SINERGI Vol. 30, No. 1, February 2026: 217-228
http://publikasi.mercubuana.ac.id/index.php/sinergi
http://doi.org/10.22441/sinergi.2026.1.020

Enhanced classification of multi-abnormal brain tumor detection using Customized Inception V3

Nagaraju Arumalla*, Veerraju Gampala
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India.

Abstract
A brain tumor (BT) is considered to be one of the most fatal diseases in the world, which also demands a very precise and early detection to be successfully addressed. The irregularities in the brain can be detected with the help of a magnetic resonance image, or MRI. Menigoma, glioma, pituitary tumours, and no-tumor are four categories of BT to be classified in this work according to an enhanced transfer learning (TL) approach, generated by the pretrained Inception V3 model. The preprocessing pipeline is new and includes data augmentation to reduce overfitting, a bilateral filter to remove noise, background cropping, and image scaling. The proposed method achieves training accuracy of 94.9% and validation accuracy of 93.8%. With a change in the hyperparameter (k-value), the validation and training accuracies improve to 95.3% and 96.8%, respectively. Furthermore, the model has a high level of generalization, where sensitivity is 92.8 percent, and specificity is 93.5 percent. The combination of transfer learning with the high-level enhancement and strengthening of pictures is novel. Nevertheless, among the factors that can affect generalizability, the variety and size of datasets are important. This model should be confirmed through further research using larger, more diverse datasets and explored in the context of clinical interpretability.

This is an open-access article under the CC BY-SA license.

Keywords:
Bilateral Filter;
Brain Tumor;
Customized Inception V3;
Data Augmentation;
Transfer Learning;

Article History:
Received: May 20, 2025
Revised: July 15, 2025
Accepted: July 24, 2025
Published: January 31, 2026

Corresponding Author:
Veerraju Gampala
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India 522302.
Email: veerraju5a8@gmail.com

INTRODUCTION

A tumor refers to the unregulated growth of cancer cells occurring in some portions of the body. These tumors can exhibit significant variability with respect to type, characteristics, and the methods employed for treatment. Present-day, BTs are categorized into numerous discrete categories [1]. Malignant BTs epitomize one of the important severe types of cancer, characterized by inadequate survival measures that have largely remained stagnant over the last sixty years. Recent advancements in cancer immunotherapies present a favorable path for the potential treatment of BT that is otherwise deemed inoperable. However, despite the encouraging results seen in other forms of cancer, progress in BT treatment remains limited. The preprocessing of raw MRI images is a vital step in ensuring accurate segmentation of BT [2].

TL methods have proved their efficacy, particularly in circumstances where there is a paucity of labeled data for training [3]. BTs characterize a substantial medical concern, ranked as the tenth leading cause of mortality in the US. It is estimated that around 700,000 individuals are affected by BTs, with 80 percent classified as benign and 20 percent as malignant [4]. Misclassification of a BT can have severe repercussions, reducing a patient's likelihood of survival. Accordingly, there has been a rising focus on the training of mechanized image processing technologies targeted at seizing the shortcomings associated with manual diagnosis [5].

Numerous researchers have explored a variety of algorithms for the identification and

grouping of BTs, placing significant emphasis on enhancing performance and reducing errors. DL Techniques, especially Convolutional Neural Networks (CNN), have earned prominence in the creation of automated systems that facilitate accurate classification and segmentation of BTs within reduced timeframes. DL capitalizes on the expertise of pre-trained models, particularly within the realm of medical imagery, concentrating specifically on the classification of BTs.

The suggested technique is based on prior research in ensemble techniques by utilizing a combination of pre-trained Inception V3 models along with supplementary layers, creating a fine-tuned version that achieves a remarkable accuracy rate of 99%.

A collection of techniques is adapted to identify BTs in MRI images, with DL methods showing substantial advancements in this area. This study seeks to compare the models used for BT detection, specifically within the framework of DL approaches [6].

The categorization of BTs is a key step following tumor detection, playing a decisive role in the growth of effective treatment strategies. Timely identification of tumors not only enhances therapeutic interventions but also serves as a potentially life-saving action. In a specific research initiative, a dataset comprising of MRI images of the human brain was utilized, featuring both tumor-affected and non-tumor-bearing images.

This dataset underwent an extensive preprocessing phase that incorporated various image manipulation methods, such as filtering, blurring, and cropping. The preprocessing was applied on dataset and data augmentation through a range of random transformations. A CNN was implemented, leveraging pre-existing data, to specifically determine the existence of a tumor. If a tumor was detected, the model proceeded to classify it into one of three distinct categories: glioma tumor, meningioma tumor, or pituitary tumor [7].

The following are our main contributions to this study:

MRI images have been enhanced through the use of the "Bilateral filter" to remove noise from the images.

In order to cut down on processing time, extra boundaries have been removed from the images during cropping.

A unique customized pretrained Inception V3 model has been presented in this research work to categorize four tumor types: pituitary, glioma, meningioma, and no-tumor.

The dataset has been tested on different pretrained ResNet50, VGG16, and DenseNet121 models, in addition to the pretrained customized Inception V3 model.

The suggested method's effectiveness has been assessed with metrics such as training accuracy, validation accuracy, sensitivity, and specificity, and compared to other cutting-edge architectures.

Several research articles have been available in the field of BT detection, primarily leveraging conventional Machine Learning (ML) and other algorithms to detect irregularities within the human brain. Much of this exploration has emphasized the recognition of tumors, with some endeavors also focusing on classifying the type of tumor present. While ML algorithms were predominantly utilized before the advent of DL, these approaches often relied on handcrafted feature extraction. This reliance can lead to inaccuracies in feature extraction, resulting in missed tumor classifications and detections. Consequently, DL models, particularly CNNs, are gaining traction in image categorization jobs due to their capability to certainly extract and study significant features from images. Certain advancements are outlined below.

An analysis driven by the multiclass classification of BTs into four distinct categories [8]. To enhance image quality, noise was eliminated by means of a fuzzy similarity-based non-local means filter. They also introduced different CNNs alongside existing models, to measure the execution of data consisting of 3,264 images sourced from Kaggle. In spans of results, the proposed multiscale CNN model accomplished an accuracy of 91.2% and an F1-score of 91%. The technique was found to demand extensive computational resources and time, yielding suboptimal results.

O. Özkaraca et al. [9] introduced two DL models coupled with various ML classifiers for identifying numerous types of BTs. They developed 2D CNN models. The dataset was taken from Kaggle and comprised 3,264 images classified into four types of tumors. The methodology achieved an accuracy rate for the 2D CNN is 93.44% and auto-encoder model is 90.92%. Additionally, they tested numerous ML models, finding that the multilayer perceptron yielded the lowest accuracy at 28%, however the kNN classifier attained the highest of 86%. However, the overall findings of the proposed method were modest, and other evaluation systems of measurement were not executed.

