Bayanika Journal of Artificial Intelligence and Data Science

journal homepage: https://bayanika.org/index.php/bjaids

Lightweight CNN-Based Brain Tumor Classification from MRI

Research Article

Journal: BJAIDS, Vol. 1, No. 1, pp. 48–57, 2026

Zupash Awais*
Faculty of Information Technology & Computer Science, University of Central Punjab, Lahore, Pakistan
E-mail: zupash.awais@ucp.edu.pk
ORCID iD: https://orcid.org/0009-0009-7910-6249
*Corresponding Author

Madiha Awais
Lahore Business School, The University of Lahore, Lahore, Pakistan
E-mail: madiha.awais@lbs.uol.edu.pk
ORCID iD: https://orcid.org/0000-0002-3474-6025

Tehreem Rizwan
Department of Computer Science and IT, The University of Lahore, Pakistan
E-mail: tehreemrizwan280@gmail.com
ORCID iD: https://orcid.org/0009-0000-0016-9509

Muhammad Salman Khan
FAST-NUCES (National University of Computer and Emerging Sciences), Pakistan
E-mail: muhammad.salman@lbs.uol.edu.pk
ORCID iD: https://orcid.org/0009-0009-1144-8998

Received: 18 July 2026; Revised: 29 July 2026; Accepted: 5 August 2026; Published: 6 August 2026

Abstract: Automated brain-tumor classification from Magnetic Resonance Imaging (MRI) can support image-based screening and clinical assessment. This paper presents a compact Convolutional Neural Network (CNN) for classifying two-dimensional brain MRI images into Glioma, Meningioma, Pituitary Tumor, and No Tumor. The model was trained and evaluated on a publicly available, multi-source collection of 7,200 images. The supplied test directory was retained as a held-out image-level test set, while the supplied training directory was divided into stratified training and validation subsets. The proposed model achieved an overall accuracy of 91.9%, macro-averaged precision, recall, and F1-score of approximately 91.9%, and a mean one-vs-rest AUC of 0.95. Under the retained normalized timing records, the proposed CNN was used as the $1.00\times$ throughput reference, compared with $0.60\times$ for DenseNet169 and $0.65\times$ for VGG19. These ratios correspond to 66.7% and 53.8% higher throughput for the proposed CNN, respectively. The trained model was also integrated into an implemented Flutter–Flask–Firestore prototype that supports authenticated image upload, server-side preprocessing and inference, prediction storage, and report retrieval. Functional testing verified these operations, although end-to-end latency, concurrent-user capacity, patient-level generalization, external validation, and clinical utility remain to be established.

Keywords: Magnetic Resonance Imaging (MRI), Brain Tumor Classification, Lightweight CNN, Deep Learning, Medical Image Analysis

1. Introduction

Brain-tumor detection and classification from Magnetic Resonance Imaging (MRI) have been studied using convolutional, transfer-learning, and transformer-based architectures [1, 2, 3, 4]. MRI provides structural information that can support the assessment of tumor location, morphology, and surrounding tissue. Image interpretation nevertheless requires careful review by trained clinicians and may vary with image quality and clinical experience. These constraints motivate computer-aided methods that provide consistent image-level predictions without the computational cost of substantially deeper architectures.

Convolutional Neural Networks (CNNs) learn hierarchical spatial representations directly from image data. This makes them suitable for MRI-based classification tasks in which the relevant features are not defined manually. In the present paper, the purpose of the CNN is limited to four-class image classification. It is not presented as a clinical diagnostic system.

This paper evaluates a lightweight CNN using a publicly available, multi-source dataset of two-dimensional brain MRI images. The evaluation reports overall accuracy, class-wise precision, recall, F1-score, one-vs-rest AUC, and normalized inference throughput. DenseNet169 [5] and VGG19 [6] are evaluated using the same saved image partitions and training procedure. On the reported held-out image-level test set, the proposed CNN achieved 91.9% accuracy and a mean one-vs-rest AUC of 0.95.

The main contributions are as follows:

- We evaluate a compact depthwise-separable CNN for four-class brain-tumor MRI classification. The reported test accuracy is 91.9%, with a mean one-vs-rest AUC of 0.95.
- We retain the supplied 1,600-image test directory and divide the supplied 5,600-image training directory into stratified training and validation subsets using an 80/20 ratio and random seed 42.
- We compare the proposed CNN with DenseNet169 and VGG19 using the same saved partitions, augmentation policy, optimizer, batch size, validation criterion, and checkpoint-selection procedure. The retained timing record reports normalized throughput values of $1.00\times$, $0.60\times$, and $0.65\times$, respectively.
- We implement a deployment prototype comprising a Flutter frontend, Flask middleware, TensorFlow model-serving component, and Firestore data layer. Functional testing verifies authenticated scan upload, preprocessing, four-class inference, prediction storage, and report retrieval.

