

## Design of a Predictive Modeling System for MRI Brain Image Classification

Gnanaprakasam Sriram<sup>ID</sup>, Rajasekaran Praveena<sup>\*ID</sup>

Department of Electronics and Communication Engineering, Muthayammal Engineering College, Rasipuram 637408, India

Corresponding Author Email: [praveena.r.ece@mec.edu.in](mailto:praveena.r.ece@mec.edu.in)



Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410642>

### ABSTRACT

**Received:** 3 November 2023

**Revised:** 25 September 2024

**Accepted:** 31 October 2024

**Available online:** 31 December 2024

#### **Keywords:**

*image classification, brain cancer, deep learning, machine learning, medical image diagnosis*

An accurate identification and characterization of any potential abnormalities in the MRI brain image is required for efficient classification of brain cancer. Recent advancements by computerized diagnosis eradicate the possibility of radiologists making an incorrect diagnosis based on their skills to perceive and interpret data. This work proposed an efficient deep learning based Predictive Modeling System for Brain Cancer Classification (PMS-BCC). It consists of three modules: feature extraction module, feature selection module and classification module. To localize the patterns for brain cancer, PMS-BCC uses stacked convolution and pooling layers along with random skip connections. The main advantage of using this combination is that it allows the model to learn hierarchical and spatially invariant features efficiently, while also addressing common issues like vanishing gradients and overfitting. In the feature extraction module, the convolution layer is responsible for the extraction of locally relevant features, while the pooling layers minimize the feature dimension. In the subsequent module, an AntLion Optimization (ALO) is used to choose the optimal subset of features, and a neural network using Greedy Layerwise Training (GLT) is used in the classification module to do the classification. The gradient based optimization techniques suffer from local minima. ALO does not depend on gradients and can explore search space more effectively to avoid local minima compared to gradient based techniques and its variations. ALO has a better exploitation-exploration balance due to its specific hunting strategy than other metaheuristic algorithms. Results showed that the proposed PMS-BCC architecture with GLT improves the classification accuracy from 93.1% (convention training) to 98.6% for binary classification (normal/abnormal) and 93% (conventional training) to 98.2% for multiclass (normal/low-grade/high-grade) classification of MRI brain images obtained from REpository of Molecular BRAin Neoplasia DaTa (REMBRANDT) database. Though the proposed PMS-BCC system provides promising results, only the axial views of brain in the REMBRANDT database is used for the analysis.

## 1. INTRODUCTION

The primary organ of the human body is the brain, and it is still unclear what causes brain cancer. Consequently, an early diagnosis is necessary to lower the fatality rate from brain cancer. One diagnostic method utilized in the medical profession that does not involve segmentation is image classification. It is regarded to be one of the essential tasks for computer vision. It is accompanied by supervised learning, which is a method of learning in which unstructured data, such as images, are categorized into predetermined classes (also known as labels), which are accessible throughout the training process. Image classification methods have traditionally been based on humanly built features, which need a high degree of domain expertise while displaying poor cross-domain adaptability. More recently, automated methods have begun to replace these manual methods. MRI brain images have several challenges which makes the conventional machine learning techniques less effective. They are summarized as below:

- High Dimensionality: MRI scans are high-resolution

resulting in a substantial quantity of data for each sample.

- Complex features: The abnormalities are often subtle and may vary from patient to patient.

- Intra-class Variability: Different patients with the same condition can have variable degrees of abnormalities in MRIs, which makes classification more difficult.

- Inter-class Similarity: Due to the similar visual appearances of healthy brain tissue and abnormal tissue structures, the classification task is very challenging. In recent years, Deep Learning (DL) has been used to solve complex issues that are found in the real world. The abovementioned challenges are addressed by the DL and are as follows:

- Automatic Feature Extraction: Feature extraction by traditional machine learning algorithms is subjective and incomplete, particularly when dealing with complicated data such as MRI images. The ability of DL architecture to automatically learn hierarchical features directly from raw images is a significant advantage.

- Handling High-Dimensional Data: DL models can process three-dimensional data efficiently, using CNNs or

autoencoders, to learn spatial relationships within volumetric data.

- Generalizing Across Intra-Class Variability: DL models have the ability at generalizing across big datasets, learning robust features that can account for intra-class variability.

- Discriminating between Inter-class Similarity: DL models learn highly discriminative features, which means that they can discover small distinctions across similar classes of raw images.

DL is a machine learning method that processes information non-linearly through hierarchical structures [1]. Its goal is to identify patterns and extract features from data. DL utilizes Artificial Neural Networks (ANN) [2] to mimic human brain functions and learn complex characteristics from input data. The proposed system applies DL for both binary (normal/abnormal) and multiclass (normal/low-grade/high-grade) classifications in brain image analysis, using non-invasive Magnetic Resonance Imaging (MRI) [3].

The main objective is to create a classification system for brain cancer using MRI scans with high accuracy and capable of early detection. The rest of the paper is as follows. Section 2 reviews the related work of MRI brain image classification models and methods which utilize DL. The development of a classification technique to classify MRI brain images is covered in Section 3. In section 4, it is demonstrated that the proposed PMS-BCC system is the most accurate system for classifying brain cancer. The key components of the suggested system are outlined in the last section.