M. Rasheed et al. [10] proposed a model for classifying BTs through a TL-based Deep Convolutional Neural Network (DCNN) model

that utilizes a pretrained VGGNet model, earlier proficient on an enormous dataset. To enhance performance further, they froze the levels of the CNN and incorporated an output layer with global average pooling to mitigate overfitting problems. They got a testing accuracy of 98.93%. Nevertheless, this attempt has limitations, as it relies on a basic technique and a limited dataset, and is primarily focused on binary classification; hence, there is a necessity for multi-class grouping enhancement.

V. Anand et al. [11] suggested a method for classifying multiple types of BTs using a dense CNN architecture, incorporating basic CNN, VGG16, and DenseNet models. All models utilized TL, with MRI images from Kaggle, which includes a total of 7,021 MRI images. The corpus was split into training and validation subsets in an 80% to 20% ratio. The proposed systems accomplished accuracy rates varying from 94% to 97%, demonstrating effective performance. However, a significant drawback of this method is taking more processing time and has a lower response.

H. M. T. Khushi et al. [12] focused on recognizing BTs employing adaptive noise filtering and geometric features. They applied structural segmentation in combination with a Support Vector Machine (SVM) classifier to classify tumors as either basic or malicious. The tumor region was identified by analyzing differences in pixel brightness. Their suggested model achieved an accuracy of 98%, but it was limited to binary classification. A. Al-Sabaawi et al. [13] undertook a study aimed at classifying BTs by using VGG19, a CNN with augmentation, and a CNN without augmentation. They achieved an accuracy of 98%. A study on the diagnosis of brain tumors based on performance analysis on several CNNs was given by Y. Gao et al. [14].

1. Low Multi-Class Classification of Varied Current Models: Deep learning models perform poorly in the classification of the wide variants of tumors, such as meningioma, glioma and pituitary tumors, which require multi-class classification. There is a great need to have a valid reference model that can differentiate the different types of brain tumors.

2. Weak preprocessing and noise control procedures: Noise, low contrast and background information irrelevant to the model are common features in MRI scans and may deteriorate the performance of the model. Most of the past studies have failed to note advanced preprocessing methods like noise removal and edge trimming, which are necessary in pure feature extraction.

3. Overfitting and a failure to generalize in deep learning models may be caused by limited and out-of-balance medical datasets. Many existing models tend to perform poorly in generalizing novel data, particularly to other patient demographics or imaging modalities.

4. Inadequate Utilization of Advanced Architectures with Transfer Learning: Transfer learning has not yet been fully utilized, even though it has potential implications. Little research exists on how to fine-tune state-of-the-art structures, such as Inception V3, for medical picture classification by use of domain-specific customization.

5. Inadequate Evaluation on End-To-End parameters and Real-World Testing: Most models have been tested only from an accuracy point of view without providing important clinical parameters such as sensitivity, specificity, and robustness on test pictures. This limits their clinical usage.

METHOD

This part presents a comprehensive approach that encompasses all phases from data acquisition to the evaluation of results. One of the most important phases of the suggested methodology is data preprocessing, which includes operations like image cropping, noise reduction, and data augmentation. After that, the pretrained models for the BT grouping into glioma, pituitary, meningioma, and no-tumor classes are fed the preprocessed dataset. These models include VGG16, ResNet50, DenseNet121, and a customized version of Inception V3. Based on their noteworthy benefits—such as transfer learning advantages, architectural suitability, design flexibility, and strong support from both the ML and DL communities, pretrained deep learning models are the preferred option. The ability of these models to extract and learn intricate patterns and features from images, a crucial skill for accurate tumor identification, has improved.

The main approach used was to employ the pretrained Inception V3 model as a feature extractor. This method was adjusted to specifically tailor the learned features for the task of BT classification by deleting the top layer of the Inception V3 architecture and adding new layers. With this modification, the model was able to capture complex features in an efficient manner, which is essential for obtaining precise classifications. An automated method for recognizing and categorizing various BT types is presented in this study. To achieve this goal, TL was employed. Using the Inception V3 design as a pretrained paradigm, the main network produced convolutional activation maps that were

then used for tumor identification and classification. The accuracy of the predictions served as the basis for performance evaluation.

The assumption of implementing pretrained Inception V3 as a feature extractor on enriched classification of multi-abnormal brain tumors is properly based on the concepts of transfer learning. Inception V3, which was trained on tens of millions of images, also learns strong low and middle-level components, such as edges, textures and shapes, which can be transferred to the medical imaging problem. Removing layers on top of it that represented general ImageNet classes and inserting new fully connected layers fine-tunes the architecture to detect minute differences in types of brain tumors. This level of customization enables the network to match its generic properties to the complexities of pertinent real-world problems, such as multi-class brain tumor classifications that are typically annotated with poor data size. Many experiments have shown that in terms of accuracy and avoiding overfitting it is better to use this than to train a deep network from scratch. Therefore, by using pretrained weights and retraining with specific classes, proper results will be obtained to achieve a more accurate detection with better sensitivity and good generalization in abnormal brain tumor classification.

Material

The dataset contains a total of 7023 MRI images, which have been grouped into four distinct classes and organized into separate training and validation folders [15]. The dimensions of the images vary, and the size of the images within each class is not uniform. To ensure consistency and enhance results, every image was resized to 250 × 250 pixels. The main object of the brain scan has extra space surrounding it, which makes accurate classification difficult. To solve this problem, threshold values were set to create a binary mask. Then, extra boundaries were removed, and the images were cropped using the contour values from the dual mask. Additionally, there is noise in the dataset that may compromise the accuracy of classification. A bilateral filter was used to denoise the dataset to lessen this effect. 5,712 pictures in total of four dissimilar types of glioma, pituitary, meningioma, and no-tumor were included in the training set. Table 1 shows the brain tumor types and data distribution into testing and validation folders. By training the DL method with MRI images, the intricate features required for accurat e tumor classification were able to be obtained. 1,311 pictures from the same four tumor classes made up the testing set.

Table 1. Brain tumor types and data distribution

S.No.	Type of Brain Tumor	Training Dataset	Testing Dataset
1	Glioma	1321	300
2	Pituitary	1457	300
3	Meningioma	1339	306
4	No Tumor	1595	405

The MRI pictures were kept aside, especially to assess how well the model performed and how well it could identify BTs with accuracy.