The remainder of the paper is organized as follows. Section 2 reviews closely related brain-tumor classification studies. Section 3 describes the dataset, model, baseline protocol, computational environment, implemented deployment topology, and evaluation metrics. Section 4 reports the results and limitations. Section 5 concludes the paper.

2. Related Work

CNN-based methods remain common in MRI-based brain-tumor classification because they learn spatial features directly from the input images. Srinivasan et al. [1] proposed a hybrid deep CNN for multiclass brain-tumor classification. Khaliki and Başarslan [2] compared a three-layer CNN with transfer-learning architectures. These studies support the use of compact CNNs as practical baselines, although differences in datasets, preprocessing, class definitions, and validation procedures prevent direct numerical comparison across papers.

Transformer and hybrid models have been used to capture broader spatial relationships in MRI scans. Ahmed et al. [3] combined a Vision Transformer with a Gated Recurrent Unit. Krishnan et al. [4] developed a rotation-invariant Vision Transformer, while Hong et al. [7] used relative positional encoding and a residual multilayer perceptron in a ViT-B/16 model. Al Bataineh et al. [8] combined a Swin Transformer with ResNet50V2. These models provide alternatives to conventional CNNs but may require more computation and memory.

Explainability and multimodal analysis form two additional research directions. Musthafa et al. [9] applied Grad-CAM to a ResNet-50 classifier, and Rasool et al. [10] incorporated an explainability component into CNN-TumorNet. Maqsood et al. [11] examined multimodal learning for brain-tumor detection. These studies are relevant to future extensions, but the present experiments evaluate image classification only and do not apply an explainability method.

Although Alzheimer's disease is outside the experimental scope of this paper, a concise neuroimaging context is retained to identify methodological developments that also arise in brain MRI analysis. CNN-based studies have examined early classification and multistage disease assessment [12, 13], while reproducibility-focused evaluations have shown that results depend strongly on data preparation and validation design [14].

Overall, the literature reports strong predictive results across CNN, transformer, hybrid, and multimodal models. The reported numbers are not directly comparable because the experimental protocols differ. The present paper therefore limits its comparison to three models trained and evaluated on the same saved image partitions.

Table 1. Representative deep-learning studies on MRI-based brain-tumor classification.

Paper	Architecture	Task	Relevance to the our paper
Srinivasan et al. [1]	Hybrid deep CNN	Multiclass brain-tumor classification	Provides a convolutional multiclass comparison for MRI-based tumor classification.
Khaliki and Başarslan [2]	Three-layer CNN and transfer-learning models	Brain-tumor detection	Compares a compact CNN with deeper transfer-learning architectures and motivates an accuracy–efficiency analysis.
Ahmed et al. [3]	Vision Transformer and GRU	Brain-tumor detection and classification	Combines transformer-based image representation with recurrent processing.
Krishnan et al. [4]	Rotation-invariant Vision Transformer	Brain-tumor detection	Studies model sensitivity to image orientation.
Hong et al. [7]	ViT-B/16 with relative positional encoding and residual MLP	Brain-tumor classification	Examines positional representation and residual processing in a transformer classifier.
Al Bataineh et al. [8]	Swin Transformer and ResNet50V2	Brain-tumor classification	Integrates convolutional feature extraction with hierarchical transformer processing.
Musthafa et al. [9]	ResNet-50 with Grad-CAM	Brain-tumor detection	Uses post-hoc saliency visualization to inspect individual predictions.
Rasool et al. [10]	CNN-TumorNet	Brain-tumor diagnosis	Combines CNN-based classification with an explainability component.
Maqsood et al. [11]	Multimodal deep neural network and multiclass SVM	Brain-tumor detection	Examines the use of complementary modalities for tumor classification.