## 2. RELATED WORKS

An automated brain image analysis is described in study [4] using MRI images of the brain. When it comes to providing an accurate early diagnosis, the classification of MRI scans as normal, low grade, or high grade is of critical importance. To extract features, the non-sub-sampled Shear let transform, which is superior to standard wavelet transforms is employed. Due to the number of shearing parameters, orientations, and scales, it is more computationally intensive than wavelets and often results in highly redundant representations of data.

An MRI brain image classification which consists of many phases, such as pre-processing MRI images, segmenting images, extracting features, and classifying images are discussed in study [5]. An adaptive filter is utilized to eliminate background noise during the pre-processing stage of an MRI scan. Both the local-binary grey level co-occurrence matrix and enhanced fuzzy c-means clustering are used for extracting features from the clustered samples. However, appropriate number of clusters must be specified for fuzzy c-means algorithm. Convolutional recurrent neural networks are used for the classification.

A variant of the VGG-16 architecture is discussed in study [6] for brain image classification. To reduce the effects of overfitting and reduces the number of hyperparameters required, the VGG-16 variation adds dropout layer and activating it with SoftMax. The convolution filters of varying sizes are layered to abstract features from the input data. It may be difficult for weight updates and can cause overfitting. Thus, Inception Module (IM) is designed which is wider rather than deeper to overcome the limitations. Using MRI, a Deep Neural Network (DNN) with a Pyramid Design of Inception Module (PDIM) is trained for the objective of brain image classification in study [7]. PDIM units are stacked on top of

one another to provide more depth in the architecture, and their respective performances are analyzed. The stacking approach requires more computational power and a time-consuming process during training compared to a conventional network.

An efficient method for multiclass classification of brain images via deep feature fusion is discussed in study [8]. MR images have been pre-processed using min-max normalization, and then considerable data augmentation is employed to the MRI images to overcome the problem of a lack of data. Features obtained from transfer learned architectures such as AlexNet, GoogleNet, and ResNet18 are fused to produce a single feature vector, which is then fed into Support Vector Machine (SVM) and K-nearest neighbor (KNN) to predict the final output. Due to feature fusion, the performance can be affected by the redundant features from different feature sets. It is very difficult to find optimal parameters of SVM and KNN requires explicit training for the classification. Techniques for image classification, especially using DNN have shown accurate results for MRI brain image classification is described in study [9].

An automated method for classifying brain images is described in study [10]. First, a  $5 \times 5$  Gaussian filter is used to the brain MR images to do the pre-processing. In the second step of the process, a deep feature extraction is carried out with the assistance of the Alex Net and VGG16 models. A combination of the acquired feature vectors is performed. The Extreme Learning Machines (ELM) classifier is used for the study of MR classification of images using these feature vectors. As ELM has a single hidden layer, it may not perform well with more complex problems, and it tends to overfit when handling smaller datasets or noisy data. The convolutional dictionary learning with local constraint algorithm to classify MRI of brain is discussed in study [11]. It integrates multi-layer dictionary learning into a convolutional neural network (CNN) structure. Encoding a vector on a dictionary can be considered as multiple projections into new spaces, and the obtained coding vector is sparse. The local constraint of atoms is generated using k-nearest neighbour graph. This ensures that the discrimination of the dictionary that is ultimately produced is of a high quality. However, choosing an appropriate sparsity level is very challenging and involves non-convex optimization. Convolutional neural network (CNN) is used in study [12] for multi-classification of brain images for the early diagnosis. Three distinct CNN models are designed which can classify brain images into three different classes, which are Grade II, Grade III, and Grade IV separately.

The non-invasive diagnostic assistance system discussed in study [13] is regarded as an image classification system that classifies as either normal or abnormal. DL has the capacity to enable the use of a single model for both the extraction of features and the classification of data, while rational methods need the use of separate models. Using transfer learning, an effective modified VGG architecture for brain the classification of images is developed. To improve the classification capabilities of the VGG architecture, several changes have been made to the pooling layer.

## 3. MATERIALS AND METHODS

This section discusses the proposed PMS-BCC architecture for MRI brain image analysis. Figure 1 presents the workflow of the proposed system that takes an MRI brain image as input. The proposed PMS-BCC architecture uses a DL concept

which employs hierarchical structures and several steps of nonlinear information processing to classify patterns and learn features. To localize the image features, stacked convolution and pooling layers with random skip connections are employed. The convolution layer extracts localized features whereas pooling layers reduces feature dimension in the feature extraction module. In the next module the best subset of features is selected by ALO and finally neural network with

GLT is employed for the classification.

### 3.1 Feature extraction module

Figure 2 shows the feature extraction module of the proposed PMS-BCC architecture. It shows the structural details of the arrangement of convolution and pooling layers along with other modules of the proposed system.