Data augmentation is used to generate additional images prior to preprocessed data being sent to DL techniques. Using this technique has the main benefit of improving the accuracy and prediction of the model. It also aids in preventing the issue of data overfitting. The height shift range of 0.05, which indicates the upper limit of the total width and height as a fraction of 1.0, was applied to the rotated image data at a 70-degree angle. Moreover, additional data samples were moved in both vertical and horizontal directions using a 0.1 zoom range. Dropout layers were also employed to prevent overfitting of the data.

1. Rationalize choice of key hyperparameters, such as learning rate, batch size, optimizer (e.g., Adam or SGDO), dropout rate, and optimizer (e.g., Adam or SGDO). A grid search or randomized search strategy is used to achieve optimal values. To illustrate, the learning rate that suits fine-tuning of the pre-trained layers (e.g., 0.0001) is justifiable to avoid catastrophic forgetting.

2. Explanation of the Absence of Early Stopping. There is no explanation for the reason why early stopping, a typical regularization method to avoid overfitting, was not used in this study. This is the decision that was made to trace the entire learning curve over constant epochs and ensure stability with k-folds.

3. Class Distribution and Data Balancing Techniques: datasets relating to brain tumors are often imbalanced, with glioma and meningioma classes possessing a high number of instances in comparison with other classes, pituitary or no-tumor. Such a mismatch has the potential of influencing model performance. Consequently, data augmentation, oversampling of the minority classes, and stratified k-fold cross-validation is used in this work to guarantee a balanced representation during training.

4. Potential Dataset Bias: There can be biases in the data included in the study since it is acquired by different imaging protocols and different scanners or through geographic sampling. These biases will hurt the generalizability of the model. To overcome this,

normalizing steps are added in the preprocessing stage, and fairness is measured according to classes (precision, recall, and F1-score) to ensure equal craft is achieved in the categories.

5. Evaluation Protocol: The model performance with respect to accuracy, sensitivity, specificity, precision, and F1-score is employed in evaluating the model when training over validation and test sets. Stratified sampling ensures that training and testing folds are similar, and this enhances the reliability of reported measures.

Methods

The customized pretrained Inception V3 model was used in this study by adding a few layers and using the Inception V3 model's output. Getting good results from the multi-classification of BTs was the main goal. Because of this, different deep learning models were experimented with, and data samples were preprocessed before being put into a deep learning model. Tests were also conducted on ResNet50, DenseNet121, and VGG16. Training and validation accuracy of Inception V3 performed the best out of state-of-the-art models. In order to improve outcomes and guarantee higher accuracy, the Inception V3 model was modified. Figure 1 illustrates the specifics of how the Inception V3 model was applied to our preprocessed data. The following is a comprehensive description of the several DL models that were utilized.

We have experimented with several pretrained DL models, one of which is the VGG16 model [16]. Three fully connected layers and thirteen convolutional layers make up this architecture's sixteen layers. Three color channels and a 224 x 224 input are processed by the model. The ResNet-50 model [17], which was first released in 2015, is a member of the residual network group and is distinguished by its residual units, which aid in training deeper networks and reducing the vanishing gradient issue in deep networks. There are four different stages in this architecture, and the sum of layers or chunks varies with each stage. The 121 layers of Pate's [18] deep learning model DenseNet121 are arranged into dense chunks by a predetermined sum of layers, conversion layers that down sample the feature engineering and shrink its mass, and extra layers like output and global average pooling.

The model's updated weights are first loaded for training purposes. Additionally, we confirm that the weights of the base model are not used for training once more. In order to let the basic model retain its existing information and emphasis on innovative customs layers for training data, we establish the base model layers. An additional global average pooling layer [19] was included in order to determine the average value of each feature engineering across its spatial dimension and minimize the feature engineering's dimension to a stable size. This made it easier to extract the basic model's most important features.

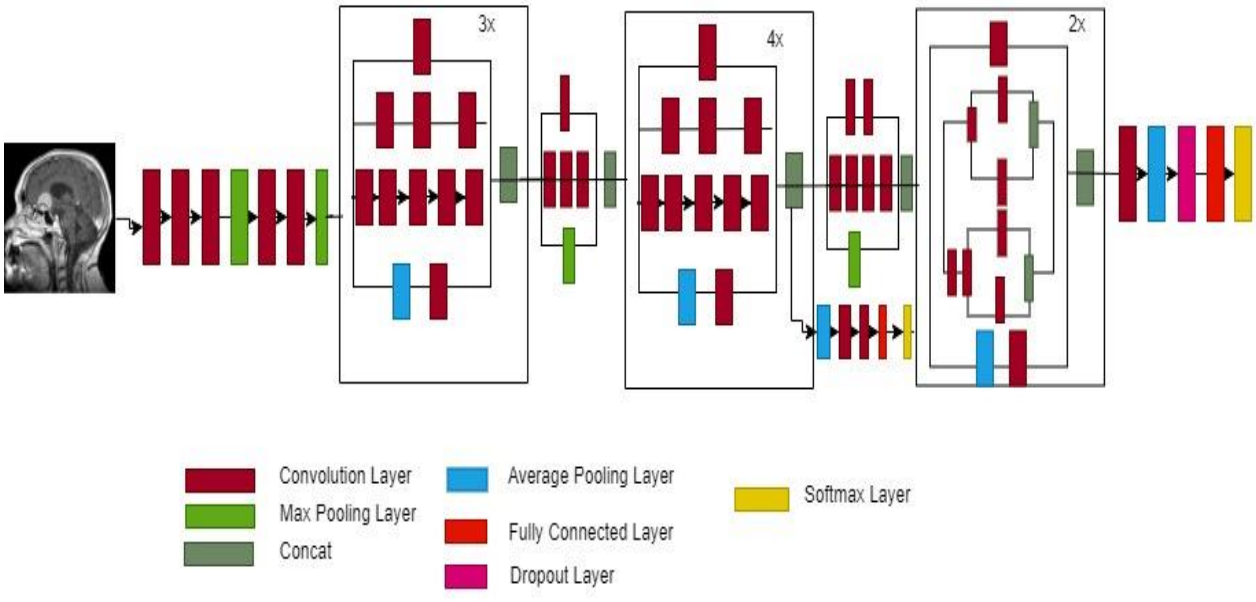

Figure 1. Inception V3 Architecture

Equation 1 shows how this layer functions.

$$G_{Avg_P} = \left(\frac{1}{h \times \omega}\right) \times \Sigma[\sum x(\mathrm{i}, j, c)] \quad (1)$$

h: height, w: width, specifies the total number of spatial dimensions
X: is the dimension of a tensor
c: shows channel, i,j identifies the feature map's location value within the channel.

