DenseNet201-Based Deep Transfer Learning Framework for Brain Tumor Classification in MRI Scans

Doaa Sami Khafaga[1,*]

[1]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Email: dskhafga@pnu.edu.sa

Abstract

The classification of brain tumors is crucial in the context of early intervention, as the appropriate and timely diagnosis can significantly influence the treatment plan and patient outcomes. Radiologists have long relied on their own judgment and have read these medical images through their own eyes, which is often subjective, time-consuming, and inter-observer variability is also likely to occur. Applications built on artificial intelligence (AI), or more specifically, deep learning (DL)-based algorithms, have radically changed the medical imaging field over the last couple of years and could potentially be used to automate the diagnosis process, offering prompt, trustworthy, and unbiased assessments. Despite such developments, most existing systems that rely on AI are constrained, especially when it comes to classification accuracy and robustness across different datasets. To overcome these problems, the article in this chapter presents a more effective DL model with a specifically designed architecture that aims to improve the classification of brain tumors. The specified methodology is based on preprocessing and data normalization steps that reduce noise and level out the data intensity, enabling effective feature extraction from the MRI images. This will increase the accuracy of the later classification. The primary component of the proposed methodology is an adapted version of DenseNet-201, designed explicitly for the four-class brain tumor classification. To achieve optimal performance, the conventional output layer of DenseNet-201 was replaced with a Global Average Pooling (GAP) layer, designed to address the issues of vanishing gradients and overfitting commonly encountered during the training of deep networks. The architectural adjustment helps to combine the features and increase the overall generalization capacity of the model. The model was thoroughly tested using two datasets: one publicly available dataset on Figshare and a locally available dataset comprising a total of 3,504 T1-weighted contrast-enhanced MRI (T1-w MRI) images. The results of the experiment provided the proposed model with a general accuracy of 100 percent, which was higher than that of the existing comparative methods. Such results support the idea that complex architectural adjustments with the broader preprocessing strategy can be effective, and why deep neural networks can be viewed as trustworthy diagnostic tools in clinical neuro-oncology, potentially achieving extremely high accuracy.

Keywords: Brain tumor; Deep Learning; DenseNet201; Multi Classification; Transfer Learning

1 Introduction

A brain tumor is characterized as an abnormal proliferation of cells within the brain, which may be either malignant (cancerous) or benign (non-cancerous) [1]. Malignant tumors are aggressive in nature, expanding uncontrollably, infiltrating nearby tissues, and interfering with normal brain function, whereas benign tumors generally remain localized and do not metastasize to adjacent structures. Despite their less aggressive behavior, benign tumors can still cause severe complications due to the confined space of the cranial cavity. Brain tumors,

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

therefore, represent a diverse and complex class of diseases, varying in biological behavior, prognosis, and treatment response. According to Behin et al. [2], brain tumors are categorized into different sorts and grades based on histological features and clinical manifestations, making their diagnosis and classification critical to effective treatment planning. Multi-class classification approaches extend beyond simple identification by also determining the specific tumor type, a process with significant implications for both treatment strategies and patient outcomes. Among the most prevalent brain tumors are meningiomas (35%), gliomas (16%), and pituitary tumors (14%). Meningiomas usually originate from the meninges, pituitary tumors arise in the region of the pituitary gland, and gliomas infiltrate glial cells, often spreading diffusely across the brain [3]. The epidemiological statistics underscore the urgent need for improved diagnostic systems. The Central Brain Tumor Registry of the United States (CBTRUS) reveals that an estimated 700,000 Americans currently live with a brain tumor, with almost 84,000 new cases of primary brain tumors diagnosed in 2021 alone (CBTRUS) [4]. Porter et al. [5] also observed that the prevalence of brain tumors in children in the United States exceeds 28,000 children, which highlights its effects on children of all ages. Worryingly, primary malignant brain tumors are likely to result in the death of about 18,000 people annually. Primary brain and central nervous system (CNS) tumors are more than 120 in number, with almost one-third (29.7%) being malignant. Radiologically, the classification of brain tumors is especially not easy because it is heterogeneous, has overlapping radiological features, isointense and hypointense, and perilesional edema. T1-weighted contrast-enhanced MRI (T1-Weighted MRI) is the most widely used imaging modality in tumor imaging. Contrast agents, including gadolinium (0.15-0.20 mmol/kg), are more effective at making lesions visible, thereby increasing the accuracy of diagnosing both primary (e.g., meningiomas) and secondary malignancies [6]. The classification of brain tumors is complex, and therefore, it cannot be assessed solely by manual computations in the form of radiology. Traditional methods have typically employed handcrafted feature extraction and region-based segmentation methods, which require considerable effort from radiologists and are prone to inter-observer variation [7]. Nevertheless, with the introduction of deep learning (DL), this paradigm has undergone significant changes. DL models are especially well-suited for radiological image analysis due to their capability to learn hierarchical feature representations directly from the raw image data [8]. These models can be used to represent increasingly abstract and discriminative patterns through a series of nonlinear transformations, which enables highly accurate classification. Notably, with pre-trained networks, transfer learning allows training on even comparatively small medical datasets, and, therefore, DL methods are effective and robust in clinical settings. DL-based Computer-Aided Diagnosis (CAD) systems can alleviate the workload of radiologists by providing a fast, objective, and dependable tumor classification, which facilitates early diagnosis and enhances patient management as well [9] [10]. Considering these opportunities and challenges, the primary objective of this chapter is to design an automated system for brain tumor detection and classification using deep learning, which will utilize both publicly accessible and locally gathered MRI data. The main contributions of this work can be summarised as following: (*i*) the modified transfer learning framework using DenseNet201 architecture is proposed, which includes a Global Average Pooling (GAP) layer, which interferes with the overfitting and vanishing gradient issues; (*ii*) it is proposed to use pre-trained ImageNet weights and a convolutional neural structure due to which the robust feature extraction is conducted; and (iii) the performance of the proposed work is evaluated in extensive experiments to critically test the performance of the model against the state-of-the-art methods The remaining part of the paper is organized in the following way.A review of related studies on the classification of brain tumors is provided in Section 2. The contents and data of this work can be found in the section on reasons. Section 3 describes the suggested methodology (preprocessing, feature extraction, and classification pipeline). Section 4 provides the results of the experiment and discussions. Finally, Section 5 is the conclusion of the paper, where statements and prospective research directions are made.