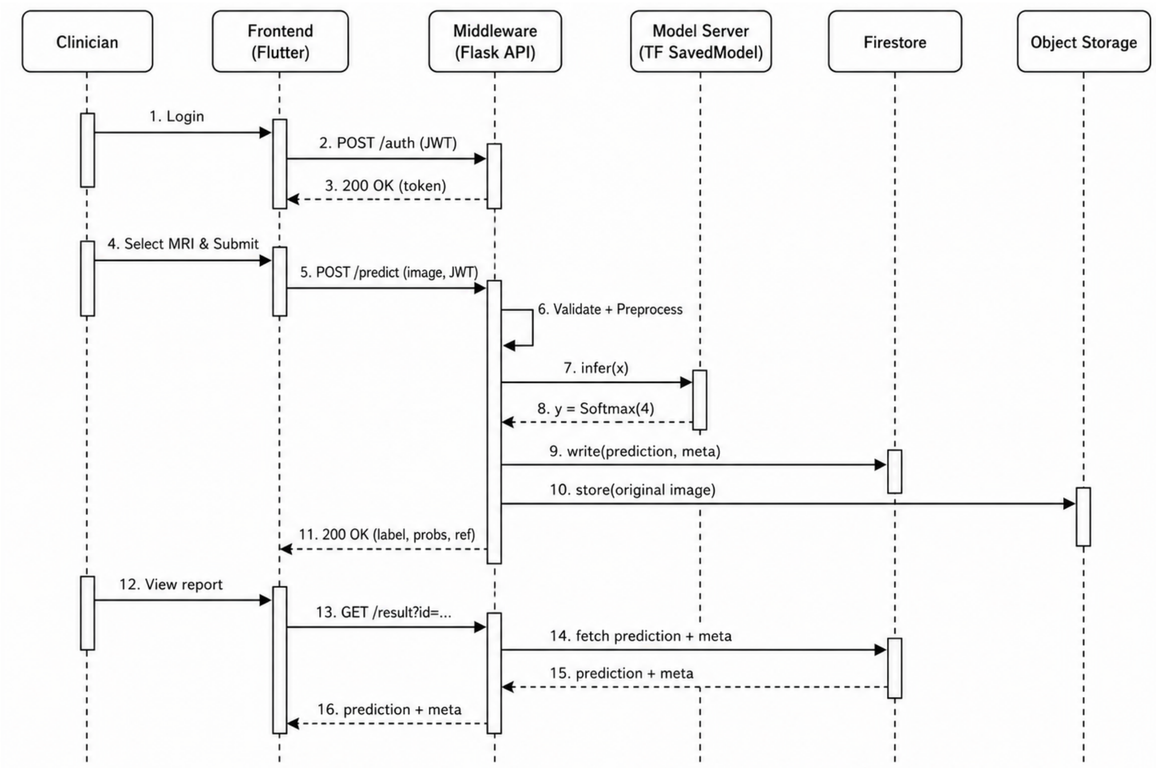


Fig. 1. Sequence diagram of the implemented MRI classification and report-retrieval prototype. The workflow connects the Flutter frontend, Flask middleware, TensorFlow model service, and Firestore data layer.

Table 2. Class distribution across the training, validation, and test subsets.

Class	Training	Validation	Testing	Total
Glioma	1,120	280	400	1,800
Meningioma	1,120	280	400	1,800
Pituitary Tumor	1,120	280	400	1,800
No Tumor	1,120	280	400	1,800
Total	**4,480**	**1,120**	**1,600**	**7,200**

3. Methodology

The paper includes both the four-class CNN evaluation and an implemented deployment prototype. Figure 1 shows the operational sequence from authenticated image upload to preprocessing, inference, prediction storage, and report retrieval. Model performance is evaluated on the held-out image-level test set, while the application layer is assessed through functional end-to-end testing.

3.1. Dataset Description

The experiments use the publicly available Brain Tumor MRI Dataset [15]. The accessed release contains 7,200 two-dimensional brain MRI images distributed equally among Glioma, Meningioma, Pituitary Tumor, and No Tumor. The collection combines images from the Figshare, SARTAJ, and Br35H datasets. It is therefore heterogeneous with respect to source, preprocessing history, image dimensions, and potentially acquisition conditions.

The distributed files are processed PNG or JPG images. Consistent patient identifiers, scanner information, acquisition-site information, and MRI-sequence metadata are not available for all images. The images therefore cannot be verified as uniformly T1-weighted. We refer to them more generally as two-dimensional brain MRI images.

The supplied test directory contains 400 images per class and was retained as a held-out image-level test set. The supplied training directory contains 1,400 images per class. It was divided into stratified training and validation subsets using an 80/20 ratio and random seed 42. This produced 4,480 training images, 1,120 validation images, and 1,600 test images. The partition was created before augmentation and saved for use by all three models.