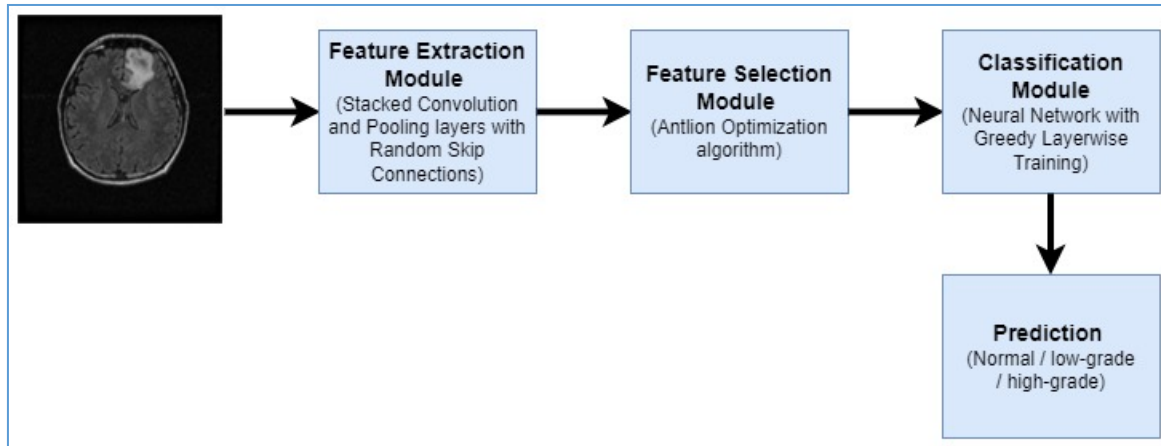


Figure 1. Proposed ISM-DR classification system

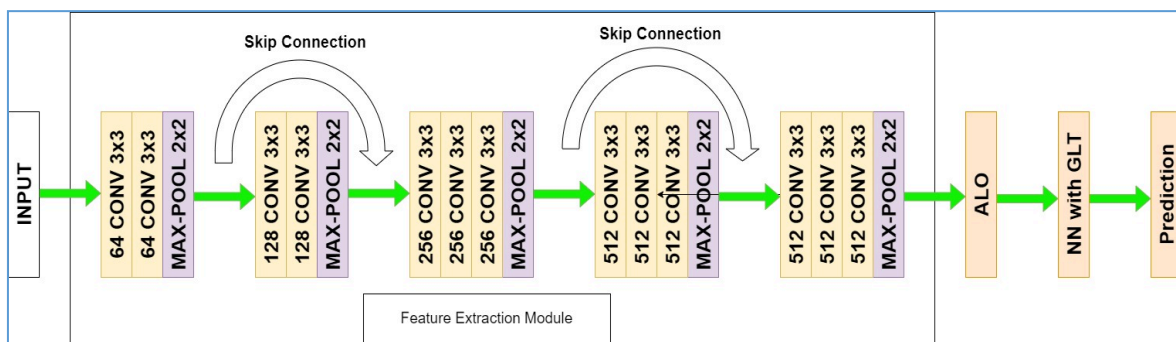


Figure 2. Feature extraction module of the proposed PMS-BCC architecture

The proposed PMS-BCC architecture combines the arrangement of convolution and pooling layers to localize image features from VGG16 architecture [14] with residual mapping from ResNet design [15]. It organizes its layers into five distinct blocks. Each block has two or three 3x3 convolutional layers that are stacked one after the other, followed by a 2x2 max-pooling layer with a stride of 2. This arrangement can gradually capture and understand the complex patterns in the brain images while compressing the spatial dimensions. As the blocks get deeper and deeper, this hierarchical feature extraction is important to the network's success in MRI brain image classification. However, adding more layers (i.e., making structures deeper) results in a vanishing gradient problem, making it harder and harder to train the architecture and causing its accuracy to saturate and then degrade. When the deep networks are trained using gradient-based learning approaches, the vanishing gradient issue arises.

To address the problem of vanishing gradient, Microsoft Research in 2015 [15] introduced the concept of skip connection between the layers. Using the skip connection, training may be skipped for a few levels and linked straight to the output. Hence, the network fits the residual mapping rather than having the layers learn the underlying mapping. If any

layer causes the model's performance to suffer in any way, the skip connection will allow regularization to work around the problem. This is one of the advantages of incorporating this kind of connection. Because of this, networks may be trained without the problems that are often produced by vanishing gradients.

### 3.2 Feature selection module

In this module, ALO is employed which is an artificial intelligence-based feature selection algorithm. The hunting behavior of antlions, which are insects that are often found in sandy areas, was the idea for the ALO method [16]. These insects are famous for their clever approach to seeking their prey. Antlions dig conical holes in sandy areas and wait at the bottom of the holes for unsuspecting insects to fall in. Simulating the hunting behavior of antlions is employed in this work to obtain the best feature subset for brain image classification. The random walk of ants is defined by:

$$X(t) = [0, \text{cumsum}(2\gamma(t_1 - 1)), \text{cumsum}(2\gamma(t_2 - 1)), \dots, \text{cumsum}(2\gamma(t_T - 1))] \quad (1)$$

where,  $t$  represents the iteration number,  $T$  is the maximum number of iterations, and  $\gamma$  is defined by:

$$\gamma(t) = \begin{cases} 1 & \text{if rand} > 0.5 \\ 0 & \text{if rand} \leq 0.5 \end{cases} \quad (2)$$