2 Related work

The past few years have seen a surge in research interest in the use of magnetic resonance (MR) imaging for the detection and classification of brain cancers. To address the complexity of medical imaging, researchers have employed a diverse range of algorithms, including traditional machine learning (ML) algorithms and more advanced deep learning (DL) architectures. In the form of a case study, publicly available brain tumor data on Figshare [11] and a local dataset have been widely used to assess the efficacy of these techniques. The given section describes essential progress in the classification and detection of brain tumors, both hand-created feature-based methods and neural networks. Conventional ML techniques have worked in brain tumor analysis mainly in terms of feature engineering and dimensionality reduction. For example, a Support Vector Machine

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

(SVM) classifier was used in conjunction with principal component analysis (PCA) or linear discriminant analysis (LDA) to classify brain tissues into white matter, gray matter, cerebrospinal fluid, and abnormal or normal regions [12]. This technique, based on the intensive use of handcrafted elements such as intensity, texture, and shape descriptors, achieved a remarkable classification accuracy of 98.87 percent. Nevertheless, one drawback of this method is that PCA has been used as a dimension reduction technique, which may result in the loss of discriminative texture information. Likewise, a hybrid approach was proposed, which combines grey wolf optimization (GWO) and multiclass SVM to classify brain tumors [13]. Using GWO to minimize and optimize features, the experiment demonstrated a higher level of performance compared to other metaheuristics, such as the firefly algorithm and particle swarm optimization, achieving a classification accuracy of 95.23%. The present works highlight the effectiveness of metaheuristic optimization to improve the performance of the classical ML classifiers. Other works focused on new partitioning and feature extraction methods. An example of this approach is the adaptive spatial division and ring-form partitioning techniques applied to regions of interest within MR images, where features such as histograms of intensity, gray-level co-occurrence matrices, and bag-of-words descriptors are extracted in those regions of interest within the MR image [14]. The classification accuracies varied between 87.54% and 91.28%, depending on the feature set, which was achieved by this approach. Similarly, a back-propagation multilayer perceptron neural network was used on wavelet-based features and Gabor filters with 91.9 % accuracy in the classification of brain tumors [15]. On the same note, the statistical and texture-based analysis of data, involving the use of gray-level co-occurrence matrix (GLCM) features in conjunction with probabilistic neural networks, also achieved an accuracy of 83.33% [16]. These studies collectively demonstrate the significant role of feature extraction in previous research on brain tumor image classification. The shift to deep learning models represents a significant paradigm shift, as these techniques do not rely on manually designed features but instead learn from raw images. For example, convolutional neural networks (CNNs) and their derivatives have been gaining popularity. Capsule networks (CapsNet), which aim to maintain spatial hierarchies of tumor structures and their surrounding tissues, obtained an overall discontinuity of 90.89% in classification problems [17]. In a different work, several CNN models were evaluated, the best of which combined ReLU activation layers with max-pooling with a training error of 98.5 and a validation error of 84.19% [18]. There has also been a high success in transfer learning strategies. As an illustration, with a pre-trained model, researchers achieved a patient-level accuracy of 98% in classifying images from the Figshare dataset [19]. On the same note, the image classification of a VGG19 network was fine-tuned to an accuracy of 94.82% [20]. Alternative hybrid methods, such as artificial neural networks (ANN) and k-nearest neighbors (KNN), showed a high level of performance with accuracies of 97% and 98%, respectively [21]. Metaheuristic optimization has also been performed in deep learning workflows. Most recently, hybrid feature extraction of normalized GIST descriptors and PCA, followed by a regularized extreme learning machine, showed a 94.23% increase in accuracy of the correct output [22]. In addition, high classification rates of up to 94% could also be achieved by genetic algorithm-optimized CNNs, which may also lead to high classification accuracies in the future [23]. These papers demonstrate that optimization may be utilized to complement deep neural networks and improve their quality and accuracy. Although these have been developed, conventional CNNs are likely to encounter issues such as fading gradients when training deep networks. To overcome these drawbacks, later studies have introduced more recent deep transfer learning systems, supported by DenseNet201. Using the pre-trained DenseNet201 models and global average pooling, they achieved impressive gains in both training and validation accuracy, reaching up to 100% in some instances. This represents a significant advancement in strong automated brain tumor classification systems, which may ultimately surpass manually based schemes, while addressing some of the deficiencies found in previous handcrafted feature-based and standard deep learning schemes.

3 Materials and Methods

In this section, we present the datasets used in the analysis and provide an overview of the proposed approach, which comprises several processing steps. The flow diagram of our analysis pipeline is presented in Fig. 1 and works in the following order: preprocessing and normalization, CNN model training and parameter optimization and, lastly, multi-label classification. In the subsequent subsections, all of the steps will be depicted.