The split was performed at the image level because patient identifiers were unavailable. The dataset provider reports removing duplicate images and direct overlap between the supplied training and testing directories. This does not establish patient-level independence. Different slices from the same patient may occur in more than one subset, and the reported performance may therefore overestimate generalization to unseen patients.

Each image was converted to a single channel, resized to 224×224, and normalized using a per-image z-score:

$$\tilde{x} = \frac{x - \mu(x)}{\sigma(x) + \epsilon}. \tag{1}$$

Augmentation was applied only to the training subset. It included random rotation ($\pm 15^\circ$), isotropic scaling in the range 0.9–1.1, horizontal and vertical flipping, brightness and contrast adjustment, and mild Gaussian noise. Vertical flipping was used as a generic image-space regularizer in the reported experiment. It was not supported by a clinically validated acquisition model and may produce an anatomically unrealistic superior–inferior orientation. This choice is treated as a limitation.

3.2. CNN Architecture

The proposed classifier uses depthwise-separable convolutions. It contains a stem block with a 3×3 convolution, batch normalization, ReLU activation, and 32 output channels. Three feature stages use depthwise-separable convolution, batch normalization, ReLU activation, and 2×2 max pooling, with 64, 128, and 192 output channels. Global average pooling is followed by dropout ($p = 0.3$) and a four-class Softmax output layer.

Let $x^{(l-1)}$ denote the input feature map of layer l. A generic convolutional feature is

$$f_{i,j}^{(l)} = \sigma\left(\sum_{m=1}^{M}\sum_{n=1}^{N} w_{m,n}^{(l)} x_{i+m,j+n}^{(l-1)} + b^{(l)}\right), \tag{2}$$

where $\sigma(\cdot)$ is ReLU, and $w^{(l)}$ and $b^{(l)}$ are learnable parameters.

The network is optimized using categorical cross-entropy:

$$\mathcal{L} = -\sum_{c=1}^{4} y_c \log \hat{p}_c. \tag{3}$$

The training subset is balanced, so uniform class weights are used.

The Adam optimizer uses an initial learning rate of 10^{-3}, a weight-decay parameter of 10^{-5}, and a batch size of 32. Training is limited to 100 epochs. Early stopping monitors validation loss with a patience of 10 epochs. The learning rate is reduced by a factor of 0.1 when validation loss reaches a plateau. The checkpoint with the lowest validation loss is retained. It is evaluated once on the held-out image-level test set after model selection.

3.3. Baseline Models and Training Protocol

The proposed CNN, DenseNet169 [5], and VGG19 [6] use the same saved training, validation, and test partitions. The same training-only augmentation process, Adam optimizer, initial learning rate, weight-decay parameter, batch size, maximum number of epochs, early-stopping criterion, learning-rate schedule, and validation-based checkpoint-selection procedure are applied to all models.

DenseNet169 and VGG19 are initialized from random weights. ImageNet-pretrained weights and transfer learning are not used. Their original classification heads are removed and replaced with global average pooling, dropout ($p = 0.3$), and a four-unit Softmax layer. The single-channel MRI input is replicated across three channels to form a $224 \times 224 \times 3$ tensor. All layers are trained end-to-end.

3.4. Computational Environment

The experiments were executed in a hosted Google Colab GPU runtime. Google Colab assigns hardware and preinstalled software dynamically. The retained timing record used the proposed CNN as the normalized throughput reference at $1.00\times$. DenseNet169 and VGG19 achieved $0.60\times$ and $0.65\times$ the proposed model's throughput. These ratios imply that the proposed CNN had 66.7% higher throughput than DenseNet169 and 53.8% higher throughput than VGG19. Equivalently, its relative per-image latency was 40% lower than DenseNet169 and 35% lower than VGG19.

3.5. Deployment Topology

The deployment prototype in Figure 1 was implemented using a Flutter frontend, a Flask-based REST middleware service, a TensorFlow SavedModel inference component, and Firestore for prediction metadata and report retrieval. The frontend supports authenticated image submission and presents the returned four-class prediction to the user. The Flask service validates the request, applies the same resizing, grayscale conversion, and normalization used during model evaluation, invokes the trained classifier, and returns the predicted class together with the four Softmax probabilities.

JWT-based authentication is used to associate requests with authenticated users, and HTTPS is used for communication between the client and the hosted service. Request validation rejects unsupported file types and malformed requests before inference. The model service records the selected model version, prediction timestamp, predicted class, and class-probability vector. The corresponding prediction record is stored in Firestore so that the user can retrieve the report in a later session. Request logging and basic rate limiting are applied at the middleware layer.