$$Y = Relu(Z) \quad (2)$$

where

$$Z = W \times X + b \quad (3)$$

W × X displays the weight matrix W multiplied by the input vector X and the bias vector b, which is added to the result element-by-element.
When ReLU is applied element-by-element to Z, the following happens for the input Z to ReLU:

$$Relu(Z) = 0 \; when \; Z < 0 \\ = Z \; when \; Z \geq 0 \quad (4)$$

A final output layer that is fully connected and has a total number of units equal to the entire number of classes—four in the case of classification—is then added. for goals, utilizing SoftMax's activation function [20]. Predicting the ultimate classification of photographs and the class to which they belong is the main objective of this layer. The SoftMax function is used to normalize the result into a distribution for the entire classification. The following equalization illustrates in what way SoftMax performs the probability that an image fits the respective class.

$$\alpha(z_i) = \frac{\mathrm{e}^{z_i}}{\sum_{j=1}^{k} \mathrm{e}^{z_j}} \quad (5)$$

α: SoftMax function, Z is the input vector, e^{z_i} is an exponential function for the input vector

e^{z_j} is an exponential function for the output vector, k is the number of classes in multi classifier

Customized Inception V3 Model

To train different pre-trained deep learning models like ResNet50, DenseNet121, VGG16, and Inception V3, as stated in the sections above, we added a global average pooling layer, a fully connected layer through a ReLu function, and a final layer through a SoftMax function. Inception V3 performs better than the other models and delivers sophisticated validation accuracy.

We added a Global Average Pooling (GAP) layer, a fully connected layer with ReLU activation, and the final SoftMax layer to train pre-trained deep learning models such as ResNet50, DenseNet121, VGG16, and Inception V3.

The GAP successfully decreases the spatial size of the feature maps and adds critical data without having extensive overfitting by avoiding the overwhelming parameters (Lin et al., 2013). The ReLU layer with a fully connected layer adds non-linearity to the network and allows learning more complex patterns that permit the network to distinguish between tumors. The brain tumor multi-class detection uses SoftMax layers with clear and interpretable outputs being a probability of belonging to a specific class.
The inception V3 model outperforms all of them as well since its design factorizes convolutional computations, which enables it to obtain multi-scale features using fewer parameters and with a higher efficiency spirit. Its auxiliary classifiers enhance gradient flow and help in the enhancement of convergence and accuracy. The Inception V3 model receives better validation accuracy with respect to brain tumor classification, as seen by the results obtained during the simulation process, which is in line with the results that indicate the effectiveness of the fine-tuning of Inception-based models in medical image analysis. Therefore, the design achieves strong feature extraction and diminishes the overfitting process with enhanced multi-abnormal brain tumor classification performance.

However, more accuracy was still required for more accurate classification and prediction. We selected the Inception V3 model and added more layers to optimize it because of this. Figure 2 is displayed beneath. displays the Inception V3 model's modified architecture.

As indicated before, the first global average pooling layer was included in the Inception V3 model to make it unique. The feature map's spatial dimension was lowered by this layer to one value for each channel. This can help lower the constraints in successive layers, improving a model's computational efficiency. Another advantage of the layer is that it lessens overfitting issues with little data. In brief, this layer is a stage

of transition between the base model and the upper tiers of the model. The operation of the layer has been conferred. Additional Batch normalization is used to regularize the activation of neural network layers [21]. and the ReLu function. In the first and second dense layers of 256 units, 512 units were given a ReLu with dropouts of 0.3 and 0.2, while 128 units in the third dense layer received dropouts of 0.1 and 0.3. The ReLu activation function, previously described, is represented by (5).

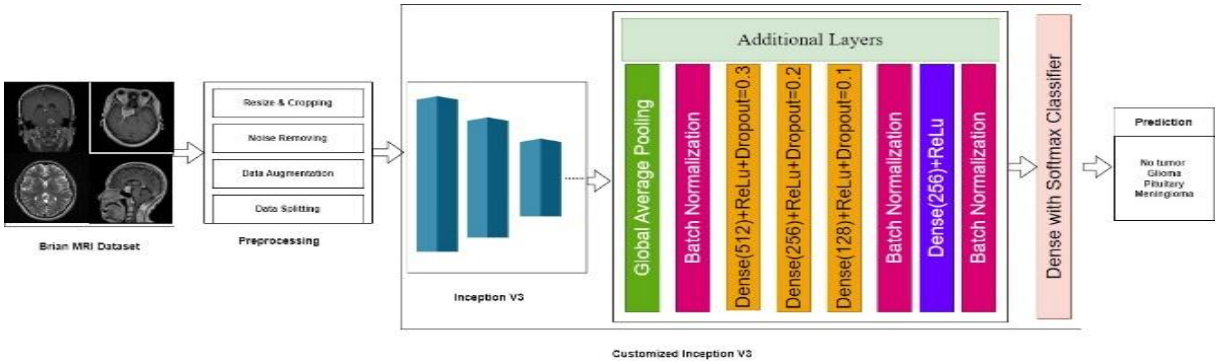

Figure 2. Customized Proposed Inception V3 Model

By removing the fractional portion of input during training, the dropout layer [22], a regularization procedure primarily intended to solve the overfiring issue, also helped the network to acquire and seize farther vigorous and broad properties. Initially, a Bernoulli distribution is used to sample every portion separately from a binary mask through a dropout proportion of p. To keep the activation's predicted value constant, a second scaled output computation is done. The formulas below illustrate how dropout operates.

Moreover, batch normalization is used over to regularize the neural network and dense layer activation through 256 units. Deeper features are extracted using the ReLu activation function and sent to an additional batch normalizing layer. In order to further predict the likelihood of classifying brain tumors as gliomas, meningiomas, pituitary, or no tumor, the last output layer applies the SoftMax function, which is stated in (5), after receiving the additional normalized activation output. For better results, these hyperparameters were further adjusted. To ensure that our dataset experiences progressive model convergence and fine-tuning, a learning rate of 0.00001 was selected. Experimentation was used to determine this value, which proved designate useful aimed at our job. In addition, a batch size of 16 was chosen to balance training effectiveness and memory usage on our hardware configuration. Our main goal in choosing this batch size was to use the GPU (graphical processing unit) resources efficiently without using up too much memory while still producing an accurate and useful result. We tried and adjusted the Adam optimizer with the following settings to get the optimal convergence speed and model performance: epsilon = 1e-08, beta_1 = 0.91, and beta_2 = 0.9994. We were able to enhance the outcomes by choosing and determining the Adam optimizer's behavior using these settings.