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

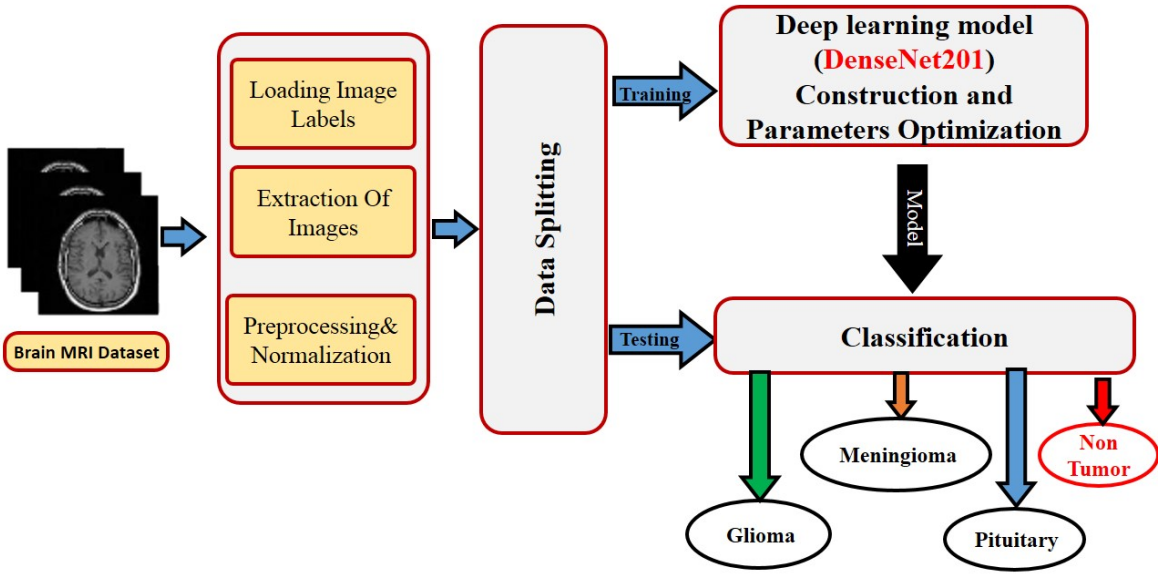


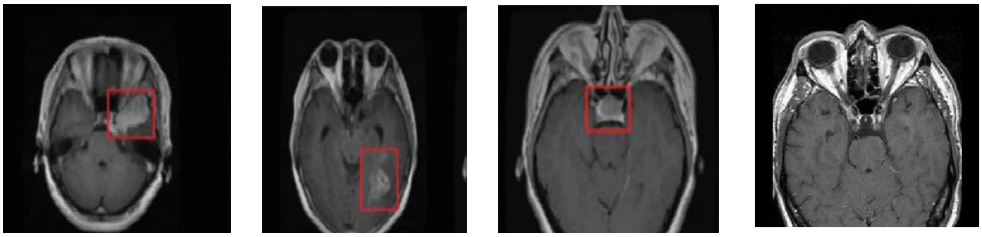
Figure 1: Overview of proposed methodology

3.1 Dataset

The dataset used in this study comprises both publicly available data, provided by Cheng [11] [24], and a locally acquired dataset, provided by the Department of Diagnostic Radiology at Mansoura University, Egypt. Both datasets contain a total of 3504 brain tumor MR images. There are three types of brain tumors in the dataset: meningioma (981images), glioma (1003 images), and pituitary (959 images). In this paper, we used an additional class for normal (or no tumor) cases, which contains 561 images. Each image contained an original size of 512 x 512 pixels. Examples of the dataset images are shown in Fig. 2, and Table 1 shows the details of the dataset distribution. Finally, the proposed model is used to identify brain tumors on test results.

Table 1: Dataset distrubtion for each class.

Class	# Images
Glioma	1003
Meningioma	981
Pituitary	959
No tumor	561
Total	**3504**



(a) (b) (c) (d)

Figure 2: Example of axial brain T1-weighted contrast-enhanced images taken from the dataset; (a) Meningioma ; (b) Glioma ;(c) Pituitary ; and (d) No tumor.

DOI: https://doi.org/10.54216/JAIM.100204
Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

3.2 Data Preprocessing

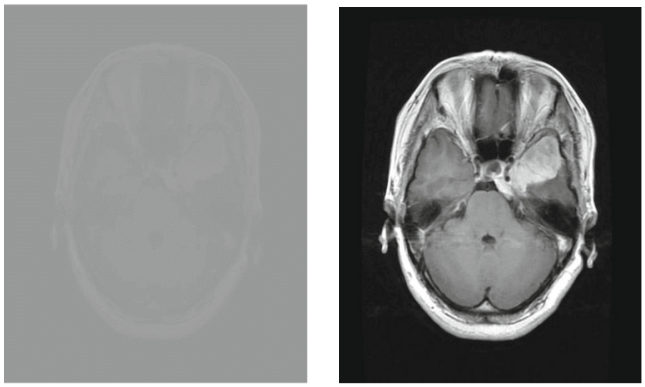
In the field of deep learning (DL) for brain image analysis, data preprocessing is a fundamental step that significantly influences the overall performance of the model. Preprocessing enhances the quality of feature extraction by ensuring that the input data is standardized, thereby allowing neural networks to learn more effectively. Raw brain MRIs often contain intensity values that fall outside the conventional grayscale range of $[0, 255]$, including negative values, due to the physics of MR image acquisition and reconstruction processes. Such irregularities can hinder the convergence of deep models and reduce the discriminative power of extracted features. To mitigate this issue, we adopt the *min–max normalization* technique, which rescales the pixel intensities of each input image into a fixed interval, typically $[0, 1]$. This step not only eliminates the presence of negative and excessively large values but also produces dense, uniform representations of the data that improve both numerical stability and learning efficiency. Furthermore, by standardizing the dynamic range of input images, the method ensures that the contribution of each pixel is balanced during training, which is particularly important when subtle intensity differences are crucial for differentiating tumor regions from normal tissues. The mathematical formulation of min–max normalization is expressed as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{1}$$

where x denotes the original pixel intensity, $x_{\min}$ and $x_{\max}$ represent the minimum and maximum intensity values within the image (or dataset), respectively, and x' is the normalized pixel intensity constrained to the interval $[0, 1]$. In a more generalized form, normalization can be performed over any interval $[a, b]$ as:

$$x' = a + \frac{(x - x_{\min})(b - a)}{x_{\max} - x_{\min}}. \tag{2}$$

In our approach, we specifically adopt the interval $[0, 1]$ since it is well-suited for deep learning models that utilize activation functions such as sigmoid and ReLU. This normalization reduces the likelihood of numerical instability during training, accelerates the convergence of optimization algorithms, and improves the discriminative capability of the network when analyzing brain MRIs for tumor detection and classification. As a result, the contrast of brain edges and regions will be enhanced and improved.Fig.5 shows an example of the input and output of this step.



(a) (b)

Figure 3: The input and output of image preprocessing step; (a) An input image of brain raw data values and (b) An output image of brain data intensity values.

3.3 Proposed Model Architecture

Transfer learning has emerged as an effective technique for addressing classification problems, especially in cases where the available dataset is limited. By reusing knowledge from pre-trained models, significant

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

improvements in classification performance can be achieved. In this work, we propose a hyper-tuned model based on the DenseNet201 architecture. DenseNet201 is chosen due to its compact design, efficient parameter usage, and ability to reuse features across multiple layers. This property enhances the diversity of input to deeper layers, thereby improving the model's discriminative capability and overall accuracy [25]. The general structure of the proposed DenseNet201-based model is illustrated in Fig. 4.

