softmax. ReLU is used as the activation function in all hidden layers. The number of trainable parameters in this network is about 143 million.[4,23]

Inception_v3

Inception_v3 is another model of a CNN for image classification tasks. This network represents the third version of the Inception model. It was pre-trained on the ImageNet dataset, making suitable for use in multiple applications with a reasonable degree of accuracy.[24] This network is a deep network consisting of a set of primitive units interconnected with each other. Each inception module merges different types of convolutional and pooling layers to extract features of the image. This network also applies a combination of 1×1, 3×3, and 5×5 convolutional filters to extract features of the image. When using 1×1 convolutional filters, the input dimensions are reduced, while when using 3×3 and 5×5 convolutional filters, more detailed features of the image are extracted. The initial layer in this network consists of three standard convolutional layers, followed by a max-pooling layer, two convolutional layers, and a max-pooling layer, then the last fully connected layer. The initial convolution of the network, by using a different filter size for each convolution, convolves the input at the same time to further merge or stack the results together and pass them through the network. Then, it executes a learning stop layer to randomly drop the weights to make the filter value equal to zero to prevent overfitting.[25]

Proposed methodology

This section explains the mechanism and pre-trained models utilized to build the proposed model. Since the research community has paid great attention to areas related to human health, this study proposes an effective method for detecting and classifying BTs. As shown in Fig. 2, the proposed model is

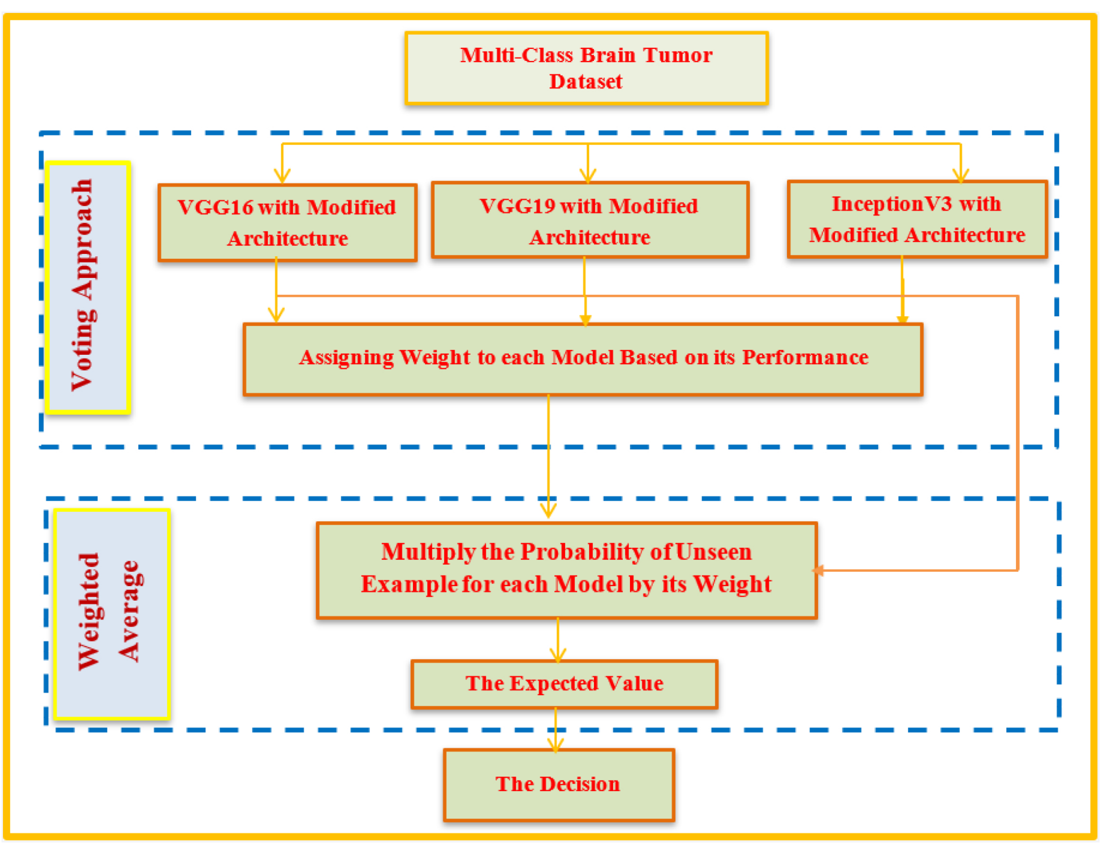

Fig. 2. The proposed system steps.

