Journal of Artificial Intelligence and Metaheuristic

♦ ISSN: 2833-5597 Vol: 10, No. 02, 2025 ♦ PP. 27–35

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



FULL LENGTH ARTICLE

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: dskhafaga@pnu.edu.sa

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

ABSTRACT

The classification of brain tumors is crucial in the context of early intervention, as appropriate and timely diagnosis can significantly influence treatment plans and patient outcomes. Radiologists have long relied on their own judgment to read medical images, a process that can be subjective and time-consuming and can lead to inter-observer variability. Artificial intelligence (AI), and particularly deep learning (DL), has transformed medical imaging and could automate diagnosis by providing prompt, reliable, and unbiased assessments. Despite this progress, many AI-based systems remain constrained in their classification accuracy and robustness across datasets. This paper presents a deep-learning model designed to improve brain-tumor classification. Its preprocessing and normalization steps reduce noise and equalize image intensity to support feature extraction. The proposed method adapts DenseNet201 for four-class classification and replaces its conventional output layer with Global Average Pooling (GAP), aiming to address vanishing gradients and overfitting while improving feature aggregation and generalization. The model was evaluated on a public Figshare dataset and a local dataset comprising 3,504 T1-weighted contrast-enhanced MRI images. It achieved 100% overall accuracy, outperforming the comparison methods. These results suggest that architectural modifications combined with preprocessing can support automated brain-tumor classification under the evaluated conditions.

Keywords: Brain tumor ▪ Deep learning ▪ DenseNet201 ▪ Multi-class classification ▪ Transfer learning

1. INTRODUCTION

A brain tumor is an abnormal proliferation of cells within the brain and may be malignant or benign [1]. Malignant tumors can expand uncontrollably, infiltrate nearby tissues, and interfere with normal brain function, whereas benign tumors generally remain localized. Even benign tumors can cause severe complications because of the confined space of the cranial cavity. Brain tumors therefore represent a diverse class of diseases with different biological behaviors, prognoses, and treatment responses. They are categorized by histological features and clinical manifestations, making accurate diagnosis and classification critical to treatment planning [2]. Multi-class classification can identify the specific tumor type, with implications for treatment strategies and patient outcomes. Common types include meningiomas, gliomas, and pituitary tumors, which arise from the meninges, glial cells, and pituitary region, respectively [3].

The epidemiological burden underscores the need for improved diagnostic systems. The Central Brain Tumor Registry of the United States reported that an estimated 700,000 Americans live with a brain tumor, with almost 84,000 new primary brain-tumor cases diagnosed in 2021 [4]. Brain tumors also affect children, and primary malignant brain tumors cause substantial mortality [5]. More than 120 primary brain

© 2025 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

and central nervous system tumor types have been identified, nearly one-third of which are malignant. Radiological classification is challenging because tumors are heterogeneous and may have overlapping imaging characteristics, including isointense and hypointense regions and perilesional edema. T1-weighted contrast-enhanced MRI is widely used for tumor imaging; contrast agents such as gadolinium can improve lesion visibility and support diagnosis of primary and secondary malignancies [6].

Tumor classification cannot rely solely on manual radiological assessment. Traditional methods often use handcrafted feature extraction and region-based segmentation, which require considerable effort and can be affected by inter-observer variation [7]. Deep-learning models can learn hierarchical feature representations directly from image data and build increasingly abstract, discriminative patterns through nonlinear transformations [8]. Transfer learning with pretrained networks can also support training on comparatively small medical datasets. Deep-learning computer-aided diagnosis systems may therefore help reduce radiologists' workload by providing rapid, objective, and dependable classifications that support early diagnosis and patient management [9, 10].

This study develops an automated brain-tumor detection and classification system using publicly available and locally collected MRI data. Its contributions are: (i) a modified DenseNet201 transfer-learning framework that incorporates a Global Average Pooling (GAP) layer to address overfitting and vanishing-gradient issues; (ii) the use of pretrained ImageNet weights for robust feature extraction; and (iii) extensive experiments comparing the proposed model with state-of-the-art methods. The remainder of the paper reviews related work, describes the methodology, reports experimental results, and concludes with future research directions.