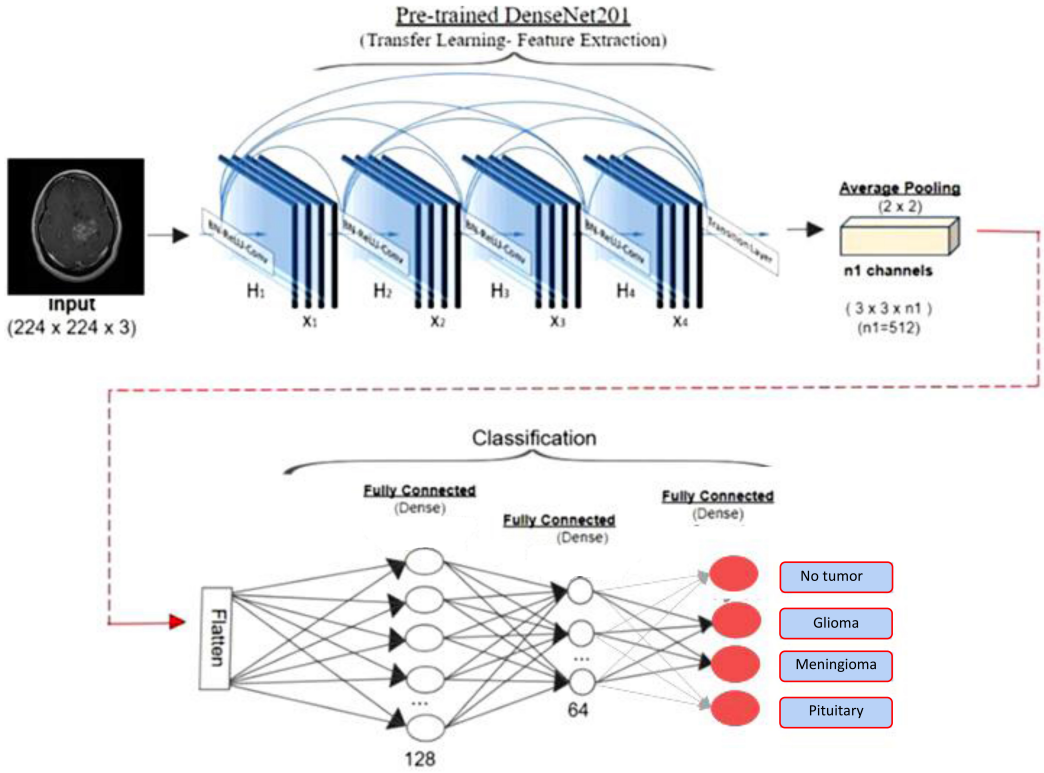


Figure 4: Proposed architecture model.

The workflow of the proposed model is further summarized in Algorithm 1. The process begins with the input of MRI images, followed by rescaling, weight initialization from a pre-trained model, and the integration of a softmax classification layer. The optimization and evaluation steps complete the pipeline.

Algorithm 1 Proposed Model

- **Input:** Image dataset with input shape equal to (224×224×3).
- **Output:** $S_n = \sum_{n=1}^{4}(P(s^{class_n}))$.
- **Begin**
  1. **Read** the images in batches of 5 of size 224×224×3.
  2. **Rescale** all images by a factor of 1/255.
  3. **Import** the weights from the pre-trained model and freeze them.
  4. **Build** the base model using the DenseNet201 architecture.
  5. **Modify** the architecture to adapt it to the specific classification problem.
  6. **Add** a classification layer (softmax) as the output layer.
  7. **Update** the weights using the categorical cross-entropy loss function and SGD optimizer.
  8. **Evaluate** the model using test data and performance metrics.
- **End**

DOI: https://doi.org/10.54216/JAIM.100204
Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

DenseNet201 introduces dense connectivity, where each layer is connected to every other layer in a feed-forward manner. This design alleviates the vanishing gradient problem, encourages feature reuse, and significantly reduces the number of parameters. Fig. 5 demonstrates the dense block structure, where feature maps from all preceding layers are concatenated and passed to subsequent layers [26].

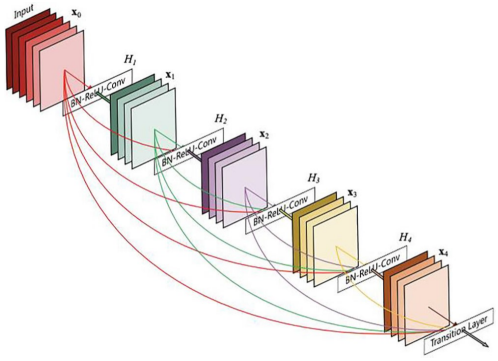


Figure 5: Layered dense block representing direct connections between layers [26].

Mathematically, the transformation within a dense block can be expressed as:

$$x' = H_1\left(\left[x^0, x^1, \ldots, x^{l-1}\right]\right) \tag{3}$$

where H_1 is a non-linear transformation consisting of batch normalization (BN), a rectified linear unit (ReLU), and a 3×3 convolution operation. The term $\left[x^0, x^1, \ldots, x^{l-1}\right]$ denotes the concatenation of feature maps from all preceding layers. Transition layers, composed of BN, 1×1 convolution, and 2×2 average pooling, are introduced between dense blocks to control complexity and improve computational efficiency. The growth rate k serves as a crucial hyperparameter, dictating how many feature maps each layer contributes to the global state. The total number of input feature maps at the l^{th} layer, denoted by $(\mathrm{FM})'$, can be formulated as:

$$(\mathrm{FM})' = \mathrm{k}^0 + \mathrm{k}(l-1) \tag{4}$$

To further increase efficiency, the implementation of bottleneck layers (1×1 convolutions) is used preceding each of the 3×3 convolutions, halving the number of input feature maps. To be classified, there are two dense layers, each with 128 and 64 neurons, respectively, followed by a multi-class prediction softmax layer. The functioning of the fully connected layer may be defined as:

$$\begin{aligned} \mathrm{T}^{l-1} &= \text{Bernoulli}(\mathrm{P}) \\ \ddot{\mathrm{y}}^{l-1} &= \mathrm{T}^{l-1} * c^{l-1} \\ \ddot{\mathrm{y}}' &= \mathrm{f}\left(\mathrm{w}^k \ddot{\mathrm{y}}^{l-1} + \mathrm{o}^1\right) \end{aligned} \tag{5}$$

In this case, the Bernoulli is randomly selected to produce a vector, denoted as T^{l-1}, with a probability distribution of 0 -1 with probability, denoted as P, and $c^{l-1}$is the dimension of the vectors. Several algorithms can be employed to optimize the model. This is the categorical cross-entropy minimization, which is usually represented as:

$$H(p,q) = -\sum_z p(z)\ln(q(z)) \tag{6}$$

where $q(z)$ is the estimated distribution of the true label $p(z)$.