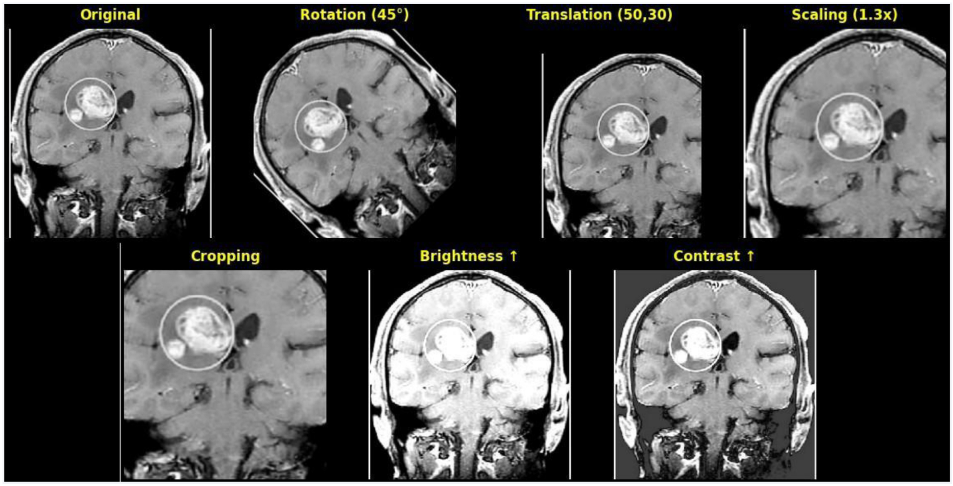

Fig. 3. Image augmentation.

achieved through three stages: image preprocessing, BTs classification, and model evaluation to obtain a decision.

Original images have been collected at (512 × 512) pixels; thus, in the preprocessing stage, these images were resized to (224 × 224) pixels, the size represented by the pre-trained model. Then, different transformations have been applied to the existing data to increase the size of a training dataset. These transformations include rotation, translation, scaling, cropping, and changing brightness and contrast. There are two main goals behind this step: 1) Enhance the brightness and contrast of the original image. 2) Train the proposed model on the largest number of images of different shapes. For more details, see Fig. 3.

The stage of classification and prediction for patients (cases) with BTs is the most important in this study. The proposed model is based on collecting the opinions of three important and effective models in TL in each case: VGG16, VGG19, and Inception_v3. The weight matrix of these pre-trained networks is used in their architecture. These weights were assigned to each model based on its individual predictive ability. Accuracy was used as the metric for assigning a weight or rank to the model. Therefore, the weighted ratio for each model was calculated by dividing its accuracy by the total (the accuracy of the model and the accuracy of the other two models). After that, each model was run individually on the dataset used in this study, which contains four categories: no tumor, meningioma, glioma, and pituitary tumor. Three layers were added to each pre-trained model in order to classify the input cases (input images), as follows: 1) dense layer (fully connected layer) with the ReLU activation function; 2) dropout layer to avoid overfitting problems; 3) classification layer with softmax activation function to classify each case. Fig. 4 shows the details of the proposed model.

The proposed model focuses on individual training for each model and weights these models based on their performance in classifiable tasks. The VGG16 model has achieved the best performance in classifying patients, followed by the VGG19 model and then the Inception_v3 model. Thus, the VGG16 model was weighted with a higher rank to have a greater role in the proposed model. On the other hand, Inception_v3 was given a lower rank compared to the rest because it achieved a lower performance in classification.