Figure 2 presents the two main user-facing interfaces of the implemented prototype. The upload interface allows an authenticated user to select a brain MRI image, preview the selected scan, verify basic file information, and submit the image for classification. The interface also displays the supported file formats, expected input resolution, deployed model version, and the four output classes. When the user initiates a prediction, the Flutter client sends the image and authentication token to the Flask REST endpoint through an HTTPS request.

3.6. Evaluation Metrics

The evaluation uses accuracy, class-wise precision, recall, F1-score, and one-vs-rest Area Under the Receiver Operating Characteristic Curve (AUC). For each class, true positives, false positives, true negatives, and false negatives are computed using a one-vs-rest formulation.

Recall and precision are defined as

$$\mathrm{Recall}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \qquad \mathrm{Precision}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c}. \tag{4}$$

The class-wise F1-score is

$$\mathrm{F1}_c = 2\frac{\mathrm{Precision}_c \mathrm{Recall}_c}{\mathrm{Precision}_c + \mathrm{Recall}_c}. \tag{5}$$

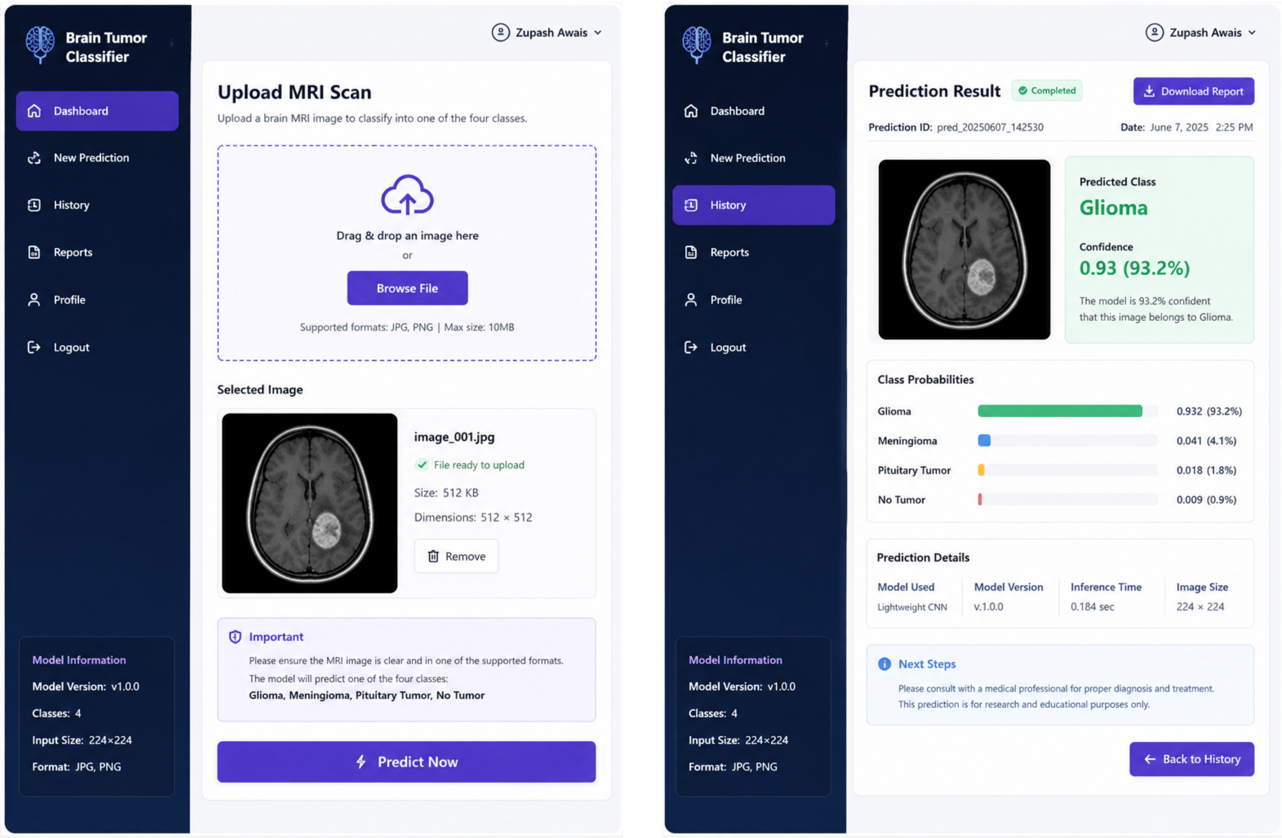
For the multiclass confusion matrix M, accuracy is

$$\mathrm{Accuracy} = \frac{\sum_{c=1}^{C} M_{cc}}{N}, \tag{6}$$

where $C = 4$ and N is the number of test images.