2. RELATED WORK

Recent research on brain-tumor detection and classification has used both traditional machine-learning techniques and deep-learning architectures. Public Figshare data and locally collected images have been used to evaluate these approaches [11].

Conventional machine-learning methods commonly rely on feature engineering and dimensionality reduction. A Support Vector Machine (SVM) combined with Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) has been used to classify brain tissues, including white matter, gray matter, cerebrospinal fluid, and abnormal or normal regions. Using handcrafted intensity, texture, and shape features, this method reported 98.87% accuracy, although PCA may remove discriminative texture information [12]. A hybrid Grey Wolf Optimization (GWO) and multiclass SVM approach achieved 95.23% accuracy and outperformed other metaheuristics in the reported experiment [13]. Other work applied adaptive spatial and ring-form partitioning to regions of interest, extracting intensity histograms, gray-level co-occurrence matrices, and bag-of-words features; reported accuracy ranged from 87.54% to 91.28% [14]. Wavelet and Gabor features used with a back-propagation multilayer perceptron achieved 91.9% accuracy [15], while gray-level co-occurrence matrix features combined with a probabilistic neural network achieved 83.33% [16]. These studies demonstrate the importance of feature extraction in brain MRI classification.

Deep-learning methods learn image features directly rather than relying on manually designed descriptors. Capsule networks intended to preserve spatial hierarchies of tumor structures reported 90.89% accuracy [17]. In a comparison of CNN models, a design using ReLU activations and max-pooling reported 98.5% training accuracy and 84.19% validation accuracy [18]. Transfer-learning studies reported 98% patient-level accuracy on Figshare data [19], while a fine-tuned VGG19 model achieved 94.82% [20]. Hybrid Artificial Neural Network (ANN) and k-nearest-neighbor (KNN) methods have also reported accuracies of 97% and 98%, respectively [21]. A hybrid method using normalized GIST descriptors, PCA, and a regularized extreme learning machine achieved 94.23% accuracy [22]; genetic-algorithm-optimized CNNs have reported classification accuracy of up to 94% [23].

Conventional CNNs can encounter vanishing gradients when training deep networks. DenseNet201 addresses this issue through dense connections, and prior work has combined pretrained DenseNet201 models with global average pooling to improve training and validation performance [24, 25]. The present study builds on this direction using a modified DenseNet201 framework and evaluates it against several transfer-learning baselines.

3. MATERIALS AND METHODS

The analysis pipeline comprises preprocessing and normalization, CNN training and parameter optimization, and multiclass classification. Figure 1 summarizes the proposed methodology.

3.1 Dataset

The study combines publicly available data provided by Cheng [11, 26] with a locally acquired dataset from the Department of Diagnostic Radiology at Mansoura University, Egypt. Together, the datasets contain 3,504 brain MR images: 1,003 glioma, 981 meningioma, 959 pituitary, and 561 no-tumor images. The original images are 512×512 pixels. Table 1 summarizes the class distribution, and Figure 2 shows examples.

Table 1. Dataset distribution for each class.

Class	Images
Glioma	1,003
Meningioma	981
Pituitary	959
No tumor	561
Total	3,504

3.2 Data Preprocessing

Preprocessing standardizes MRI inputs so that the neural network can learn from consistent data. Raw MR images may contain intensity values outside the conventional grayscale range of [0,255], including negative values. We apply min–max normalization to rescale image intensities to [0,1], re-