Evaluation metrics

The WAEL model was assessed using a confusion matrix. This matrix is defined as a table that is utilized to describe the performance of the model. A relative explanation of the evaluation metric steps is carried out, as shown in Eqs. (2) to (5). Through the confusion matrix, accuracy, precision, recall, and F1-score can be calculated. Accuracy indicates how many predictions were correct (e.g., True Positive (TP) and True Negative (TN)) compared to the overall number of instances. Precision is the ratio of TP to the total number of samples predicted as positive. Recall, or sensitivity, is the ratio of TP to the total number of actual positive samples. Finally, the F1-score is the harmonic mean of precision and recall.[1]

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TN} + \text{FN}} \tag{3}$$

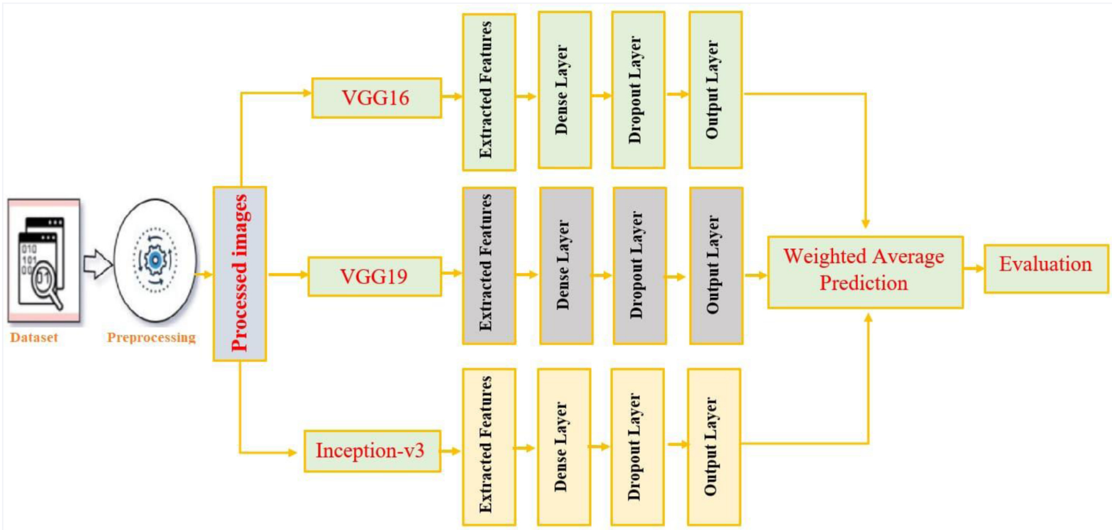

Fig. 4. The proposed model architecture.

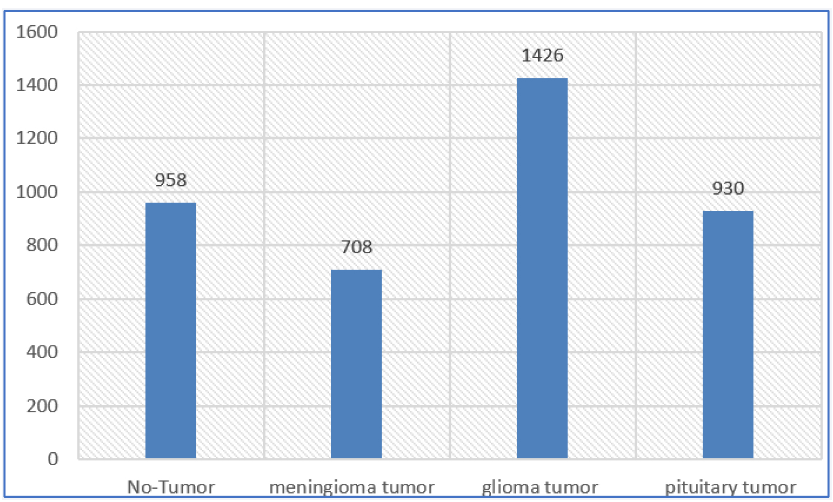

Fig. 5. Number of images in each class.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{F1} - \text{score} = \frac{2 \times TP}{2TP + FP + FN} \tag{5}$$

Result and discussion

The section discusses the performance of the WAEL model in BTs classification by conducting experiments on MR images. At first, the code is written in Python using TensorFlow and Keras, and experiments are performed on a Core i7 processor with 16 GB of RAM. The imaging dataset is utilized to measure the quality of the WAEL model. To enhance the diversity and representativeness of the dataset, additional preprocessing and expansion steps were applied before model training. The original Figshare BTs dataset lacked benign (no-tumor) MRI images; therefore, we supplemented it with publicly available benign cases from compatible MRI sources to balance the class distribution. This expansion increased the dataset to 4,022 images, categorized into glioma, meningioma, pituitary tumor, and no-tumor classes. The number of these categories in the dataset is shown in Fig. 5.