In Eq. (2), rand is a random number uniformly distributed in (0,1). The position of the ants needs to be regulated so that the ants do not violate the boundaries and move in the search space randomly. The regulation is defined by:

$$x_i^t = \frac{(x_i^t - a_i) \times (ub_i^t - lb_i^t)}{(b_i^t - a_i)} + lb_i^t \quad (3)$$

where,  $ub_i^t$  and  $lb_i^t$  are the maximum and minimum value of certain ant in the  $t^{\text{th}}$  generation.  $b_i$  and  $a_i$  are the maximum and minimum value of the  $i^{\text{th}}$  variable of all ants. The calculation for the ants that fall into the trap are given by:

$$lb_i^t = \text{Antlion}_j^t + lb^t \quad (4)$$

$$ub_i^t = \text{Antlion}_j^t + ub^t \quad (5)$$

To make the ants move towards the antlions, the ant's random walk range is lowered adaptively by:

$$lb_i^t = \frac{lb_i^t}{1 + 10^{\frac{\nu \times t}{T}}} \quad (6)$$

$$ub_i^t = \frac{ub_i^t}{1 + 10^{\frac{\nu \times t}{T}}} \quad (7)$$

where,  $\nu$  depends on the iterations and given by:

$$t > \begin{cases} 0.1T & \nu = 2 \\ 0.5T & \nu = 2 \\ 0.75T & \nu = 2 \\ 0.9T & \nu = 2 \\ 0.95T & \nu = 2 \end{cases}$$

The position of the ants is updated using the following formula:

$$\text{Ant}_i^t = \frac{R_i^t + R_E^t}{2} \quad (8)$$

where,  $R_i^t$  and  $R_E^t$  are position of the ants by route wheel selection and random position around the elite antlion respectively at the  $t^{\text{th}}$  iteration. The position of the ants is updated and replace an antlion with its corresponding ant if it satisfies Eq. (9).

$$\text{Antlion}_j^t = \text{Ant}_i^t \text{ if } f(\text{Ant}_i^t) > f(\text{Antlion}_j^t) \quad (9)$$

The new value of the ant's objective function is calculated and compared to the value of the elite. If the ant's fitness value is lower than that of the antlions, the antlions are thought to

have caught the ant, and the position of the antlions is updated. The objective function for the proposed PMS-BCC architecture is defined by:

$$f(X) = w \times CE(X) + (1-w) \times \frac{|X_n|}{N} \quad (10)$$

where,  $f(X)$  is the fitness function,  $N$  is the total number of features (population size),  $CE(X_i)$  is the classification error rate using the subset of  $X$ ,  $X_n$  is the number of selected features and  $w=0.8$  is a constant controlling the classification performance to the number of features used. The hyperparameter tuning by ALO is outlined below.

---

#### Algorithm:

---

Input: Features from the feature extraction module

Output: Selected sub-set of features

1. Randomly initialize the initial number of ants and antlions
  2. Utilizing Eq. (10), determine the fitness of ants and antlions
  3. If the end criterion is not met, identify the top antlions and take them for granted as the elite (decided optimal)
  4. for every ant
    - Use the roulette wheel to choose an antlion for each ant.
    - Update the ants walking range
    - make a random walk and normalize it
    - Adjust the position of ant
  - end for
  5. Estimate each ant's fitness
  6. If the Eq. (9) is met, replace an antlion with the appropriate ant
  7. If an antlion gets more fit than the elite, update the elite
  8. If the end requirements are met, return the elite
- 

### 3.3 Classification module

Figure 3 shows the structure of a neural network which executes data transformation-based mapping from input to output. These transformations are learnt using a variety of input training samples and the learnable parameters are weights and bias.

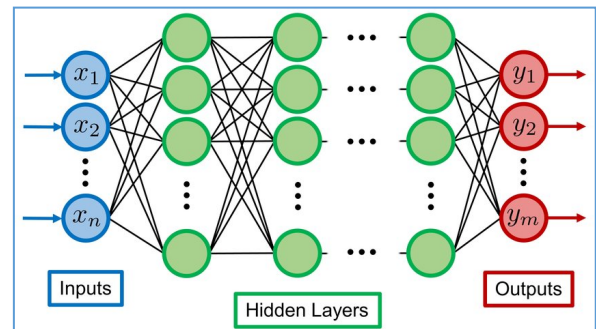


Figure 3. Structure of neural network

The proposed system randomizes the weight and bias parameters before learning starts using He initialization. For 'n' input values, this approach draws values from a distribution with 0 (zero) mean and variance. It maintains the variance of activations across the network which is a requirement for the rectified linear activation function at hidden layers. The training process is repeated until the needed

values and the desired output are attained for both parameters. The two parameters have different effects on the input data. Bias is the difference between the intended and expected values and is responsible for the difference between the network's actual and intended output. A low bias value means that the network generates fewer output form assumptions, while a high bias value means that the network generates more output form assumptions. High bias models are unable to accurately represent the salient characteristics of dataset samples and are ineffective when applied to fresh data. Weights of the networks may be thought of as the degree of neural connectivity. The weight of an input variable determines how much of an impact it has on the output. While a high weight number will have a bigger influence on the output, a low weight value will have minimal impact on the input. A single neuron in the neural network is defined as:

$$\hat{y} = f\left(\sum_{k=1}^M x_k w_k + b\right) \quad (11)$$

where,  $f$  represents the activation function,  $M$  represents the input samples,  $w$  and  $x$  denoted the weights and inputs to the neuron with bias ( $b$ ). In this work, rectified linear unit is employed as  $f$ . It might be challenging to train neural networks that are extremely deep and contain many layers.