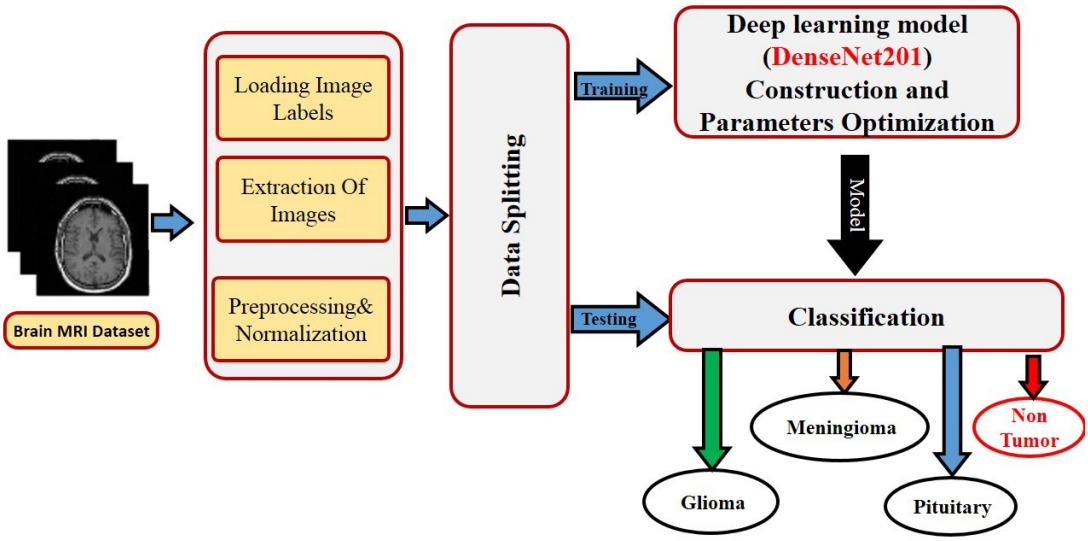
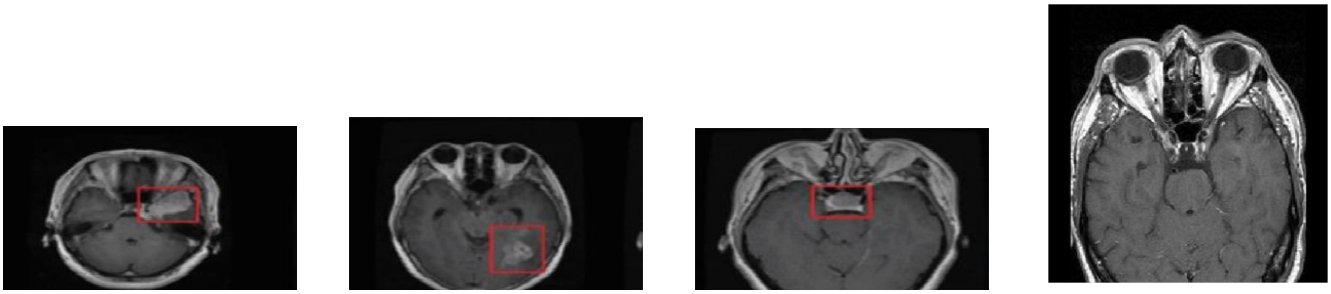


Figure 1. Overview of the proposed methodology.



(a) Meningioma **(b)** Glioma **(c)** Pituitary **(d)** No tumor

Figure 2. Examples of axial T1-weighted contrast-enhanced brain images from the dataset.

moving negative and excessively large values and improving numerical stability. The normalization is defined as

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

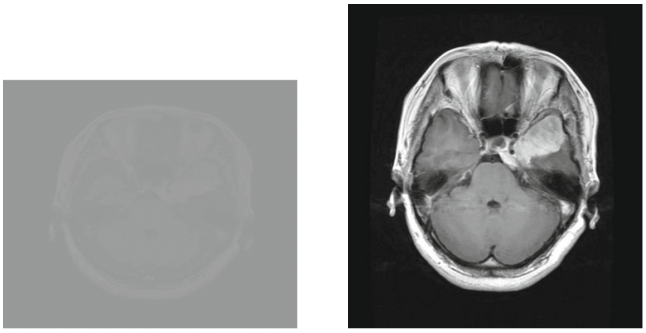
where x is the original pixel intensity, $x_{\min}$ and $x_{\max}$ are the minimum and maximum intensity values, and x' is the normalized intensity. More generally, normalization over any interval $[a,b]$ is

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

The interval [0,1] is used in this study. It is compatible with deep-learning models using activation functions such as sigmoid and ReLU, and it can improve optimization convergence and the distinction between tumor regions and normal tissue. Figure 3 shows an example of this step.

3.3 Proposed Model Architecture

Transfer learning can improve classification when the available dataset is limited. We use a hyper-tuned model based on DenseNet201 because its compact design reuses features across layers, increasing the diversity of inputs to deeper layers and improving discriminative capability [24]. The proposed model architecture is shown in Figure 4.