Macro averaging uses the unweighted mean:

$$\mathrm{Metric}_{\mathrm{macro}} = \frac{1}{C}\sum_{c=1}^{C}\mathrm{Metric}_c. \tag{7}$$



(a) MRI upload and submission interface. The user can preview the selected scan, review its file information, and initiate four-class prediction.

(b) Prediction-result interface showing the submitted MRI, predicted class, confidence score, complete four-class probability distribution, and report-management options.

Fig. 2. User interfaces of the implemented brain MRI classification prototype.

Table 3. Class-wise performance of the proposed CNN on the held-out image-level test set.

Class	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC**	**Support**
Glioma	92.0	91.5	91.7	0.95	400
Meningioma	91.8	92.5	92.2	0.96	400
Pituitary Tumor	93.2	92.5	92.8	0.96	400
No Tumor	90.5	91.0	90.8	0.94	400
Macro average	**91.9**	**91.9**	**91.9**	**0.95**	**1,600**

Because the test set has equal support for all four classes, macro-averaged recall is equal to overall accuracy, apart from display rounding.

For each one-vs-rest class, the true-positive and false-positive rates are

$$\mathrm{TPR}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \qquad \mathrm{FPR}_c = \frac{\mathrm{FP}_c}{\mathrm{FP}_c + \mathrm{TN}_c}. \tag{8}$$

The corresponding AUC is

$$\mathrm{AUC}_c = \int_0^1 \mathrm{TPR}_c(\mathrm{FPR}_c)\, d(\mathrm{FPR}_c). \tag{9}$$

4. Results

On the held-out image-level test set of 1,600 images, the proposed CNN achieved an overall accuracy of 91.9%. Macro-averaged precision, recall, and F1-score were each approximately 91.9%, and the mean one-vs-rest AUC was 0.95. Because each class contributed 400 test images, macro-averaged recall and overall accuracy are equal within the reported rounding precision.

Table 3 reports similar performance across the four classes. Pituitary Tumor has the highest reported precision at 93.2%, while No Tumor has the lowest at 90.5%. Class-wise recall ranges from 91.0% to 92.5%.

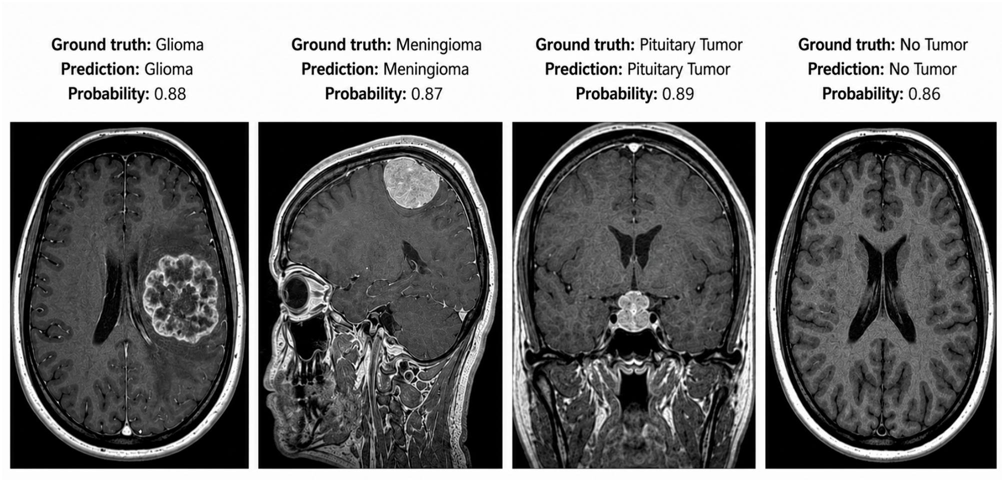


Fig. 3. Representative correctly classified MRI images from the held-out image-level test set.