All images are pre-processed by rescaling them to 224*224. As shown in Fig. 6, the data were augmented using rotation, translation, scaling, cropping and alterations in brightness and contrast. These augmentations allow for enriching the training data with realistic transformations of existing samples, thereby avoiding the necessity to acquire additional measurements. The rotation and transformation operations

Fig. 6. Percentage of training and testing set.

make the model invariant to object orientation and location, while scaling and shearing allow the system to deal with size variations and partial visibility. Brightness and contrast normalization leads to increased robustness to varying lighting conditions and sensors. Overall, these optimizations not only reduce overfitting and improve feature learning capability but also allow the model to be more practical, working well in various realistic scenarios. Finally, the dataset was split into 70% for training and 30% for testing to ensure that the model had enough data to learn the underlying patterns while maintaining an adequate portion for evaluating unbiased performance on unseen samples.

Hyperparameters play an important role in model training convergence and non-fall into local minima. As such, the batch size is set as 32, and the learning rate is set as (0.0001). Finally, the number of epochs was determined by trial and error to be 20, whereas the dropout probability is 0.25.

To classify the patients with BTs, VGG16, VGG19, and Inception_v3 are applied. These pre-trained models were tested with the Adam optimizer to extract the relevant features. After the feature extraction, three layers have been added to the end of these models to classify the cases. The first layer is a fully connected layer (dense layer) with a ReLU activation function. The second layer is the dropout layer to avoid overfitting. The last layer is the output layer (classification layer), which is adapted to the softmax activation function for classification.

The classification is achieved by utilizing a WAEL model of the three aforementioned models. From those three models, three diverse feature spaces are extracted, which are then ensembled to improve the feature space. Three different weights, namely weight_1, weight_2 and weight_3, are assigned to three different models by dividing the accuracy of a model by the sum of the accuracies for all these models. This normalization ensures that the weights add up to 1, facilitating an easy interpretation of how much a given model contributes to the ensemble predictions. For example, if the validation accuracies of the three models are 0.75, 0.80, and 0.85, respectively, the corresponding weights would be 0.75 / (0.75 + 0.80 + 0.85) = 0.31 for the first model, 0.80 / (0.75 + 0.80 + 0.85) = 0.33 for the second model, and 0.85 / (0.75 + 0.80 + 0.85) = 0.35 for the third model. This weight distribution reflects the relative performance of the individual models, enabling the ensemble to leverage the strengths of each component model effectively. Thus, the proposed method is distinguished from other traditional ensemble methods (e.g., voting) in that it assigns a weight to each model based on its predictive ability on the testing set. This means that the model that performs better is given a higher ranking.

To prove the effectiveness of the WAEL model, it has been necessary to compare this model with individual CNN models (VGG16, VGG19, and Inception_v3). In addition, much work has been done on the MRI dataset. However, it is found that, compared to other works that utilized the same dataset, the proposed model is better in terms of all evaluation metrics, as shown in Table 2 and Figs. 7 and 8.

Table 2. Performance evaluation of the proposed model on the MRI dataset.

Model	Accuracy	Precision	Recall	F1-score
VGG16	0.9809	0.9819	0.9809	0.9809
VGG19	0.9786	0.9799	0.9786	0.9786
Inception_v3	0.9832	0.9836	0.9832	0.9832
Aamir et al. [11]	**98.95**	–	–	–
Proposed model	**0.9908**	**0.9910**	**0.9908**	**0.9908**