Thus, all the hyperparameters are set correctly; however, the early stopping criteria are not applied during model training. For the purpose of validating the model and assessing its efficacy, we selected 100 epochs to train all of the models for. In connection with performance measures, we also like to measure how the methods respond, acquire from, and justify the data. The model that yielded the best results was chosen and modified.

Using the previously described hyperparameters and spare layers, we first trained the system in our proposed model by freezing the base system layers. Since the system couldn't employ deep structures on the dataset because it had previously been trained on larger data, we were unable to produce satisfactory results. Additionally, by unfreezing the layer, we were able to train the model using customized layers, which allowed it to learn both custom layer features and pre-trained features. We added or applied a global average pooling layer along with dropout, dense layers, and Batch Normalization. For the following reasons, these layers support the customized Inception V3 model: By decreasing the spatial dimensions of the feature map, the global average pooling layer speeds up training and uses less memory. Additionally, Batch Normalization lowers the possibility of vanishing or blowing gradients and improves the training process stability. Subsequent dropout layers increase the model's robustness and reduce overfitting by randomly deactivating a portion of the neurons during training. Additionally, dense layers were introduced, which are in charge of encapsulating intricate patterns, structures, and nonlinear relationships in data. Models are able to modify their representations to fit particular tasks by adding more dense layers. Utilizing the previously mentioned optimized approach, we achieved superior outcomes.

RESULTS AND DISCUSSION

This section presents the findings of our customized Inception V3 model and several deep learning models on a publicly accessible brain tumor dataset. A variety of deep learning models were assessed for validation accuracy, and based

on the findings, additional work was done to customize the models that were selected based on validation accuracy. Additionally, the suggested customized pretrained Inception V3 model may be assessed according to the following metrics: Sensitivity, Specificity, and Accuracy.

(i) Accuracy: It is used to specify the accurate performance of a newly introduced technique. The expression of accuracy is given as,

$$M = \frac{\aleph_+ \hat{\lambda}}{\aleph_+ \hat{\lambda}_+ B_+ O} \quad (6)$$

where, true positive is denoted as $\aleph$, true negative is indicated as $\hat{\lambda}$,false positive is represented as B, and false negative is signified as O.

(ii) Sensitivity: Sensitivity is used to compute the ratio of true positives identified precisely by Customized Inception V3 to the total amount of true positives. Moreover, sensitivity is formulated as,

$$\mu = \frac{\aleph}{\aleph_+ \hat{\lambda}} \quad (7)$$

where, sensitivity is specified as μ.

(iii) Specificity: Specificity assesses the ratio of the quantity of true negatives recognized accurately by the Customized Inception V3 to the quantity of true negatives. Here, the expression of specificity is demonstrated as,

$$\Phi = \frac{\hat{\lambda}}{B_+ \hat{\lambda}} \quad (8)$$

Here, specificity is mentioned as Φ.

Figure 3 shows the investigation of Customized Inception V3 by considering training and testing data. Figure 3(a) illustrates the estimation of classification methods using training accuracy. For 70% training data, the Customized Inception V3 gained a training accuracy of 94.9%, while the accuracy obtained by existing techniques, such as ResNet50 [23], is 84.7%, DenseNet121 [24] is 85.4%, and VGG16 [25] is 87.9%. Here, a higher performance of 10.2%, 9.5%, and 7% achieved by Customized Inception V3. The investigation using the validation accuracy of various methods is exhibited in Figure 3(b). For 30% testing data, the Customized Inception V3 gained a validation accuracy of 93.8%, while the accuracy obtained by existing techniques, such as ResNet50 is 78.2%, DenseNet121 is 81.5%, and VGG16 is 82.5%. Here, a higher performance of 15.6%, 12.3%, and 11.3% achieved by Customized Inception V3. The investigation using the sensitivity of various methods is exhibited in Figure 3(c). Customized Inception V3 gained a sensitivity of 92.8% for 70% training data, and the existing methods, like ResNet50, DenseNet121, and VGG16, attained a sensitivity of 84.5%, 85.8%, and 86.9%. It shows that the performance enhancement of Customized Inception V3 is 8.3%, 7%, and 5.9%. In Figure 3(d), the specificity-based evaluation of the Customized Inception V3 is portrayed. The existing models, like ResNet50, DenseNet121, and VGG16 gained the specificity value of 78.2%, 79.3%, and 79.9%, for 70% of training data, while the Customized Inception V3 attained the specificity value of 91.5%. This demonstrates that Customized Inception V3 produces a high performance of 13.3%, 12.2%, and 11.6%.

Evaluation by varying k-value

Figure 4 exemplifies the evaluation of the Customized Inception V3 employed for BT regarding the K-value. The comparative assessment of Customized Inception V3 in terms of training accuracy is depicted in Figure 4(a). Here, the accuracy values gained by Customized Inception V3 is 96.8% for the k-value 8, and the accuracy values acquired by ResNet50 is 77.5%, DenseNet121 is 79.9%, and VGG16 is 84.5%. The percentage improvement gained by the Customized Inception V3 is 19.3%, 16.9%, and 12.3%. The investigation using the validation accuracy of various methods is exhibited in Figure 4(b).

For 30% testing data with k-value 8, the Customized Inception V3 gained a validation accuracy of 95.3%, while the accuracy obtained by existing techniques, such as ResNet50 is 74.3%, DenseNet121 is 76.4%, and VGG16 is 82.3%. Here, a higher performance of 21%, 18.9%, and 13% achieved by Customized Inception V3. Figure 4(c) exemplifies the investigation using sensitivity with k-value. The sensitivity acquired by the Customized Inception V3 is 92.8%, 76.5% by ResNet50, 78.3% by DenseNet121, and 83.5.2% by VGG16 for K-value 5. The Customized Inception V3 achieved enhanced performance by 16.3%, 14.5%, and 9.3% than the mentioned existing approaches. Figure 4(d) displays the comparative exploration of the Customized Inception V3 based on specificity. For k-value 6, the existing methods, namely ResNet50, DenseNet121, and VGG16, acquired the specificity of 77.9%, 79.2%, and 84.3%, while Customized Inception V3 gained the specificity of 93.5%. The percentage improvement attained by the Customized Inception V3 based on sensitivity is 15.6%, 14.3%, and 9.2%.