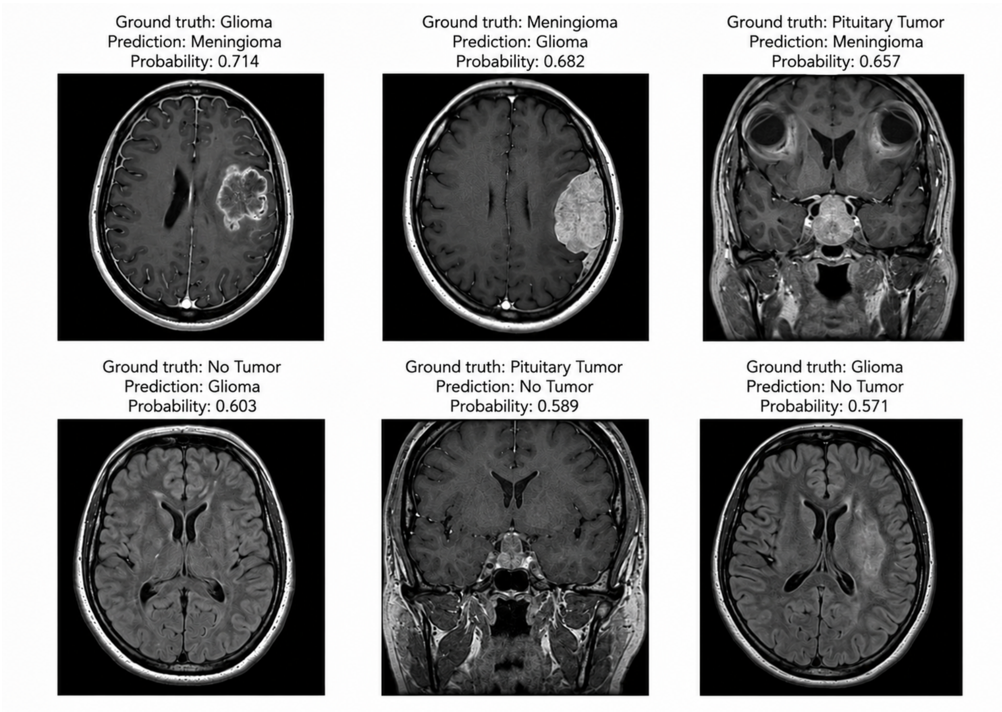


Fig. 4. Representative misclassified MRI images from the held-out image-level test set.

Table 4. Confusion matrix of the proposed CNN. Rows represent ground-truth classes and columns represent predicted classes.

Actual / Predicted	Glioma	Meningioma	Pituitary	No Tumor
Glioma	**366**	10	10	14
Meningioma	9	**370**	7	14
Pituitary Tumor	10	10	**370**	10
No Tumor	13	13	10	**364**

Table 4 reports 366 correct Glioma predictions, 370 correct Meningioma predictions, 370 correct Pituitary Tumor predictions, and 364 correct No Tumor predictions. The diagonal entries sum to 1,470 correct classifications.

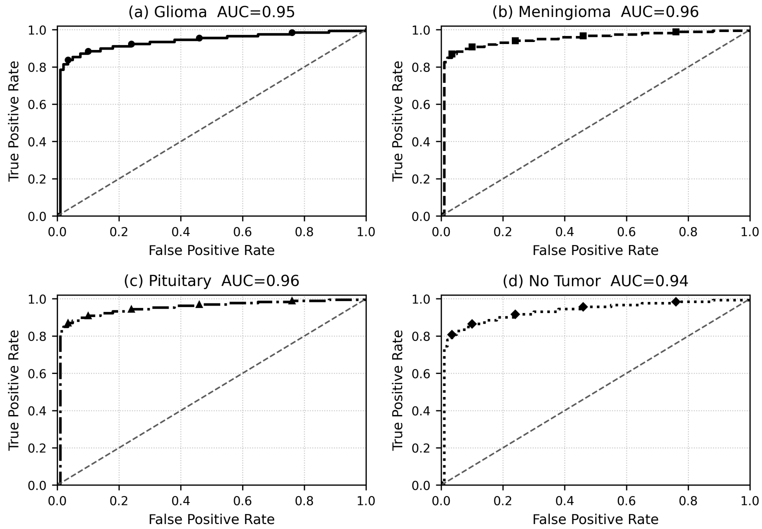


Fig. 5. One-vs-rest ROC curves on the held-out image-level test set.

The one-vs-rest ROC curves in Fig. 5 correspond to AUC values of 0.95 for Glioma, 0.96 for Meningioma, 0.96 for Pituitary Tumor, and 0.94 for No Tumor.

Table 5. Classification performance and normalized throughput on the same held-out image-level test set.