(a) Raw image **(b)** Normalized image

Figure 3. Input and output of the image-preprocessing step.

Algorithm 1. Proposed Model

- **Input:** Image dataset with input shape $(224 \times 224 \times 3)$.
- **Output:** Probability scores for four classes, $P(s_{\text{class}_n})$, $n = 1,\ldots,4$.

Begin

1. Read images in batches of 5 at size $224 \times 224 \times 3$.
2. Rescale all images by a factor of $1/255$.
3. Import and freeze the weights from the pretrained model.
4. Build the base model using the DenseNet201 architecture.
5. Modify the architecture for the specific classification problem.
6. Add a softmax classification layer as the output layer.

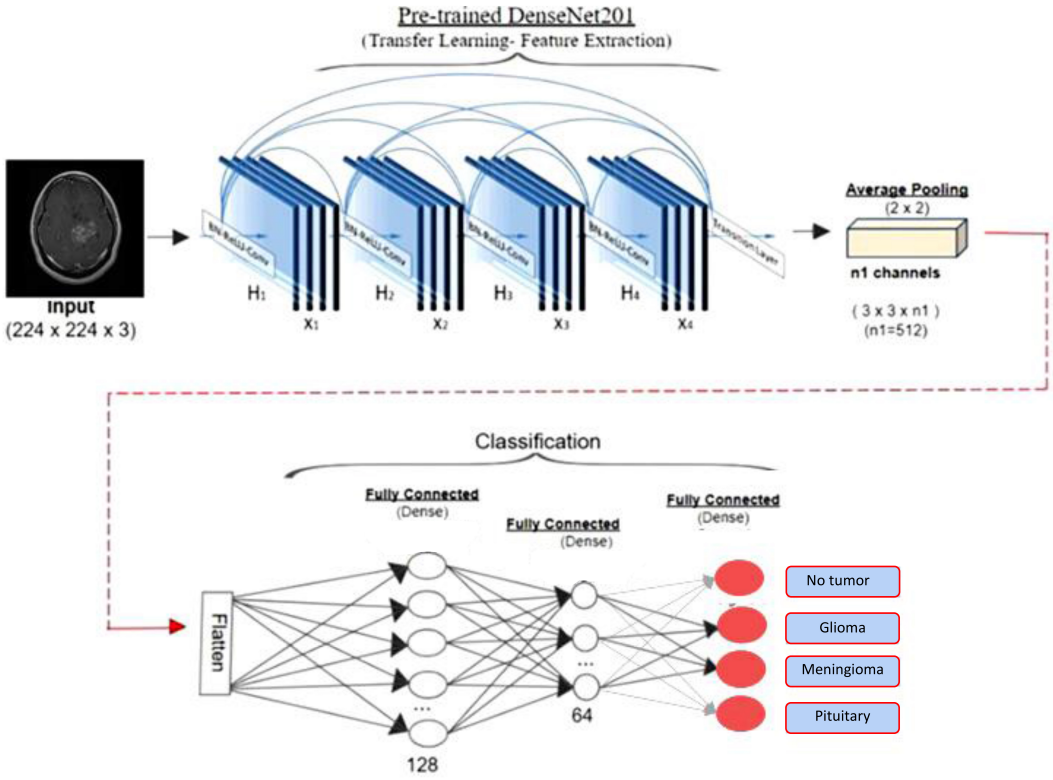


Figure 4. Proposed model architecture.

7. Update the weights using categorical cross-entropy and the SGD optimizer.
8. Evaluate the model using test data and performance metrics.

End

DenseNet201 uses dense connectivity, in which each layer connects to every subsequent layer in a feed-forward manner. This design alleviates vanishing gradients, encourages feature reuse, and reduces the number of parameters. Feature maps from preceding layers are concatenated and passed to subsequent layers, as illustrated in Figure 5 [25].