When the number of hidden layers is raised, the amount of inaccurate information that is sent back to earlier levels is dramatically reduced. This shows that although hidden layers near the input layer have scarcely had any changes, layers near the output layer have undergone the necessary updates. This problem, known as the vanishing gradient, made it difficult to train exceedingly deep neural networks. An important turning point in the revival of neural networks was the development of the greedy layer-wise pre-training strategy, which is more often referred to as simply pre-training. This approach is originally accountable for enabling the building of deeper neural network models

Pre-training entails modifying a model while maintaining the weights, adding a new hidden layer, and letting it learn from the prior layer. To overcome the more challenging problem of training a deep network, the model is trained layer by layer. The whole training procedure is divided into a series of layer-wise training using a greedy shortcut to get locally optimal answers [17]. This quick cut results in a workable, worldwide solution. A shallow network may be trained more rapidly than a deep network using pre-training, which employs a layer-wise training method. Shallow networks are thought to be simpler to train than deep networks.

Pre-training may be approached in two main ways: supervised and unsupervised. A model trained via supervised learning is then put through supervised pre-training, which involves progressively introducing hidden layers. The greedy layer-wise method is used in unsupervised pre-training to build an unsupervised auto-encoder model. A supervised output layer is added after unsupervised pre-training. Unsupervised pre-training may be the optimal option if more unlabeled cases are used to initialize the model. This is so that unsupervised pre-training may take place without any human supervision. It is customary to alter the network's weights after the preceding layer is included, even though the weights in the layers underneath the final one are always maintained constant. During training, Mean Square Error (MSE) is used as an objective function which is defined by:

$$LossFunction(MSE) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

where,  $\hat{y}$  is the network's output and  $y$  is the ground truth data. Table 1 shows the parameter setting for the proposed PMS-BCC architecture.

**Table 1.** Parameter setting for the proposed PMS-BCC architecture

Parameters	Setting
Epochs	500
Learning rate	0.1
Activation function (Hidden layer)	Rectified Linear
Prediction function (Output layer)	SoftMax
Momentum	0.9
Training	Gradient descent and GLT
Loss function	MSE

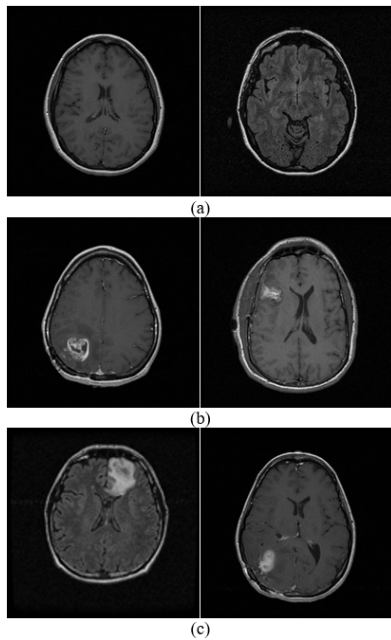
At first, the learning rate is set to 0.1. It is reduced by a factor of 10 (0.01, 0.001, 0.0001) when the validation accuracy stopped improving. It is observed that the learning rate is reduced 4 times during training of PMS-BCC architecture and the learning of deep features for the classification of brain images stopped at 245 epochs. There are many activation functions employed in DL architecture for computer vision applications. The commonly used hidden layer's activation functions are rectified linear, leaky and parametric rectified linear, tanh, exponentially linear unit. The rectified linear function allows only the positive values and rejects all negative values to avoid vanishing gradient problem. Among different functions, rectified linear is widely used due to its low complexity and simplicity. If any dying problem (neurons outputting zero for all inputs) due to rectified linear, the modified rectified linear functions can be used. The proposed PMS-BCC architecture uses rectified linear at hidden layer and SoftMax layer at the output layer as the system is a multiclass classification system.

#### 4. RESULTS AND DISCUSSIONS

The performances of the PMS-BCC architecture for brain image classification are evaluated by using REpository of Molecular BRAin Neoplasia DaTa (REMBRANDT) [18-20]. It has MRI scans of the brains of 130 participants. Every MRI scan has a resolution of 256 by 256 pixels and is saved in the DICOM format. As the REMBRANDT database has an enormous collection of images, the selection of images is crucial for the system's performance. The proposed PMS-BCC architecture uses 200 images which was carefully selected for MRI brain image classification in study [21]. The high quality and clarity of the selected images are well annotated and have traces of tumours. The proposed system considers the same set of images for the performance evaluation. Figure 4 shows samples from each category.