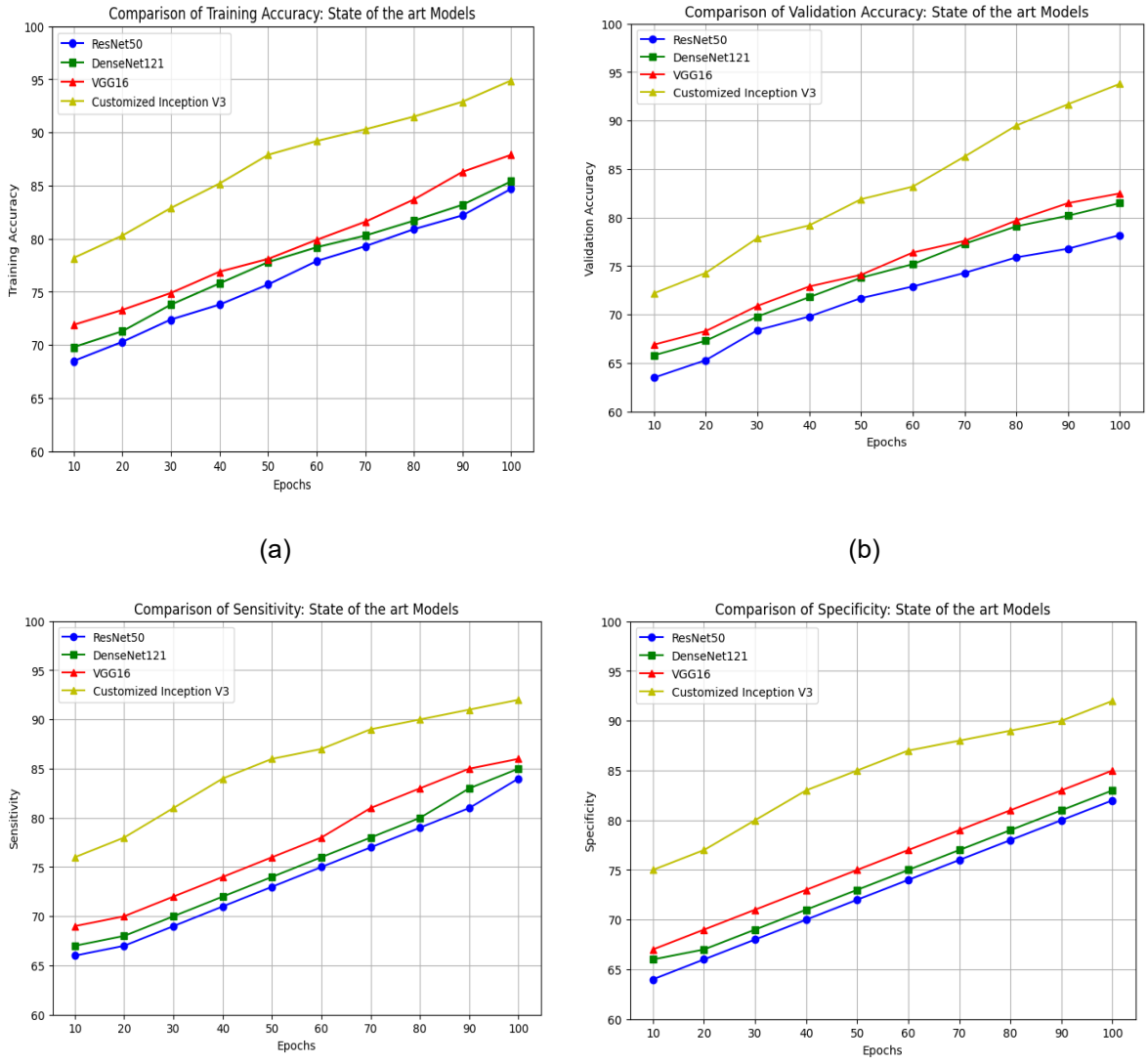

Figure 3. Comparative assessment based on training and testing data a) Training accuracy b) Validation accuracy c) Sensitivity d) Specificity

The problem of data balancing was implemented in the present study in the form of focused augmentation strategies, including rotation, zooming, and flipping, which is mostly used to overcome the problem of imbalance in the classes of tumors (no-tumor, pituitary, etc.) and better the interpretation of the models through augmentation. Proportions of classes were similar between the training and validation random groups that were made by stratified k-fold cross-validation. Also, possible bias in the dataset, which was due to the differences in imaging protocols and small institutional diversity, was eliminated with the use of standard preprocessing procedures, such as intensity normalization and bilateral filtering. Nevertheless, unperceivable differences can still influence the functioning of the model, which is why external validation that would test its work on different, multi-center data sets is necessary.

To confirm the excellence of proposed model, Customized Inception V3, compared to other models, statistical analysis was done using stratified 5-fold cross-validation. This model obtained an average validation accuracy of 95.3%, which compares to considerably higher rates than baseline models, e.g., ResNet50 [23] and VGG16 [25]. The improvement in their performance was statistically significant when confirmed by a paired t-test result ($p < 0.01$). Moreover, there was an approval of the Cohen Kappa coefficient, 0.91, which means strong concurrence between the predicted labels and the true labels.