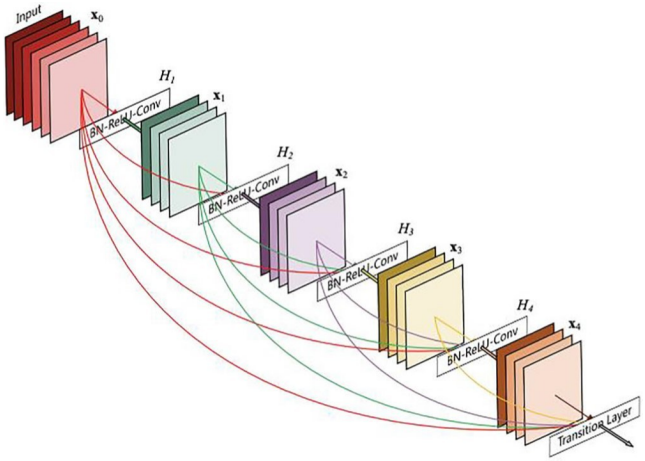


Figure 5. Layered dense block representing direct connections between layers [25].

The transformation within a dense block can be expressed as

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

where H_l is a nonlinear transformation consisting of batch normalization, a rectified linear unit (ReLU), and a 3×3 convolution. The concatenated term contains feature maps from all preceding layers. Transition layers, comprising batch normalization, a 1×1 convolution, and 2×2 average pooling, control complexity and computational cost. The growth rate k determines the number of feature maps added by each layer. The total number of input feature maps at layer l is

$$(\mathrm{FM})'_l = k_0 + k(l-1). \tag{4}$$

Bottleneck layers using 1×1 convolutions precede the 3×3 convolutions to reduce the number of input feature maps. The classifier has two dense layers with 128 and 64 neurons, respectively, followed by a multiclass softmax layer. The fully connected layer is described by

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

Here, the Bernoulli variable generates a vector T^{l-1} with probability P, and c^{l-1} denotes the dimension of the vectors. The model is optimized by minimizing categorical cross-entropy:

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

where $q(z)$ is the estimated distribution of the true label $p(z)$. Figure 6 shows the layered DenseNet201 architecture [25].

Different optimization strategies were considered to improve convergence.

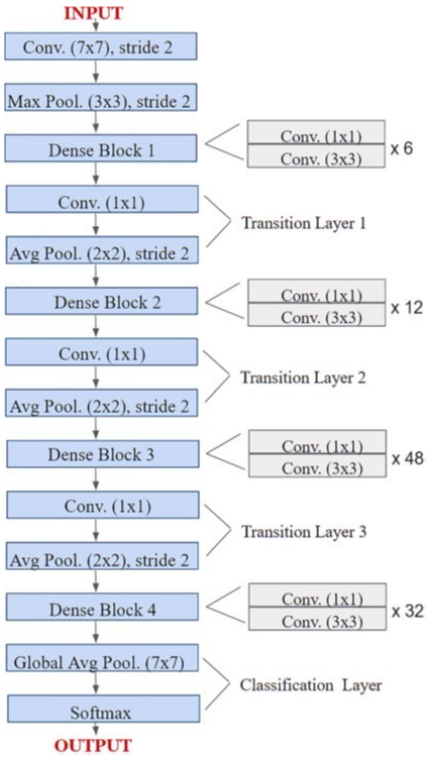


Figure 6. Layered architecture of DenseNet201 [25].

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]:

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

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

AdaGrad [27]:

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

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

AdaDelta [29]:

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

AdaDelta improves AdaGrad by limiting the gradient accumulation window.

Adam [30]:

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

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

Decisions on the hyperparameters for the proposed DenseNet201-based model are summarized in Table 2. These values were selected empirically to balance performance and computational efficiency.

Table 2. Hyperparameters of the proposed classification model.

Hyperparameter	Value	Input size
Optimizer	SGD	224×224
Loss	Categorical cross-entropy	–
Batch size	5	–
Pooling	Global Average Pooling	–
Initial learning rate	3×10^{-4}	–
Epochs	36	–
Random state	123	–

4. EXPERIMENTAL RESULTS

We evaluate the proposed framework on the T1-weighted MRI dataset described in Section 3.1. The objectives are to measure generalization to hidden data and assess multiclass tumor discrimination. We compare DenseNet201 with InceptionV3, Xception, VGG19, and ResNet50, using the softmax layer of each pretrained model for classification. Experiments were performed in Python 3.6 with Keras 2.2.4 and TensorFlow 1.13. The optimizers SGD, RMSProp, Adam, AdaGrad, and AdaDelta were evaluated while other hyperparameters were held constant.