Different optimization strategies have been explored to enhance convergence:

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

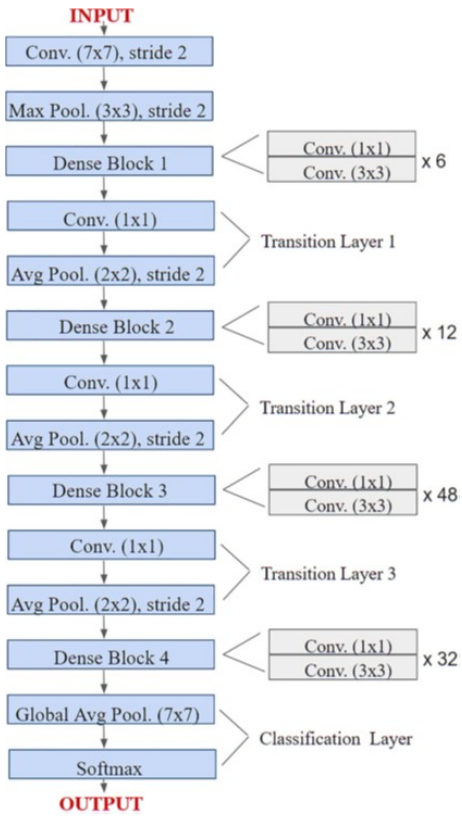


Figure 6: Layered architecture of DenseNet201 [26].

- **Stochastic Gradient Descent with Momentum (SGD)** [27]:

$$\begin{aligned} k_t &= \gamma k_{t-1} + \eta \nabla_\theta H(\theta) \\ \theta_t &= \theta_{t-1} - k_t \end{aligned} \tag{7}$$

This method accelerates convergence by combining gradient descent with a momentum term.

- **RMSProp** [28]:

$$\left.\begin{aligned} E\left|G^2\right|_T &= \gamma E\left|G^2\right|_{T-1} + (1-\gamma) G_T^2 \\ \theta_{T+1} &= \theta_T - \frac{\eta}{\sqrt{E|G^2|_T + \epsilon}} G_T \end{aligned}\right\} \tag{8}$$

RMSProp adapts learning rates by scaling them with a moving average of recent gradients.

- **AdaGrad** [27]:

$$\theta_{T+1,u} = \theta_{T,u} - \frac{\eta}{\sqrt{G_{T,uu} + \epsilon}} G_{T,u} \tag{9}$$

This method adjusts learning rates per parameter, making it particularly effective for sparse features.

- **AdaDelta** [29]:

$$\begin{aligned} \Delta\theta_t &= -\frac{rms[\Delta\theta]_{t-1}}{rms[G]_t} G_t \\ \theta_{t+1} &= \theta_t + \Delta\theta_t \end{aligned} \tag{10}$$

AdaDelta improves AdaGrad by limiting the gradient accumulation window.

- **Adam** [30]:

$$\left.\begin{aligned} L_T &= \beta_1 L_{T-1} + (1-\beta_1) G_T \\ S_T &= \beta_2 S_{T-1} + (1-\beta_2) G_T^2 \end{aligned}\right\} \tag{11}$$

Adam combines the benefits of AdaGrad and RMSProp, offering robust performance with minimal hyperparameter tuning.

DOI: https://doi.org/10.54216/JAIM.100204
Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

Decisions on the hyperparameters in training the proposed DenseNet201-based model are summarized in Table 2. These values were selected through empirical testing to strike a balance between performance and computational efficiency.

Table 2: Hyperparameters of the proposed model for classification

Hyper-parameters	Value	Size of input image
Optimizer	SGD	224×224
Loss	Categorical crossentropy	–
Batch size	5	–
Pooling	GlobalAveragePooling	–
Initial learning rate	3e-4	–
Epochs	36	–
Random state	123	–

4 Experimental Results

This part assesses the framework proposed on the T1-weighted MRI (T1w-MRI) dataset defined in Section 3.1. We have two objectives: (i) to measure the generalization to hidden data and (ii) to estimate multi-class tumor discrimination. Strong assessment entails having a suitable validation protocol as well as informative performance measures. We therefore compare our DenseNet201-based model to strong transfer-learning baselines (InceptionV3, Xception, VGG19, and ResNet50) to make a final classification using the softmax layer of each pre-trained model, thereby creating a similar decision function across models. Experiments were performed in Python 3.6, using Keras 2.2.4, which is backed by TensorFlow 1.13. We used multiple popular optimizers (SGD, RMSProp, Adam, AdaGrad, and AdaDelta) in training each model, while holding other hyperparameters constant, to find the optimizer that best balances convergence rate and generalization. The measurement of accuracy, sensitivity and specificity: We report accuracy, sensitivity and specificity to give augmentary perspective of predictive behavior:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \qquad \text{Sensitivity} = \frac{TP}{TP+FN}, \qquad \text{Specificity} = \frac{TN}{TN+FP}, \quad (12)$$

In which TP, TN, FP and FN represent true positives, true negatives, false positives and false negatives, respectively. The overall performance is captured with accuracy, and the detection of the positive class and rejection of the negative class are characterized by sensitivity and specificity, respectively. Figure 7 shows the training/validation accuracy and loss of the proposed model and the baselines in epochs. Two trends are evident. First, the proposed architecture has a significantly faster convergence than rival networks; second, it achieves a low validation loss at all times, which is not the case with some baselines (in particular, ResNet50). Combined, these dynamics signify efficient optimization and generalization that is favorable. The proposed method achieves a quantitative accuracy of 100% on the training, validation, and test splits.

The end-of-training metrics and processing times are summarized in Table 3. The offered model has a very high score in all criteria and a significantly low processing time (P_t=5.4 s). Although VGG19 and Xception are both competitive in terms of accuracy, they may be significantly more computationally expensive. ResNet50 does not perform well in terms of accuracy and particularly sensitivity, making it challenging to predict positive cases on this dataset accurately. To isolate the contribution of architectural decisions, we ablated the GAP

Table 3: Comparison of the proposed model with transfer-learning baselines under identical hyperparameters. P_t denotes processing time (seconds).