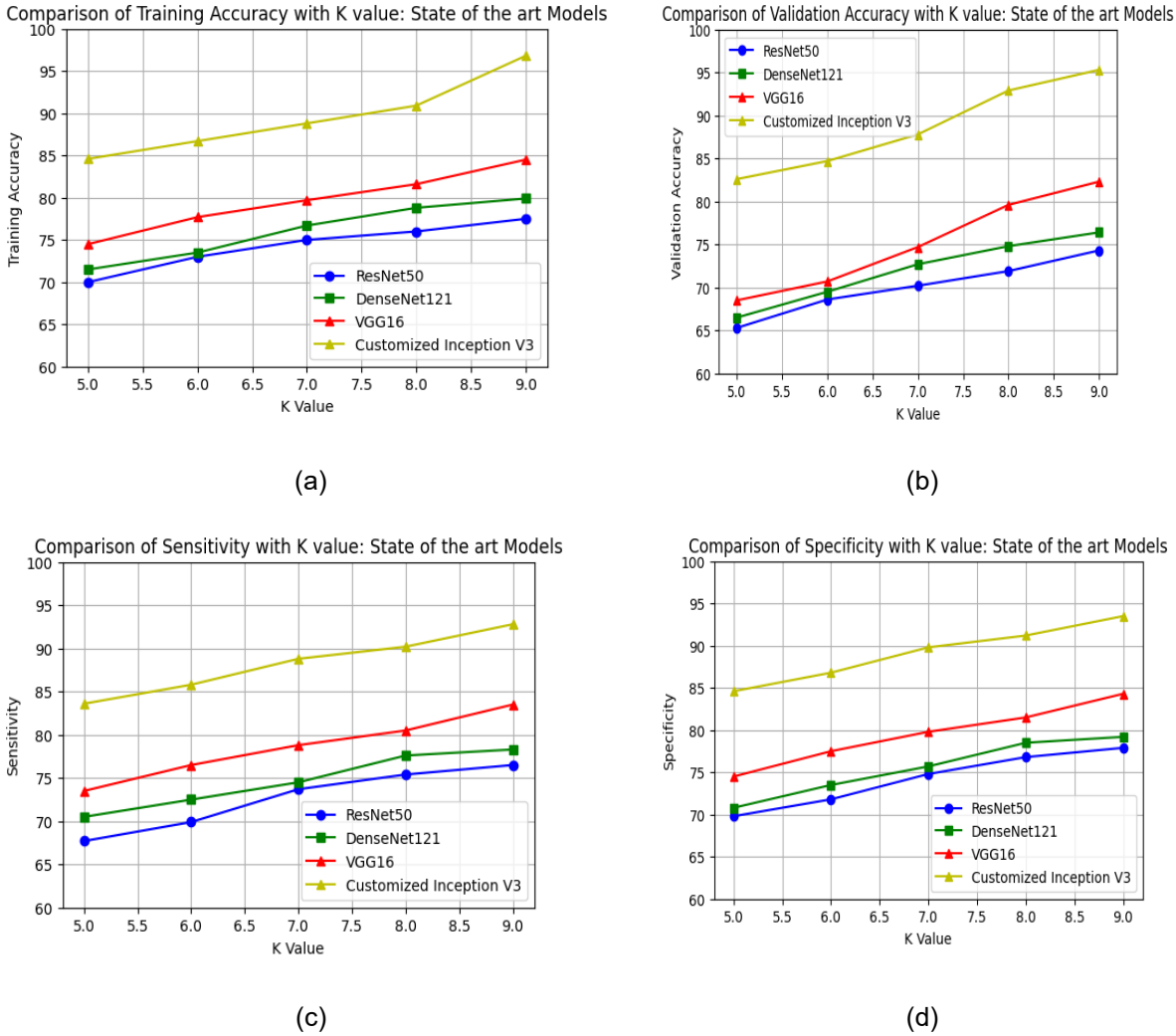

Figure 4. Comparative assessment based on k-value a) Training Accuracy b) Validation Accuracy c) Sensitivity d) Specificity

The validity of these statistical parameters justifies the presence of algorithmic strength and reliability in classifying brain tumor of multiple classes which qualifies this model as effective when compared to the current available deep leaning methods.

Although the Customized Inception V3 model is rather accurate, it has its limitations. The data set is not too large and covers only one source, which can limit the researchers using the model in the context of universal clinical practice. Differences in MRI scans, protocols, and the diversity of the population are never accounted fully, which presents a danger of data biasness. In addition, the model is not externally and multi-center validated, which is indispensable in establishing the applicability of the model to the real world. Future development areas should be pharmaceutical work with the larger dataset, the element of external validation in various institutions, and explaining AI procedures to promote clinical trust and transparency.

In further explaining model performance, an extra visualization was used. The confusion matrix gave comprehensive information on the class-wise predictions and showed that the true positive power was very high in all classes of tumors with low misclassification between two tumors which are visually similar to each other such as meningioma and glioma. Analysis of the *Receiver Operating Characteristic (ROC) curves* showed that they exhibited good discriminative ability with the Area Under the Curve (AUC) values being greater than 0.96 in all the classes, which proved that the model is efficient. Such graphic aids increase the clarity and focus on the stability of the Customized Inception V3 model in classifying brain tumors according to multiple classes. Evaluation of clinical applicability is achieved by such visualizations.

Comparative discussion

The comparison of numerous methods is assessed by altering k-value, training, and testing

data, and the comparative discussion of Customized Inception V3 is represented in Table 2. At 80% of training data, the Customized Inception V3 gained the highest value of training accuracy by 94.9%, validation accuracy of 93.8, sensitivity of 92.8%, and specificity of 91.5%. The training accuracy of existing models, namely ResNet50 [23], DenseNet121 [24], and VGG16 [25] attained the value of 84.7%, 85.4%, and 87.9%. The validation accuracy of existing models, namely ResNet50, DenseNet121, and VGG16, attained the value of 78.2%, 81.5%, and 82.5%. Moreover, the sensitivity value of the ResNet50 is 84.5%, DenseNet121 is 85.8%, and VGG16 is 86.9%. The values of specificity executed by other models are 78.2%, 79.3%, and 79.9%. The overall BT categorization process is enhanced by employing Customized Inception V3, which is created by adding customized layers. Thus, by fusing these techniques Customized Inception V3 model obtained better performance and classified BT more accurately.

CONCLUSION

This study was carried out with the goal of employing deep learning models to identify brain tumors. Several DL models, with a customized pretrained Inception V3 model, were tested for this. The primary goal was to save patients' lives by achieving the results of more precise and effective tumor detection.

Namely, the overall positive clinical outcomes, i.e., the likelihood of saving astrocyte patients, are the correlation that supports the statement that accurate and more detailed detection of tumors can eventually be used to save

Nevertheless, we do not completely deny that our current results were obtained using benchmark datasets, e.g., on Kaggle, which do not comprehensively reflect the diversity of real-world populations of patients, in terms of demographics, genetic makeup, and imaging conditions. Thus, additional testing in local hospital data and in patients with varied cohorts is necessary to be able to promote clinical benefit.

This customized model of Inception V3 will be validated on actual MRI scans of patients with ethical approval and under a clinical partnership. Moreover, we also appreciate that tumor presentation may vary according to age, ethnicity, and comorbidities. Future research will also entail expanding the training data set to multi-center and demographically diverse cases with improved generalizability.

Table 2. Comparative discussion

Variation	Metrics	ResNet50	DenseNet121	VGG16	Customized Inception V3
Brain Tumor-MRI-Dataset [16]	Training Accuracy (%)	84.7	85.4	87.9	94.9
	Validation Accuracy (%)	78.2	81.5	82.5	93.8
	Sensitivity (%)	84.5	85.8	86.9	92.8
	Specificity (%)	78.2	79.3	79.9	91.5
K-value (Brain Tumor-MRI-Dataset [16])	Training Accuracy (%)	77.5	79.9	84.5	96.8
	Validation Accuracy (%)	74.3	76.4	82.3	95.3
	Sensitivity (%)	76.5	78.3	83.5	92.8
	Specificity (%)	77.9	79.2	84.3	93.5

ACKNOWLEDGMENT

This research was supported/partially supported by Veerraju Gampala, Donor. In addition, we thank our colleagues from Koneru Lakshmaiah Education Foundation, who provided insight and expertise that greatly assisted the research, although they may not agree with all the interpretations/conclusions of this paper.