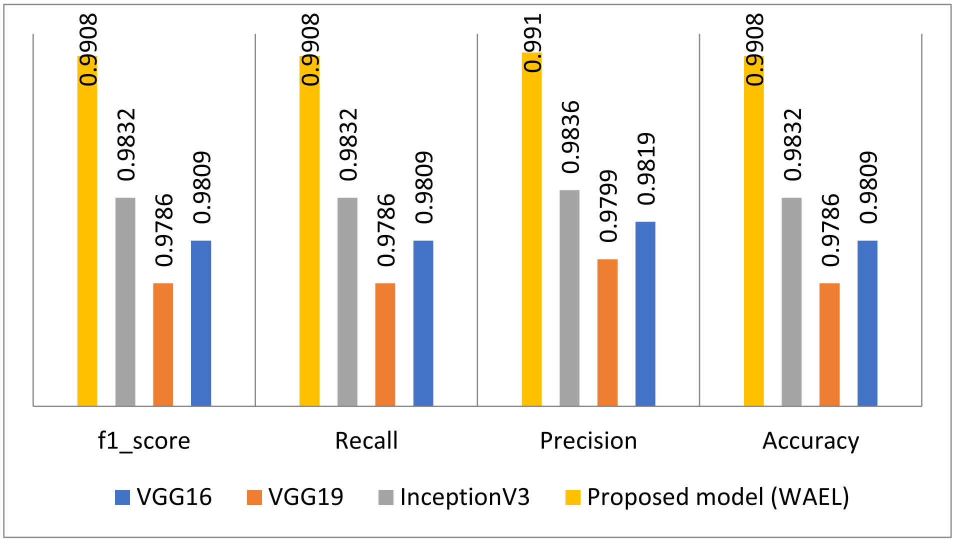

Fig. 7. The performance scores for proposed model and other deep CNNs.

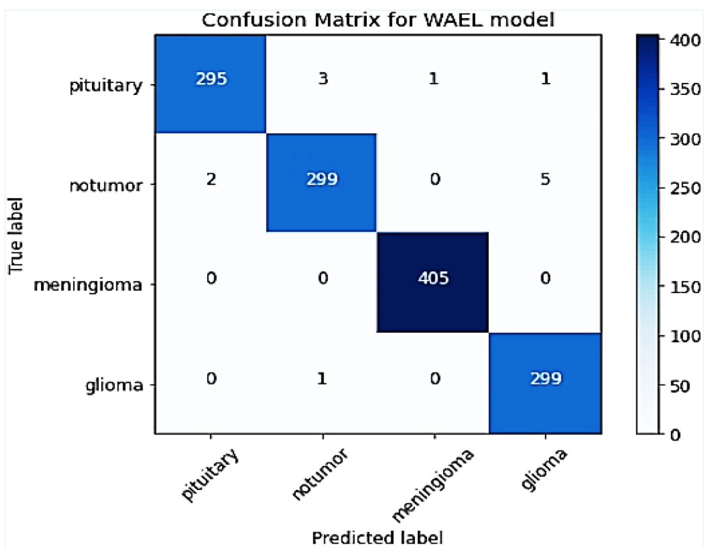

Fig. 8. The confusion matrix for the proposed model (WAEL).

The confusion matrix of the WAEL model, as illustrated in Fig. 8, demonstrates the performance of the classifier via four categories: pituitary tumor, no tumor, meningioma, and glioma. The proposed model shows strong overall classification ability, particularly in detecting meningioma (405 correct classifications) and glioma (299 correct classifications), both with zero false positives. However, some minor errors in classification were noted. For example, three pituitary tumor samples were misclassified as no tumor, and one as meningioma, while two no-tumor samples were misclassified as pituitary and five as glioma. These small mistakes are the result of some tumor tissue partially overlapping healthy tissue, a general problem in medical imaging. However, the WAEL model has great reliability in distinguishing different classes of BTs.

Recent advances in deep learning-based BTs classification have focused on hybrid transformer architectures, multimodal fusion, and lightweight CNNs for real-time diagnosis. For instance, Alam et al.[6] and Nassar et al.[8] have demonstrated that transformer-based networks can obtain high precision; however, in practical clinical applications, such models often require remarkable computational power and large annotated datasets. Contrary to this, the proposed WAEL model framework provides a balance between computational efficiency and diagnostic performance by employing weighted ensembles of several CNN models. Such a design enhances robustness to data variance and promotes consistent tumor delineation on any MRI modality. Being a reliable and interpretable ensemble system, it can be used as a second-opinion tool in a clinical setting to assist radiologists, which may decrease the workload of clinicians or improve early detection abilities. As a result, the WAEL model provides an attractive bridge between high-performance computational methods and practical tools suitable for routine clinical radiology.