As DL models include a significant number of parameters, it is necessary to learn these characteristics from the data to recognize a wide variety of patterns and features present in the MRI scans. Thus, deep learning models need large amounts of training data to generalize successfully to testing data. The selection of training images for each class is mostly determined by the availability and inherent distribution of the classes within the dataset. The objective is to guarantee that

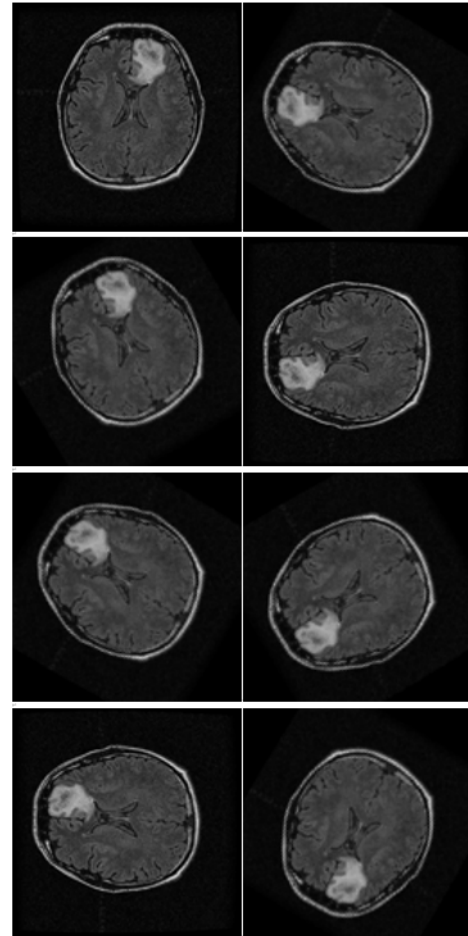
every class had a representative sample size to train the model efficiently without introducing substantial bias. The training process becomes more consistent when the dataset is more evenly distributed. It reduces the chances of the model experiencing substantial variations in its learning patterns during training. To balance the number of images, data augmentation is used with flipping and rotation in multiples of 30 degrees. Figure 5 shows the data augmented images. For binary classification, 2400 normal images are generated by data augmentation and 2400 abnormal images (1200 low-grade and 1200 high-grade) images are generated. To avoid class imbalance, only 1200 normal images are randomly chosen from 2400 images that are used for multiclass classification.



**Figure 4.** (a) Normal (b) Low grade (c) High grade

For both classification approaches, random split technique (70:30) is employed to separate samples for training and

testing the proposed architecture. The experiments on the REMBRANDT database are evaluated in terms of sensitivity, specificity, and receiver operating characteristics. For evaluating performance, the associated 95% confidence interval (95% CI) is also computed. These metrics are listed in Table 2.



**Figure 5.** Data augmented images

**Table 2.** Performance metrics used by the proposed PMS-Bcc architecture

Accuracy (Overall System Accuracy)	Sensitivity (Ability to Correctly Identify Positive Cases)	Specificity (Ability to Correctly Identify Negative Cases)
$\frac{T_P + T_N}{T_P + F_P + T_N + F_N}$	$\frac{T_P}{T_P + F_N}$	$\frac{T_N}{T_N + F_P}$

**Table 3.** Confusion matrix

Output Class	Target Class	
	Abnormal	Normal
Abnormal	$T_P$ (abnormal as abnormal)	$F_P$ (normal as abnormal)
Normal	$F_N$ (abnormal as normal)	$T_N$ (normal as normal)

where,  $T_P$  is the classification of abnormal images as abnormal,  $T_N$  is the classification of normal images as normal,  $F_N$  is the classification of abnormal as normal and  $F_P$  is the classification of normal images as abnormal. These metrics are generated by considering the number of accurate and wrong classifications of normal and abnormal images respectively. Based on the prediction results of input normal and abnormal

images, a confusion matrix for binary classification is drawn which is shown in Table 3.

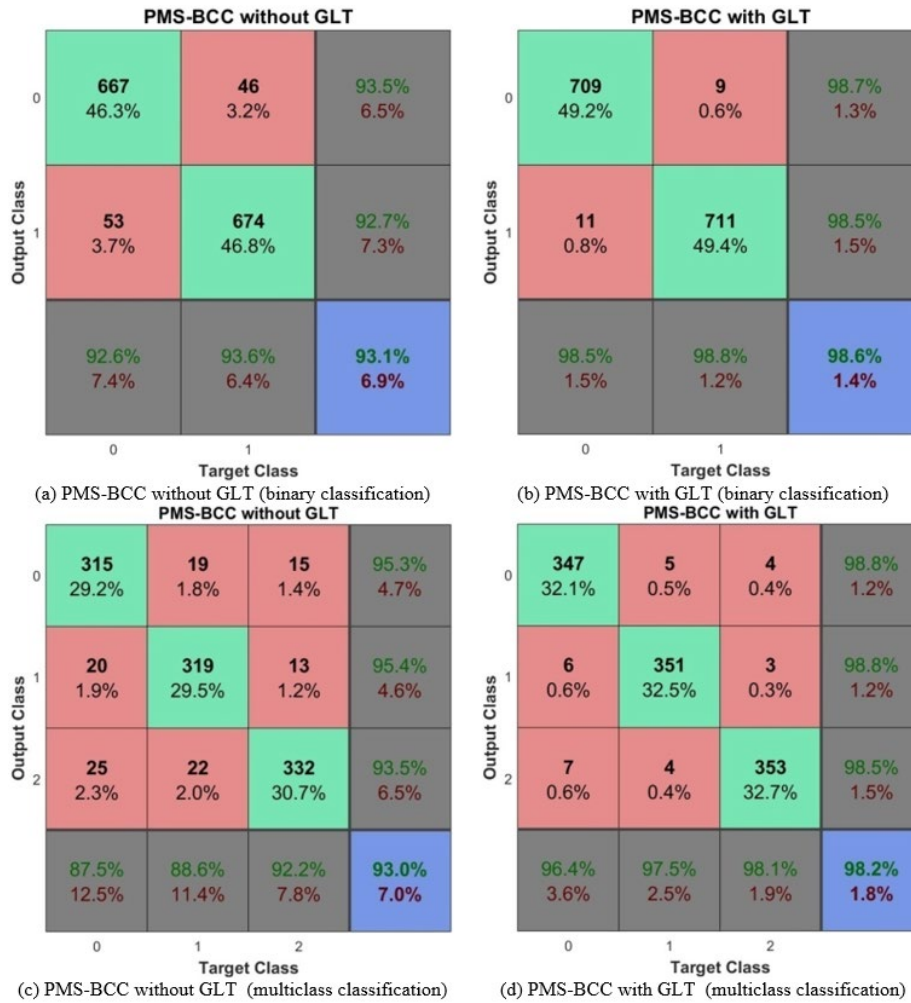
Figure 6 shows the confusion matrix obtained by the proposed PMS-BCC architecture for both binary and multi-class classification. The performances of the proposed PMS-BCC architecture based on the confusion matrices in Figure 5 are summarized in Table 4.