Model	Train acc.	Test acc.	Sensitivity	Specificity	P_t(sec)
Proposed	100%	100%	100%	100%	5.4
Xception	87.3%	96.3%	98.2%	96.3%	3.9
InceptionV3	87.6%	90.3%	91.0%	94.6%	4.2
VGG19	95.6%	96.0%	97.0%	95.8%	1390.7
ResNet50	52.9%	54.7%	15.0%	92.6%	1657.6

DOI: https://doi.org/10.54216/JAIM.100204
Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

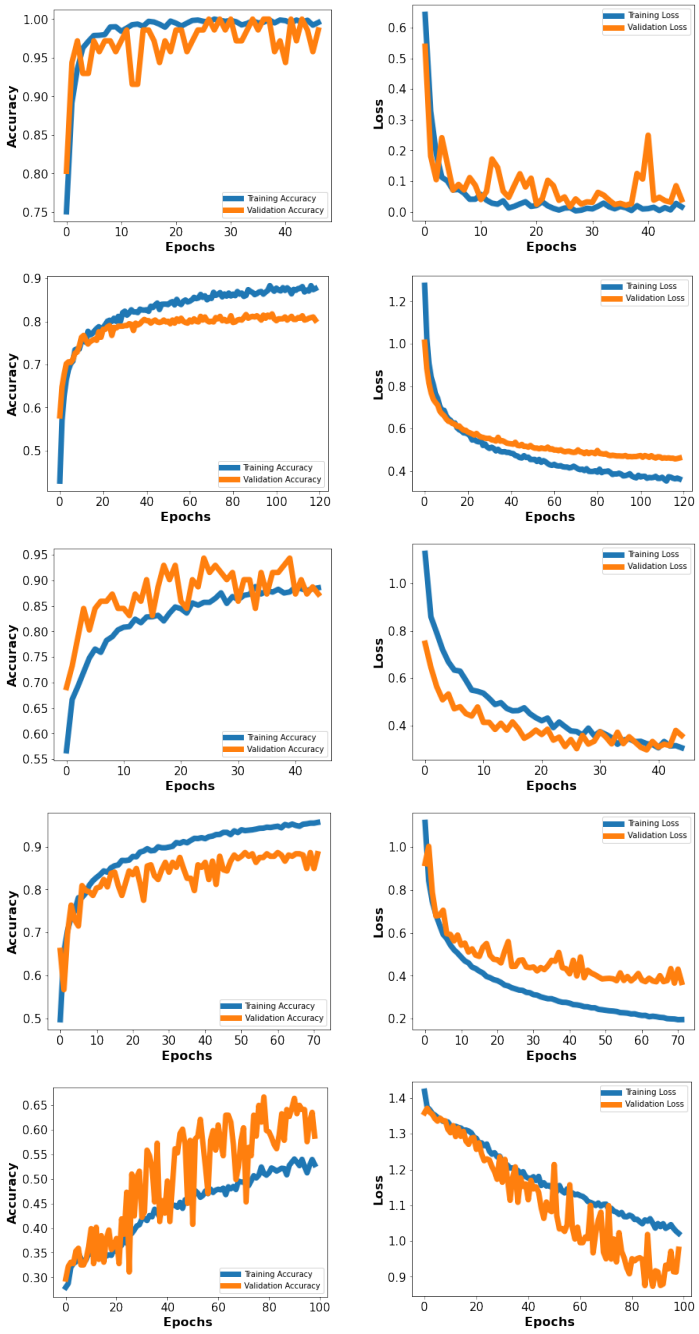


Figure 7: Training versus validation accuracy (left column) and loss (right column) over epochs for the proposed model (top) and baselines: InceptionV3, Xception, VGG19, and ResNet50.

layer. As indicated in Table 4, the introduction of GAP significantly increases accuracy and sensitivity, while reducing overfitting, compared to the DenseNet201 baseline without GAP, which suffers on both the training and testing splits. The fact that GAP was one of the main contributors to stable generalization in our pipeline is indicated by this ablation.

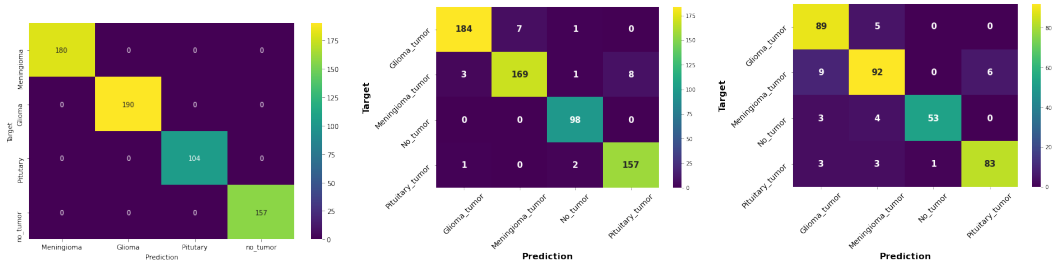
Table 4: Ablation study: DenseNet201 with and without global average pooling (GAP).

Model	Train acc.	Test acc.	Sensitivity	Specificity
Proposed model	100%	100%	100%	100%
DenseNet201 (w/o GAP)	79.5%	78.3%	68.9%	98.7%

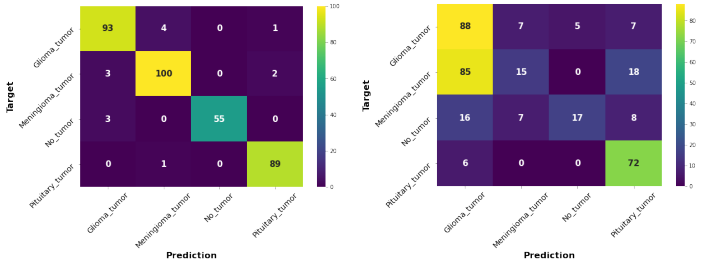
Figure 8shows class-wise predictions through confusion matrices. The suggested model has an ideal diagonal

DOI: https://doi.org/10.54216/JAIM.100204
Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

structure in all types of tumors, and the baseline models have different levels of confusion (once again, the most evident is ResNet50). This supports the fact that in our designs, we are making valid discrimination on a class level.



(a) Proposed (b) Xception (c) InceptionV3



(d) VGG19 (e) ResNet50

Figure 8: Confusion matrices for the proposed model and baselines. Perfect diagonal dominance for the proposed method indicates error-free class assignments.

We next examine the optimizer's role in convergence and final performance. Figure 9 shows learning curves for Adam, RMSProp, AdaGrad, AdaDelta, and SGD. Visually, SGD yields the most stable descent with the fastest approach to low validation loss. Tables 5 and 6 corroborate these observations numerically: SGD attains the lowest loss and perfect accuracy, while alternative adaptive methods are competitive but generally less consistent under our setting.

Table 5: Final losses for different optimizers.

Optimizer	Loss
SGD	0.01
RMSProp	0.14
AdaGrad	0.04
Adam	0.05
AdaDelta	0.20

Table 6: Performance of the proposed architecture under different optimizers.