Limitations

Even though the WAEL model achieved high performance and outperformed the single CNN classifiers as well as other related works, there are limitations to consider. First, the model was only trained and externally tested with one publicly available database, which may narrow its generalizability in different scenarios (other databases) or clinical settings. Second, the processed and balanced datasets do not express complete real-world situations such as imbalanced/noisy data. Finally, it is to be noted that while the model demonstrated encouraging accuracy in the classification of three tumor types (and differentiating these from benign pathologies), rare and atypical BTs are not accounted for by this model and may limit its clinical utility within a broader diagnostic context.

Conclusion

This study presented the WAEL model, a powerful TL-based model for automatic classification of BTs from MRI images. This model was realized by employing three popular CNN architectures. The experimental results presented in the study support and demonstrate the potential of such a model to classify BT types, using this data set as training and testing sets. In the AWEL, three models were used for decision making, namely VGG16, VGG19 and Inception_v3. Subsequently, models were weighted based on classification accuracy. The precision, accuracy, F1-score, and recall of the proposed model reached as high as 99.08%. The model's performance was compared with that of previous studies, which employed this data and achieved better accuracy in classification. These results suggested the promising use of the WAEL model as an adjunctive and reliable decision-assisting tool for radiologists, providing more prompt interventions.

In future work, we can design a novel model using powerful deep CNNs to classify BTs and further promote the field of MRI image analysis to support more efficient and accurate BT diagnosis. Another direction for future work would be to incorporate interpretation techniques such as (LIME or SHAP) in order to provide medical professionals with more transparent and interpretable predictions of the WAEL model.

Authors' declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any figures or images that are not produced by us have been included after citing their respective references, which are attached to the manuscript.
- No animal studies are present in the manuscript.
- Author(s) signed on ethical considerations' approval.
- Ethics Statement: This study used a publicly available brain tumor dataset that is freely accessible for research purposes. No human participants were directly recruited, and no identifiable personal data were collected by the authors.
- Ethical Clearance: The project was approved by the local ethical committee at University of Babylon.

Authors' contributions statement

LH.A, T.A.W. and H.A.A. suggested the idea by Conception, design, execute and drafting the manuscript,A.F.H. and A.K.A. take a part of analysis, interpretation of the results, revision and proofreading the manuscript, and the reviewers provided some suggestions which improved the quality of the work.

Data availability

The datasets analyzed during the current study are publicly available in the following repositories: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427.

References

1. Srinivas C, K. S. NP, Zakariah M, Alothaibi YA, Shaukat K, Partibane B, *et al.* Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images. M.A B, editor. J. Healthc. Eng. 2022 Mar 8;2022:1–17. http://dx.doi.org/10.1155/2022/3264367.
2. Ahmmed S, Podder P, Mondal M, Rahman S, Kannan S, Hasan M, *et al.* Enhancing Brain Tumor Classification with Transfer Learning across Multiple Classes: An In-Depth Analysis. BioMedInformatics. 2023 Dec 6;3(4):1124–44. http://dx.doi.org/10.3390/biomedinformatics3040068.
3. Matin Malakouti S, Bagher Menhaj M, Abolfazl Suratgar A. Machine learning and transfer learning techniques for