While comparing the two binary classification models, it is evident from Table 4 that PMS-BCC with GLT outperforms PMS-BCC without GLT across key performance metrics. PMS-BCC with GLT demonstrates a significantly higher accuracy of 98.61% compared to PMS-BCC without GLT's 93.13%, signifying its ability to make a larger proportion of accurate predictions overall. Furthermore, PMS-BCC with GLT provides sensitivity (true positive rate), with a rate of



98.47%, compared to PMS-BCC without GLT's 92.64%, indicating that it is more effective at correctly identifying positive cases. Similar trend is observed in terms of specificity also, with PMS-BCC with GLT achieving a rate of 98.75% as opposed to PMS-BCC without GLT's 93.61%. Thus, PMS-BCC with GLT stands out as the stronger candidate for the binary classification task, offering higher accuracy and a

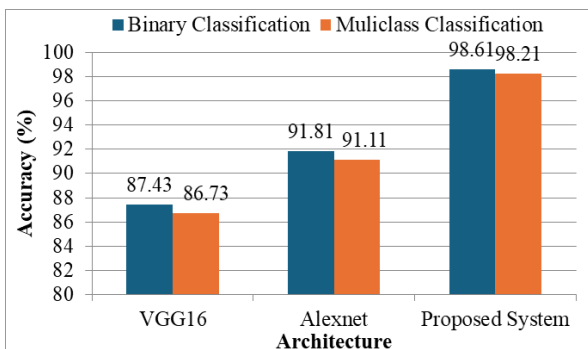
greater capability to discern both positive and negative instances accurately. For multiclass classification, it is clearly demonstrated that the GLT with the proposed architecture gives more accurate results for the classification of MRI brain images than PMS-BCC without GLT. A comparative analysis is provided in Figure 7.



**Figure 6.** Confusion matrices obtained by the proposed PMS-BCC architecture for both binary and multi-class classification

**Table 4.** Performance analysis of the proposed PMS-BCC architecture

Architecture	Accuracy (%)	Sensitivity (%)	Specificity (%)
PMS-BCC without GLT (binary classification)	93.13	92.64	93.61
PMS-BCC with GLT (binary classification)	98.61	98.47	98.75
PMS-BCC without GLT (multiclass classification)	92.96	89.44	94.00
PMS-BCC with GLT (multiclass classification)	98.21	97.31	98.66



**Figure 7.** Comparative analysis with other architectures

## 5. LIMITATIONS OF THIS STUDY

As the REMBRANDT database has an enormous collection of images, the selection of images is crucial for the system's performance. The proposed PMS-BCC architecture uses 200 images which was carefully selected for MRI brain image classification in study [21]. The high quality and clarity of the selected images are well annotated and have traces of tumors. Axial, coronal, and sagittal views are commonly employed in medical imaging to indicate the position and direction of structures within the body. The axial view provides a detailed cross-sectional image of the brain compared to others. The coronal and sagittal views provide a clear view of the frontal

structure and structures in the left-right orientation respectively. The main limitation is that the proposed PMS-BCC system uses only axial views of brain in the REMBRANDT database.

As the proposed PMS-BCC system is trained on REMBRANDT database, it might not perform well on other datasets due to differences in image acquisition, scanning view, population size, diversification of images per classes. Though the proposed system uses data augmentation to avoid overfitting issuing, regularization methods, cross-validation, and the utilization of larger, and more varied datasets can effectively mitigate the problem of overfitting. Regular assessment on an independent validation set during the training process can also aid in identifying and addressing overfitting at an early stage.

## 6. CONCLUSIONS

A well-performing computerized system must be able to assist the radiologists in the detection of brain cancer and offer them a valuable second opinion. DL-based medical image analysis has been a developing area of study for several years. The proposed PMS-BCC architecture presented in this paper is mainly focused for medical image classification, specifically for grading MRI brain images using DL. The developed method combined the architecture of VGG16 and ResNet and then utilizes GLT for effectively training the network. To achieve more accurate results, ALO is introduced between the feature extraction module and classification module. Experimental results on REMBRANDT MRI brain images clearly demonstrated that the GLT outperforms the conventional training for MRI brain image classification. The system has attained the accuracy of 98.6% (95% CI: 97.7%-99.5%) for binary classification with 98.5% (95% CI: 97.6%-99.4%) sensitivity and 98.8% (95% CI: 98%-99.6%) of specificity and for multiclass classification; the obtained average classification accuracy is 98.2% (95% CI: 97.3%-99.1%).