Optimizer	Train acc.	Test acc.	Sensitivity	Specificity
Adam	98.4%	99.8%	100%	100%
RMSProp	97.8%	95.5%	93.3%	100%
AdaGrad	98.7%	100%	100%	100%
AdaDelta	91.9%	98.2%	100%	99.4%
SGD	100%	100%	100%	100%

Lastly, we situate our approach within the existing literature. Table 7 presents the results of representative methods in terms of accuracy (and, where possible, sensitivity/specificity). It builds upon previous work in classical ML pipelines and deep transfer learning, with reported accuracy ranging between 91.9% and 98.87%. When using the same data and assessment parameters, our DenseNet201+GAP structure achieves a 100% score in all measures, representing a significant improvement over existing systems.

DOI: https://doi.org/10.54216/JAIM.100204
Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

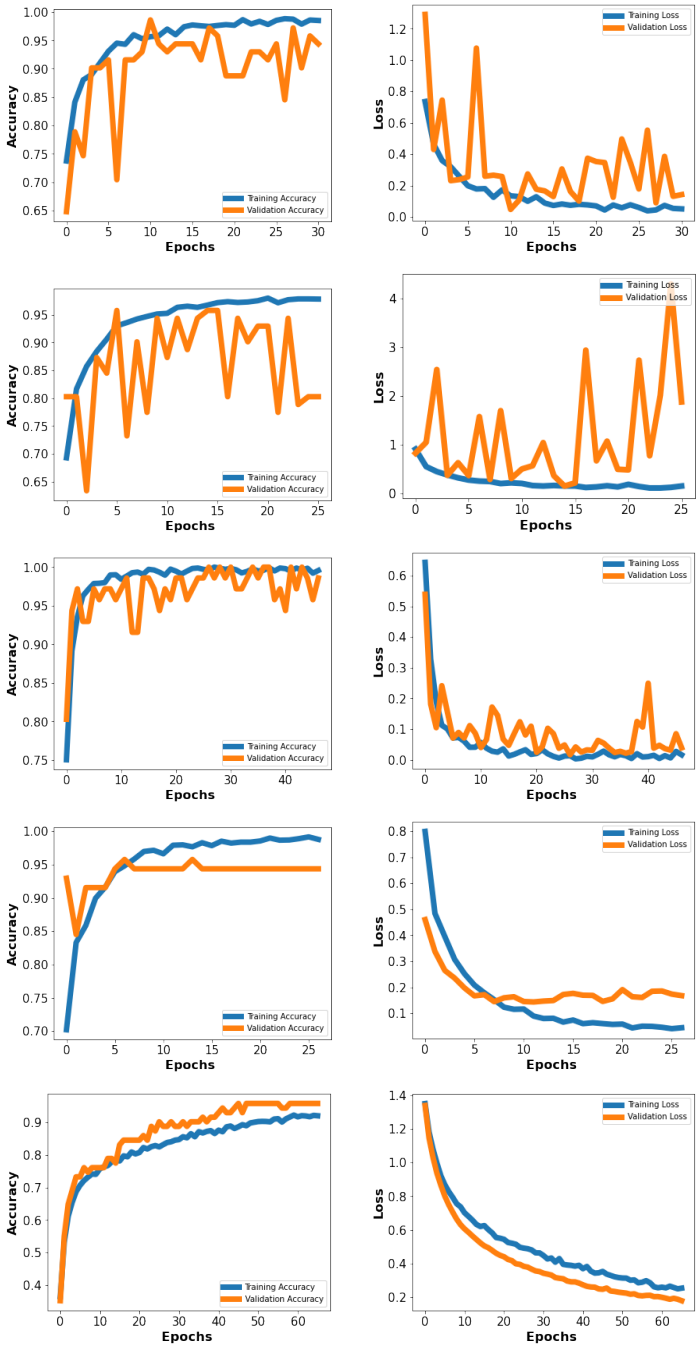


Figure 9: Training (left) and validation (right) accuracy/loss across optimizers: Adam, RMSProp, SGD, AdaGrad, and AdaDelta. SGD exhibits the fastest stable convergence and lowest final error.

5 Conclusion

This paper presents a deep neural network architecture that utilizes transfer learning to classify brain tumors on a multi-label basis. We utilized the DenseNet-201 base model, which features dense connectivity and practical feature propagation, thereby facilitating the derivation of both rich and discriminative hidden features from medical images. Specially created top layers later inherited these characteristics, including Global Average Pooling (GAP), fully connected layers, and a softmax classifier, to classify the input tumor images into three broad categories: glioma, meningioma, and pituitary tumors. The proposed framework was capable of reflecting minor spatial changes among tumor classes, as it adopted a design that increased the strength and accuracy of classification. The findings established the effectiveness of our model over other methods

DOI: https://doi.org/10.54216/JAIM.100204
Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

Table 7: Comparison with state-of-the-art models.

Ref	Model	Accuracy	Sensitivity	Specificity
[12]	SVM	98.87%	NA	NA
[13]	SVM	95.23%	NA	NA
[15]	DWT–Gabor–NN	91.9%	NA	NA
[22]	RELM	94.23%	NA	NA
[23]	CNN+GA	94.20%	NA	NA
[20]	Pretrained VGG19	94.82%	93.0%	94.6%
[19]	GoogleNet+TL	97.1%	NA	NA
Proposed model	DenseNet201+GAP	100%	100%	100%

of transfer learning and state-of-the-art deep neural network designs reported in the literature. One of the fundamental aspects of our study was the investigation of a range of network parameters, as well as the effects of different optimizers on training performance, in particular. Through extensive experimentation, it was observed that the selection of the optimizer had a significant impact on convergence behavior and model generalization. Out of the optimizers evaluated, Stochastic Gradient Descent (SGD) proved to be the most efficient, with the lowest loss rates and providing an excellent accuracy of 100 percent on the dataset used. The result underscores the optimizer's capability to balance gradient updates and achieve better stability and performance. The fact that the experiment achieved perfect accuracy highlights the power of well-tuned transfer learning pipelines in medical image analysis, where even a minor increase in prediction power can translate into significant clinical benefits to aid in diagnosis. In the future, we recognize that our work can be further developed to enhance its clinical applicability. The following steps will include incorporating more classes and subtypes of tumors to develop a more complex and easier-to-use diagnostic tool that will provide the ability to cope with the heterogeneity of brain tumors. In addition, it may be possible to discover new performance improvements by examining other deep architecture designs, such as Vision Transformers, CNN-Transformer hybrids, and more efficient designs like EfficientNet. Another important measure will be testing the model on larger and more heterogeneous multi-center data to ensure the model's strength across various imaging regimens and populations. Besides increasing the diagnostic potential of the system, these expansions will also make it more resilient to integration into the real-world clinical workflow, thereby benefiting the application of precision medicine and patient care in the field of neuro-oncology.