accurate brain tumor classification. Clin. eHealth. 2024 Dec;7:106–19. http://dx.doi.org/10.1016/j.ceh.2024.08.001.
4. Amarnath A, Al Bataineh A, Hansen JA. Transfer-Learning Approach for Enhanced Brain Tumor Classification in MRI Imaging. BioMedInformatics. 2024 Jul 22;4(3):1745–56. http://dx.doi.org/10.3390/biomedinformatics4030095.
5. Albalawi E, Thakur A, Dorai DR, Bhatia Khan S, Mahesh TR, Almusharraf A, *et al.* Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach. Front. Comput. Neurosci. 2024 Jun 12;18. http://dx.doi.org/10.3389/fncom.2024.1418546.
6. Alam N, Zhu Y, Shao J, Usman M, Fayaz M. A Novel Mearning Framework for Brain Tumor Classification Using Improved Swin Transformer_V2. ICCK Trans. Adv. Comput. Syst. 2025 Feb 10;2(1):1-0. https://dx.doi.org/10.62762/TACS.2025.807755.
7. Khoramipour S, Gandomkar M, Shakiba M. Enhancement of brain tumor classification from MRI images using multi-path convolutional neural network with SVM classifier. Biomed. Signal Process. Control. 2024 Jul;93:106117. http://dx.doi.org/10.1016/j.bspc.2024.106117.
8. Nassar SE, Yasser I, Amer HM, Mohamed MA. Comparative analysis of vision transformers and fine-tuned transfer learning models for brain tumor classification. Imaging. Radiat. Res. 2024 Nov 11;7(1):8514. http://dx.doi.org/10.24294/irr8514.
9. Ullah N, Khan JA, Khan MS, Khan W, Hassan I, Obayya M, *et al.* An Effective Approach to Detect and Identify Brain Tumors Using Transfer Learning. Appl. Sci. 2022 Jun 2;12(11):5645. http://dx.doi.org/10.3390/app12115645.
10. Disci R, Gurcan F, Soylu A. Advanced Brain Tumor Classification in MR Images Using Transfer Learning and Pre-Trained Deep CNN Models. Cancers. 2025 Jan 2;17(1):121. http://dx.doi.org/10.3390/cancers17010121.
11. Aamir M, Rahman Z, Dayo ZA, Abro WA, Uddin MI, Khan I, Imran AS, Ali Z, Ishfaq M, Guan Y, Hu Z. A deep learning approach for brain tumor classification using MRI images. Comput. Electr. Eng. 2022 Jul 1;101:108105. https://doi.org/10.1016/j.compeleceng.2022.108105.
12. Raheem AK, Dhannoon BN. Deep Learning based Models for Drug-Target Interactions. Baghdad Sci. J. 2024 Nov 1;21(11):3605-16. http://dx.doi.org/10.21123/bsj.2024.9212.
13. Musthafa MM, Manimozhi I, Mahesh TR, Guluwadi S. Optimizing double-layered convolutional neural networks for efficient lung cancer classification through hyperparameter optimization and advanced image pre-processing techniques. BMC Med Inform Decis Mak. 2024 May 27;24(1):142. https://doi.org/10.1186/s12911-024-02553-9.
14. Al-Mashanji AK, Alhasnawy LH, Hamza AH. Robust attribute selection to improve the Parkinson's disease classification: a hybrid approach. Bull. Electr. Eng. Inform. 2023 Dec 1;12(6):3365–73. http://dx.doi.org/10.11591/eei.v12i6.5247.
15. Zhang S, Liu C, Zhang Y, Liu S, Wang X. Multi-Scale Feature Learning Convolutional Neural Network for Image Denoising. Sensors. 2023 Sep 6;23(18):7713. http://dx.doi.org/10.3390/s23187713.
16. Prince R, Niu Z, Khan ZY, Emmanuel M, Patrick N. COVID-19 detection from chest X-ray images using CLAHE-YCrCb, LBP, and machine learning algorithms. BMC Bioinformatics. 2024 Jan 17;25(1):28. https://doi.org/10.1186/s12859-023-05427-5.
17. Senguttuvan D, Pichai S. Mammogram image preprocessing using intensity range based partitioned cumulative distribution function. J. Anal. 2022 Oct 26;31(3):1639–56. http://dx.doi.org/10.1007/s41478-022-00499-7.
18. Alomar K, Aysel HI, Cai X. Data augmentation in classification and segmentation: A survey and new strategies. J. Imaging. 2023 Feb 17;9(2):46. https://doi.org/10.3390/jimaging9020046.
19. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2021;54(3):2033–2044. https://doi.org/10.1016/j.neuroimage.2020.117749.
20. 1.Xu M, Yoon S, Fuentes A, Park DS. A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. Pattern Recognit. 2023 May;137:109347. http://dx.doi.org/10.1016/j.patcog.2023.109347.
21. Wotaifi TA, Dhannoon BN. Attention Mechanism Based on a Pre-trained Model for Improving Arabic Fake News Predictions. Iraqi J. Sci. 2023 Nov 30;64(11):6041–6054. http://dx.doi.org/10.24996/ijs.2023.64.11.45.
22. Minarno AE, Aripa L, Azhar Y, Munarko Y. Classification of Malaria Cell Image using Inception-V3 Architecture. JOIV: Int. J. Informat. Vis. 2023 May 5;7(2):273. http://dx.doi.org/10.30630/joiv.7.2.1301.
23. Shah SR, Qadri S, Bibi H, Shah SMW, Sharif MI, Marinello F. Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease. Agronomy. 2023 Jun 18;13(6):1633. http://dx.doi.org/10.3390/agronomy13061633.
24. Mahmood AA, Sadeq S, Aljanabi YI, Sabry AH. Developing a convolutional neural network for classifying tumor images using Inception v3. East.-Eur. J. Enterp. Technol. 2023 Jun 5;123(9):86–93. http://dx.doi.org/10.15587/1729-4061.2023.281227.
25. Ahamed J, Ahmed M, Afreen N, Ahmed M, Sameer M. An Inception V3 Approach for Malware Classification Using Machine Learning and Transfer Learning. SSRN Electron. J. 2022. http://dx.doi.org/10.2139/ssrn.4233090.