Accuracy, sensitivity, and specificity are reported to characterize predictive performance:

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN}, \\ \text{Sensitivity} &= \frac{TP}{TP+FN}, \\ \text{Specificity} &= \frac{TN}{TN+FP}. \end{aligned} \tag{12}$$

Here, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Accuracy summarizes overall performance, while sensitivity and specificity measure positive-class detection and negative-class rejection. Figure 7 shows training and validation accuracy and loss by epoch. The proposed architecture converges faster than the comparison networks and maintains low validation loss; it achieves 100% accuracy on the training, validation, and test splits.

The end-of-training metrics and processing times are summarized in Table 3. The proposed model achieves 100% across the reported metrics with a processing time of 5.4 seconds. Although VGG19 and Xception are competitive in accuracy, they may be more computationally expensive. ResNet50 reports lower accuracy and sensitivity. Table 4 isolates the contribution of GAP: adding GAP improves accuracy and sensitivity and reduces overfitting compared with DenseNet201 without GAP.

Figure 8 shows the class-wise predictions. The proposed model has a perfect diagonal in its confusion matrix, while the baselines show varying levels of confusion, most notably

Table 3. Comparison of the proposed model with transfer-learning baselines under identical hyperparameters. Processing time is reported in seconds.

Model	Train accuracy	Test accuracy	Sensitivity	Specificity	Time (s)
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

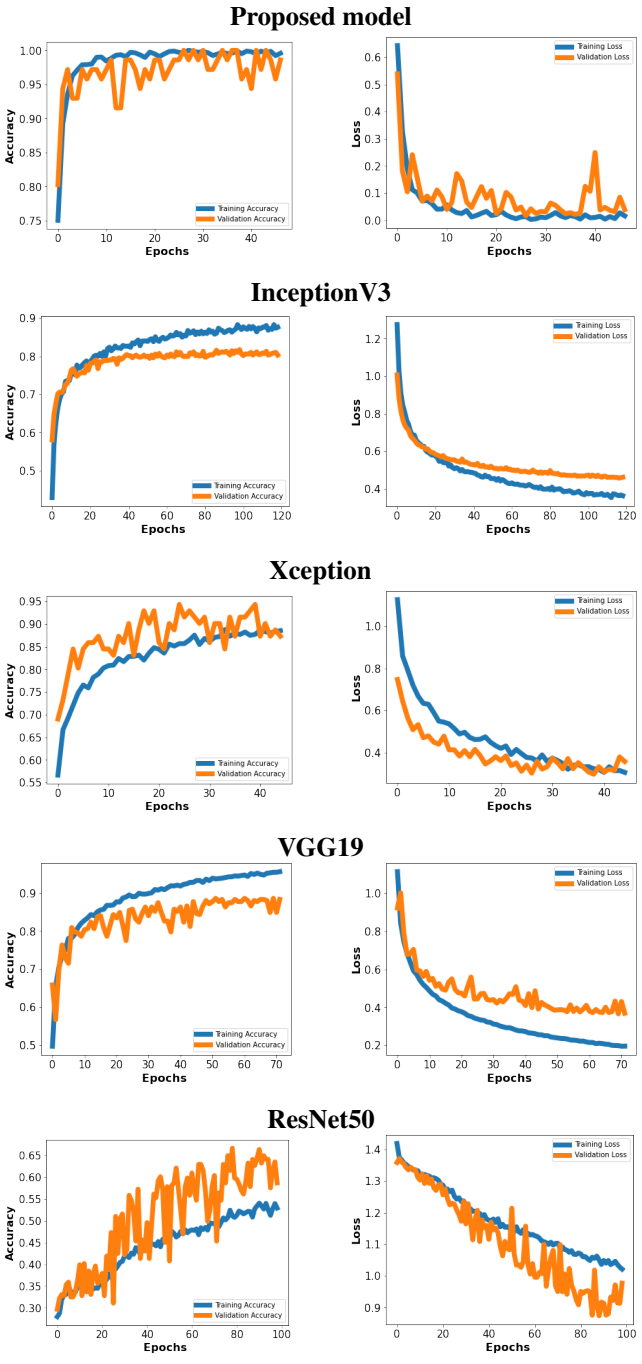
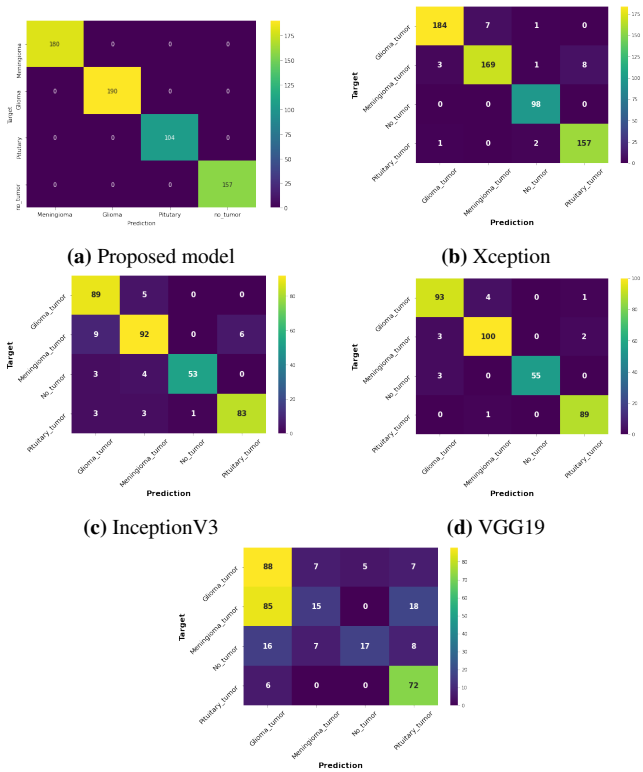