References

[1] Y.-K. Kim and J. Song, "Metabolic imbalance and brain tumors: The interlinking metabolic pathways and therapeutic actions of antidiabetic drugs," *Pharmacological Research*, vol. 215, p. 107 719, 2025, ISSN: 1043-6618. DOI: https://doi.org/10.1016/j.phrs.2025.107719. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1043661825001446.

[2] A. Behin, K. Hoang-Xuan, A. F. Carpentier, and J.-Y. Delattre, "Primary brain tumours in adults," *The Lancet*, vol. 361, no. 9354, pp. 323–331, 2003.

[3] A. Hekmat, Z. Zhang, S. Ur Rehman Khan, I. Shad, and O. Bilal, "An attention-fused architecture for brain tumor diagnosis," *Biomedical Signal Processing and Control*, vol. 101, p. 107 221, 2025, ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2024.107221. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424012795.

[4] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "Cbtrus statistical report: Primary brain and other central nervous system tumors diagnosed in the united states in 2013–2017," *Neuro-oncology*, vol. 22, no. Supplement_1, pp. iv1–iv96, 2020.

[5] K. R. Porter, B. J. McCarthy, S. Freels, Y. Kim, and F. G. Davis, "Prevalence estimates for primary brain tumors in the united states by age, gender, behavior, and histology," *Neuro-oncology*, vol. 12, no. 6, pp. 520–527, 2010.

[6] H. Yadav, S. Singh, K. K. Mishra, S. Srivastava, M. S. Naruka, and S. P. Yadav, "Brain tumor detection with mri images," in *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 2022, pp. 519–527. DOI: 10.1109/CISES54857.2022.9844387.

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

[7] V. Yamuna, P. RVS, R. Sathya, M. Dhivva, R. Lidiya, and P. Sowmiya, "Integrating ai for improved brain tumor detection and classification," in *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, 2024, pp. 1603–1609. DOI: 10.1109/ICSES63445.2024.10763262.

[8] M. Agarwal, G. Rani, A. Kumar, P. Kumar, R Manikandan, and A. H. Gandomi, "Deep learning for enhanced brain tumor detection and classification," *Results in Engineering*, vol. 22, p. 102 117, 2024. DOI: https://doi.org/10.1016/j.rineng.2024.102117.

[9] A. Batool and Y.-C. Byun, "Brain tumor detection with integrating traditional and computational intelligence approaches across diverse imaging modalities - challenges and future directions," *Computers in Biology and Medicine*, vol. 175, p. 108 412, 2024, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2024.108412. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482524004967.

[10] S. Khalighi, K. Reddy, A. Midya, K. B. Pandav, A. Madabhushi, and M. Abedalthagafi, "Artificial intelligence in neuro-oncology: Advances and challenges in brain tumor diagnosis, prognosis, and precision treatment," *NPJ precision oncology*, vol. 8, no. 1, p. 80, 2024. DOI: https://doi.org/10.1038/s41698-024-00575-0.

[11] J. Cheng, *Brain tumor dataset. figshare. dataset (2017)*, 2017.

[12] V. Rathi and S. Palani, "Brain tumor detection and classification using deep learning classifier on mri images," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 10, pp. 177–187, May 2015.

[13] A. Kumar, M. A. Ansari, and A. Ashok, "A hybrid framework for brain tumor classification using grey wolf optimization and multi-class support vector machine 7747."

[14] J. Cheng et al., "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PloS one*, vol. 10, no. 10, e0140381, 2015.

[15] M. R. Ismael and I. Abdel-Qader, "Brain tumor classification via statistical features and back-propagation neural network," in *2018 IEEE international conference on electro/information technology (EIT)*, IEEE, 2018, pp. 0252–0257.

[16] T. A. Abir, J. A. Siraji, E. Ahmed, and B. Khulna, "Analysis of a novel mri based brain tumour classification using probabilistic neural network (pnn)," *Int. J. Sci. Res. Sci. Eng. Technol*, vol. 4, no. 8, pp. 65–79, 2018.

[17] P. Afshar, K. N. Plataniotis, and A. Mohammadi, "Capsule networks for brain tumor classification based on mri images and coarse tumor boundaries," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 1368–1372.

[18] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in *World congress on medical physics and biomedical engineering 2018*, Springer, 2019, pp. 183–189.

[19] S Deepak and P. Ameer, "Brain tumor classification using deep cnn features via transfer learning," *Computers in biology and medicine*, vol. 111, p. 103 345, 2019.

[20] Z. N. K. Swati et al., "Content-based brain tumor retrieval for mr images using transfer learning," *IEEE Access*, vol. 7, pp. 17 809–17 822, 2019.

[21] E.-S. A. El-Dahshan, T. Hosny, and A.-B. M. Salem, "Hybrid intelligent techniques for mri brain images classification," *Digital signal processing*, vol. 20, no. 2, pp. 433–441, 2010.

[22] A. Gumaei, M. M. Hassan, M. R. Hassan, A. Alelaiwi, and G. Fortino, "A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification," *IEEE Access*, vol. 7, pp. 36 266–36 273, 2019.

[23] A. K. Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *biocybernetics and biomedical engineering*, vol. 39, no. 1, pp. 63–74, 2019.

[24] J. Cheng, *Brain tumor dataset. figshare. dataset*, 2018.

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025

[26] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the covid-19 infected patients using densenet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 15, pp. 5682–5689, 2021, PMID: 32619398. DOI: 10.1080/07391102.2020.1788642. eprint: https://doi.org/10.1080/07391102.2020.1788642. [Online]. Available: https://doi.org/10.1080/07391102.2020.1788642.

[27] M. Yaqub et al., "State-of-the-art cnn optimizer for brain tumor segmentation in magnetic resonance images," *Brain Sciences*, vol. 10, no. 7, 2020, ISSN: 2076-3425. DOI: 10.3390/brainsci10070427. [Online]. Available: https://www.mdpi.com/2076-3425/10/7/427.

[28] R. Mehrotra, M. Ansari, R. Agrawal, and R. Anand, "A transfer learning approach for ai-based classification of brain tumors," *Machine Learning with Applications*, vol. 2, p. 100 003, 2020.

[29] M. D. Zeiler, *Adadelta: An adaptive learning rate method*, 2012. arXiv: 1212.5701 [cs.LG].

[30] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: 1412.6980 [cs.LG].

DOI: https://doi.org/10.54216/JAIM.100204

Received: April 18, 2025 Revised: July 05, 2025 Accepted: September 10, 2025