نموذج التعلم الجماعي المتوسط المرجح باستخدام صور التصوير بالرنين المغناطيسي لتحسين تصنيف أورام الدماغ

ليث حامد حمزة، حسنين علي عبيس، تحسين علي حسن، عباس فاضل حمادي، امير علي كاظم

رئاسة جامعة بابل، جامعة بابل، بابل، العراق.

الخلاصة

يُعدّ اكتشاف أورام الدماغ (BTs) ذا فائدة كبيرة في تحديد نمو الخلايا الخطيرة في الدماغ التي تشكّل تهديداً لحياة المريض. ولا يزال الاكتشاف المبكر لأورام الدماغ تحدياً مهماً بسبب الفروقات الدقيقة بين أنواع الأورام وتعقيد البنية التشريحية للدماغ. وغالباً ما يكون التشخيص اليدوي من قبل أطباء الأشعة مستهلكاً للوقت ومعرضاً للتباين، مما قد يؤدي إلى تأخير العلاج. ومن بين تقنيات التصوير المتاحة، يُستخدم التصوير بالرنين المغناطيسي (MRI) على نطاق واسع بسبب قدرته العالية على إنتاج صور ذات دقة عالية. وقد حقق التعلم العميق (DL) نجاحاً ملحوظاً في مهام تصنيف الصور من خلال التعلم بالنقل (TL)، الذي يعتمد على استخدام نماذج مدرّبة مسبقاً. لذلك تقترح هذه الدراسة نموذجاً جديداً يسمى نموذج التعلم الجماعي المتوسط المرجح (WAEL) بالاعتماد على التعلم بالنقل لتصنيف أورام الدماغ. وقد تم استخدام معماريات الشبكات العصبية الالتفافية (CNN) مثل VGG16 وVGG19 وInception_v3 كمصنّفات داخل نموذج WAEL للتنبؤ التلقائي بأورام الدماغ. تم تدريب النموذج باستخدام مجموعة بيانات Figshare الخاصة بأورام الدماغ، والتي تحتوي على 3064 صورة من ثلاث فئات من الأورام (السحائي، الورم الدبقي، وأورام الغدة النخامية). ولتحسين القدرة على التعميم، تم إدراج صور دماغ سليمة (بدون ورم) ضمن مجموعة البيانات. أظهرت النتائج التجريبية أداءً متفوقاً، حيث حقق نموذج WAEL دقة بلغت 0.9908 مقارنةً بـ 0.9809 و0.9786 و0.9832 لكل من VGG16 وVGG19 وInception_v3 على التوالي. كما حقق النموذج قيم دقة واسترجاع ودرجة F1 مرتفعة تؤكد قوته. وقد تفوق نموذج WAEL أيضاً على الدراسات السابقة التي استخدمت نفس مجموعة البيانات. وبذلك يمكن اعتباره أداة مساعدة ثانوية لأطباء الأشعة لاكتشاف الأورام من صور الرنين المغناطيسي للدماغ.

الكلمات المفتاحية: تقنيات الأشجار، الشبكات العصبية التلافيفية، التعلم العميق، Inception_v3، التصوير بالرنين المغناطيسي، التعلم الآلي، التعلم الخطي، WAEL، VGG19، VGG16.