Figure 7. Training and validation accuracy (left) and loss (right) over epochs for the proposed model and transfer-learning baselines.

Table 4. Ablation study of DenseNet201 with and without Global Average Pooling (GAP).

Model	Train	Test	Sens.	Spec.
Proposed model	100%	100%	100%	100%
DenseNet201 (without GAP)	79.5%	78.3%	68.9%	98.7%



(a) Proposed model **(b)** Xception

(c) InceptionV3 **(d)** VGG19

(e) ResNet50

Figure 8. Confusion matrices for the proposed model and transfer-learning baselines. The proposed model's perfect diagonal indicates error-free class assignments in this evaluation.

ResNet50.

We also examine the optimizers' effects on convergence and final performance. Figure 9 shows their learning curves. SGD gives the most stable descent and fastest approach to low validation loss. Tables 5 and 6 show that SGD achieves the lowest loss and perfect accuracy; the other optimizers are competitive but less consistent in this 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 7 compares the proposed model with representative methods from the literature. Reported accuracies range from 91.9% to 98.87%; the DenseNet201+GAP model achieves 100% for accuracy, sensitivity, and specificity in the present

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

Optimizer	Train	Test	Sens.	Spec.
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%

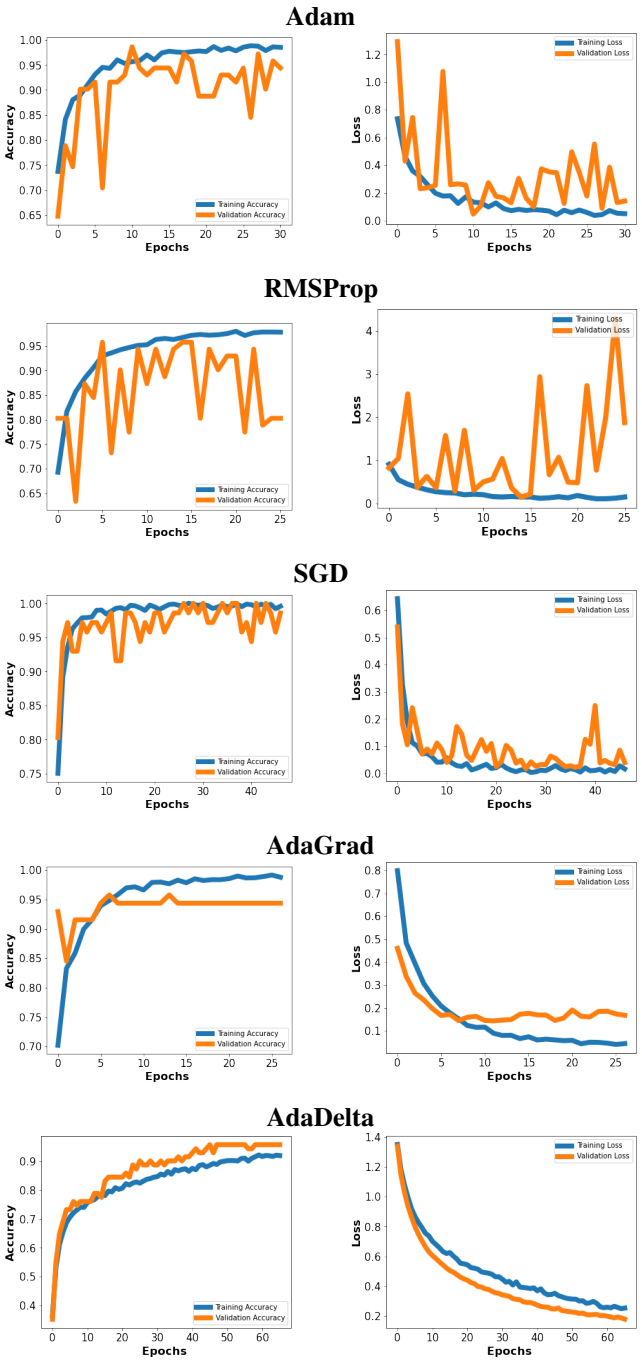


Figure 9. Training (left) and validation (right) accuracy and loss across optimizers. SGD shows stable convergence and low final error.

evaluation.

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

Ref.	Model	Accuracy	Sens.	Spec.
[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	DenseNet201+GAP	100%	100%	100%

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 of transfer learning and state-of-the-art deep neural network designs reported in the literature.

One of the fundamental aspects of this study was the investigation of a range of network parameters, particularly the effects of different optimizers on training performance. Through extensive experimentation, optimizer selection was found to have a significant impact on convergence behavior and model generalization. Among the optimizers evaluated, Stochastic Gradient Descent (SGD) proved most efficient, achieving the lowest loss and 100% accuracy on the dataset. This result underscores the optimizer's ability to balance gradient updates and achieve stable performance. The perfect accuracy also highlights the potential of well-tuned transfer-learning pipelines in medical image analysis, where improvements in prediction may benefit clinical diagnosis.