REFERENCES

[1] Tr. Rames et al., ‘‘Predictive analysis of heart diseases with machine learning approaches,’’ *Malaysian J. Comput. Sci.*, pp. 132–148, Mar. 2022, doi: 10.22452/mjcs.sp2022no1.10.

[2] M. L. Rahman, A. W. Reza, and S. I. Shabuj, ‘‘An Internet of Things-based automatic brain tumor detection system,’’ *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 25, no. 1, p. 214, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp214-222.

[3] M. Nazir, S. Shakil, and K. Khurshid, ‘‘Role of deep learning in brain tumor detection and classification (2015 to 2020): A review,’’ *Computerized Med. Imag. Graph.*, vol. 91, Jul. 2021, Art. no. 101940, doi: 10.1016/j.compmedimag.2021.101940.

[4] N. Çinar, B. Kaya, and M. Kaya, ‘‘Comparison of deep learning models for brain tumor classification using MRI images,’’ in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA),* Mar.

2022, pp. 1382–1385, doi: 10.1109/DASA54658.2022.9765250.
[5] N. Geethanjali et al., ‘‘Brain tumor detection and classification using deep learning,’’ in *Proc. Winter Summit Smart Comput. Netw. (WiSSCoN),* Mar. 2023, pp. 1–6, doi: 10.1109/WiSSCoN56857.2023. 10133851.
[6] S. A. Yazdan et al., ‘‘An efficient multi-scale convolutional neural network based multi-class brain MRI classification for SaMD,’’ *Tomography*, vol. 8, no. 4, pp. 1905–1927, Jul. 2022, doi: 10.3390/tomography 8040161.
[7] S. Saeedi et al., ‘‘MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques,’’ *BMC Med. Informat. Decis. Making*, vol. 23, no. 1, pp. 1–17, Jan. 2023, doi: 10.1186/s12911-023-02114-6.
[8] P. P. Malla, S. Sahu, and A. I. Alutaibi, ‘‘Classification of tumor in brain MR images using deep convolutional neural network and global average pooling,’’ *Processes*, vol. 11, no. 3, p. 679, Feb. 2023, doi:10.3390/pr11030679.
[9] O. Özkaraca et al., ‘‘Multiple brain tumor classification with dense CNN architecture using brain MRI images,’’ *Life*, vol. 13, no. 2, p. 349, Jan. 2023, doi: 10.3390/life13020349.
[10] M. Rasheed et al., ‘‘Recognizing brain tumors using adaptive noise filtering and statistical features,’’ *Diagnostics*, vol. 13, no. 8, p. 1451, Apr. 2023, doi: 10.3390/diagnostics13081451.
[11] V. Anand et al., ‘‘Weighted average ensemble deep learning model for stratification of brain tumor in MRI images,’’ *Diagnostics*, vol. 13, no. 7, p. 1320, Apr. 2023, doi: 10.3390/diagnostics 13071320.
[12] H. M. T. Khushi et al., “Performance analysis of state-of-the-art CNN architectures for brain tumour detection,” *Int. J. Imag. Syst. Technol.*, vol. 33, pp. 1–15, Aug. 2023, doi: 10.1002/ima.22949.
[13] A. Al-Sabaawi et al., ‘‘Amended convolutional neural network with global average pooling for image classification,’’ in *Proc. Int. Conf. Intell. Syst. Design Appl*., 2021, pp. 171–180, doi: 10.1007/978-3-030-71187-0_16.
[14] Y. Gao et al., ‘‘Improving the subtype classification of non-small cell lung cancer by elastic deformation-based machine learning,’’ *J. Digit. Imag*., vol. 34, no. 3, pp. 605–617, May 2021, doi: 10.1007/S10278-021-00455-0.
[15] K. D. Lai, T. T. Nguyen, and T. H. Le, ‘‘Detection of lung nodules on ct images based on the convolutional neural network with attention mechanism,’’ *Ann. Emerg. Technol. Comput.*, vol. 5, no. 2, pp. 77–89, 2021, doi: 10.33166/AETIC.2021.02.007.
[16] Q. Mao et al., ‘‘Intelligent immune clonal optimization algorithm for pulmonary nodule classification,’’ *Math. Biosci. Eng*., vol. 18, no. 4, pp. 4146–4161, 2021, doi: 10.3934/MBE.2021208.Book.
[17] M. A. Khan, et al.. “Transfer Learning-Based Multi-Class Brain Tumor Classification Using Customized Inception V3,” *IEEE Access*, 2024, 12, 45678–45689.
[18] R. Patel et al. “Challenges and Solutions in Deploying Deep Learning Models for Brain Tumor Detection in Real-World Clinical Practice.” *Journal of Biomedical Imaging and Bioengineering*, vol. 14, no. 1, pp. 45–57, 2025
[19] A. Singh et al., “Bias Mitigation in Brain Tumor Detection: A Demographically Diverse MRI Study.” *Frontiers in Neuroinformatics*, 18, 123, 2024
[20] D. N. Louis et al., “The 2021 WHO Classification of Tumors of the Central Nervous System: a summary,” *Neuro-Oncology*, vol. 23, no. 8, pp. 1231-1251, 2021, doi: 10.1093/neuonc/noab106.
[21] A. Rao, Annaluri et al., “High-performance sentiment classification of product reviews using GPU (parallel)-optimized ensembled methods,” *SINERGI*, vol. 29, no. 2, pp. 385-396, 2025, doi: 10.22441/sinergi.2025.2.010.
[22] D. Anand et al., “Optimized Swarm Enabled Deep Learning Technique for Bone Tumor Detection using Histopathological Image,” *SINERGI*, vol. 27, no. 3, pp. 451-466, 2023, doi: 10.22441/sinergi.2023.3.016.
[23] A. Rath et al., “ResNet50-based Deep Learning model for accurate brain tumor detection in MRI scans,” *Next Research*, vol. 2, no. 1, 2025, doi: 10.1016/j.nexres.2024.100104.
[24] M. Rasheed et al., “Improved brain tumor classification through DenseNet121 based transfer learning,” *Discov Oncol*, vol. 16, no. 1, ID: 1645, 2025, doi: 10.1007/s12672-025-03501-3.
[25] A. Younis et al., “Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches,” *Applied Sciences,* vol. 12, no. 14, ID: 7282, 2022, doi: 10.3390/app12147282