The views of MRI scans in REMBRANDT database are axial, coronal and sagittal views. The main limitation of the proposed PMS-BCC system is that it is designed to classify the axial MRI brain scans only. In future, the proposed work can be extended to classify the tumors, present in all views of MRI scans and different evolutionary algorithms can be used for selecting features to increase the performance of the proposed PMS-BCC architecture. Although the proposed PMS-BCC architecture demonstrates considerable potential in improving the accuracy and efficiency of MRI brain image classification, the integration of the proposed system in medical applications requires ethical considerations, and regulatory approval to ensure patient safety and efficacy.

## REFERENCES

[1] Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA transactions on Signal and Information Processing*, 3: e2. <https://doi.org/10.1017/atsip.2013.9>

[2] Velmurugan, A.K., Padmanaban, K., Kumar, A.M., Azath, H., Subbiah, M. (2023). Machine learning IoT based framework for analysing heart disease prediction. *AIP Conference Proceeding*, 2523(1): 020038.

<https://doi.org/10.1063/5.0110179>

[3] Badža, M.M., Barjaktarović, M.Č. (2020). Classification of brain tumors from MRI images using a convolutional neural network. *Applied Sciences*, 10(6): 1999. <https://doi.org/10.3390/app10061999>

[4] Muthaiyan, R., Malleswaran, M. (2022). An automated brain image analysis system for brain cancer using shearlets. *Computer Systems Science & Engineering*, 40(1): 299-312. <https://doi.org/10.32604/csse.2022.018034>

[5] Srinivasan, S., Bai, P.S.M., Mathivanan, S.K., Muthukumar, V., Babu, J.C., Vilcekova, L. (2023). Grade classification of tumors from brain magnetic resonance images using a deep learning technique. *Diagnostics*, 13(6): 1153. <https://doi.org/10.3390/diagnostics13061153>

[6] Minarno, A.E., Bagas, S.Y., Yuda, M., Hanung, N.A., Ibrahim, Z. (2022). Convolutional neural network featuring VGG-16 model for glioma classification. *JOIV: International Journal on Informatics Visualization*, 6(3): 660-666. <http://doi.org/10.30630/joiv.6.3.1230>

[7] Kumar, S.M., Yadav, K.P. (2021). Brain image classification by deep neural network with pyramid design of inception module. *Annals of the Romanian Society for Cell Biology*, 25(6): 1871-1880.

[8] Kibriya, H., Amin, R., Alshehri, A.H., Masood, M., Alshamrani, S.S., Alshehri, A. (2022). A novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers. *Computational Intelligence and Neuroscience*, 2022(1): 7897669. <https://doi.org/10.1155/2022/7897669>

[9] Ayadi, W., Elhamzi, W., Charfi, I., Atri, M., (2021). Deep CNN for brain tumor classification. *Neural Processing Letters*, 53: 671-700. <https://doi.org/10.1007/s11063-020-10398-2>

[10] Ari, A., Alcin, O.F., Hanbay, D. (2020). Brain MR image classification based on deep features by using extreme learning machines. *Biomedical Journal of Scientific and Technical Research*, 25(3): 19137-19144. <https://doi.org/10.26717/BJSTR.2020.25.004201>

[11] Gu, X., Shen, Z., Xue, J., Fan, Y., Ni, T. (2021). Brain tumor MR image classification using convolutional dictionary learning with local constraint. *Frontiers in Neuroscience*, 15: 679847. <https://doi.org/10.3389/fnins.2021.679847>

[12] Irmak, E. (2021). Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 45(3): 1015-1036. <https://doi.org/10.1007/s40998-021-00426-9>

[13] Veni, N., Manjula, J. (2022). Modified visual geometric group architecture for MRI brain image classification. *Computer Systems Science & Engineering*, 42(2): 825-835. <https://doi.org/10.32604/csse.2022.022318>

[14] Simonyan, K. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>

[15] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778.



- <https://doi.org/10.1109/CVPR.2016.90>
- [16] Mirjalili, S. (2015). The ant lion optimizer. *Advances in Engineering Software*, 83: 80-98. <https://doi.org/10.1016/j.advengsoft.2015.01.010>
- [17] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H. (2006). Greedy layer-wise training of deep networks. *Advances in Neural Information Processing Systems*, 19.
- [18] Scarpace, L., Flanders, A.E., Jain, R., Mikkelsen, T., Andrews, D.W. (2019). Data from REMBRANDT [data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2015.588OZUZB>
- [19] Clark, K., Vendt, B., Smith, K., Freymann, J., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of Digital Imaging*, 26: 1045-1057. <http://doi.org/10.1007/s10278-013-9622-7>
- [20] Brain MRI images. <https://www.cancerimagingarchive.net/collection/rembrandt/>.
- [21] Ayalapogu, R.R., Pabboju, S., Ramisetty, R.R. (2018). Analysis of dual tree M-band wavelet transform based features for brain image classification. *Magnetic Resonance in Medicine*, 80(6): 2393-2401. <https://doi.org/10.1002/mrm.27210>