Future work will incorporate more tumor classes and subtypes and investigate other architectures, including Vision Transformers, CNN–Transformer hybrids, and EfficientNet. Testing on larger and more heterogeneous multi-center datasets will help assess robustness across imaging protocols and populations. These extensions may improve clinical applicability and resilience in real-world workflows, supporting precision medicine and patient care in neuro-oncology.

REFERENCES

[1] P. Sabeghi, P. Zarand, S. Zargham, B. Golestany, A. Shariat, M. Chang, E. Yang, P. Rajagopalan, D. C. Phung, and A. Gholamrezanezhad, "Advances in Neuro-Oncological Imaging: An Update on Diagnostic Approach to Brain Tumors," *Cancers*, vol. 16, no. 3, p. 576, 2024.

[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] S. Srinivasan, D. Francis, S. K. Mathivanan, H. Rajadurai, B. D. Shivahare, and M. A. Shah, "A hybrid deep CNN model for brain tumor image multi-classification," *BMC Medical Imaging*, vol. 24, p. 21, 2024.

[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.

[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.

[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. 102117, 2024.

[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. 108412, 2024.

[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.

[11] J. Cheng, "Brain tumor dataset," Figshare dataset, 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," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 7746–7752, 2019.

[14] J. Cheng *et al.*, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PLOS ONE*, vol. 10, no. 10, p. 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)," *International Journal of Scientific Research in Science, Engineering and Technology*, 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 *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. 103345, 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] 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.

[25] 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.

[26] J. Cheng, "Brain tumor dataset," Figshare dataset, 2018.

[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.

[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. 100003, 2020.

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

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