Model	**Accuracy (%)**	**Macro P (%)**	**Macro R (%)**	**Macro F1 (%)**	**AUC**	**Norm. throughput**
Proposed CNN	91.9	91.9	91.9	91.9	0.95	1.00×
DenseNet169 [5]	90.7	91.0	90.7	90.8	0.94	0.60×
VGG19 [6]	89.9	90.4	89.9	90.0	0.93	0.65×

Table 5 shows that the proposed CNN has the highest reported accuracy among the three evaluated models. The normalized throughput ratios correspond to 66.7% higher throughput than DenseNet169 and 53.8% higher throughput than VGG19. These comparisons apply only to the common Google Colab session in which the normalized timing values were recorded.

Figure 3 presents one correctly classified example from each of the four output classes: Glioma, Meningioma, Pituitary Tumor, and No Tumor. For each example, the ground-truth class, predicted class, and predicted-class probability are reported to provide a concise indication of the model's confidence. Figure 4 presents representative misclassified examples from the held-out test set. These cases show that errors may occur when different tumor categories exhibit similar visual characteristics, when lesion boundaries are less distinct, or when the scan contains atypical anatomical patterns. These figures complement the quantitative interpretation provided by Table 3, Table 4, and Fig. 5.

5. Conclusion and Future Work

This paper evaluated a compact CNN for classifying two-dimensional brain MRI images into Glioma, Meningioma, Pituitary Tumor, and No Tumor. The proposed model achieved 91.9% accuracy, a macro-averaged F1-score of approximately 91.9%, and a mean one-vs-rest AUC of 0.95 on the held-out image-level test set. Its retained normalized throughput was higher than that of DenseNet169 and VGG19 under the same Google Colab session. The trained model was integrated into an implemented Flutter-Flask-Firestore prototype, and functional testing verified authenticated upload, preprocessing, inference, prediction storage, and report retrieval.

The results do not establish patient-level or institution-level generalization. Future work should use datasets with reliable patient identifiers and acquisition metadata, remove anatomically implausible augmentation, and conduct patient-disjoint and external validation. Future evaluations should also retain sample-level probabilities, exact model summaries, and complete timing logs. Grad-CAM or another interpretation method may then be added and assessed by domain experts. The implemented prototype should be evaluated under concurrent load and on target devices, with explicit measurements of end-to-end latency, availability, security, and failure recovery before claims of real-time or clinical readiness are made.

Declarations

Author Contributions Statement

Zupash Awais: conceptualization, methodology, and validation. **Madiha Awais:** supervision and writing-original draft. **Tehreem Rizwan:** writing-review and editing. **Muhammad Salman Khan:** methodology and editing. All authors reviewed and approved the final manuscript.

Acknowledgment

The authors thank the Center for Emerging Networks & Technologies (CENT) Lab at the University of Central Punjab (UCP), Lahore, Pakistan, for technical support and constructive feedback.

Conflict of Interest Statement

The authors declare no conflict of interest.

Funding Declaration

This research received no external funding.

Data Availability Statement

The Brain Tumor MRI Dataset used in this paper is publicly available on Kaggle at `https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset`.

Declaration of Generative AI in Scholarly Writing

During the preparation of this work, the authors used generative AI tools only for language editing and proofreading, where applicable. The authors reviewed and edited the content and take full responsibility for the article.

References

[1] Saravanan Srinivasan, Divya Francis, Sandeep Kumar Mathivanan, Hariharan Rajadurai, Basu Dev Shivahare, and Mohd Asif Shah. A hybrid deep CNN model for brain tumor image multi-classification. *BMC Medical Imaging*, 24(1):21, 2024.

[2] Mohammad Zafer Khaliki and Muhammet Sinan Başarslan. Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. *Scientific Reports*, 14(1):2664, 2024.

[3] Md. Mahfuz Ahmed, Md. Maruf Hossain, Md. Rakibul Islam, Md. Shahin Ali, Abdullah Al Noman Nafi, Md. Faisal Ahmed, Kazi Mowdud Ahmed, Md. Sipon Miah, Md. Mahbubur Rahman, Mingbo Niu, and Md. Khairul Islam. Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in southern bangladesh. *Scientific Reports*, 14(1):22797, 2024.

[4] Palani Thanaraj Krishnan, Pradeep Krishnadoss, Mukund Khandelwal, Devansh Gupta, Anupoju Nihaal, and T. Sunil Kumar. Enhancing brain tumor detection in MRI with a rotation invariant vision transformer. *Frontiers in Neuroinformatics*, 18:1414925, 2024.

[5] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[7] Shuang Hong, Jin Wu, Lei Zhu, and Weijie Chen. Brain tumor classification in VIT-B/16 based on relative position encoding and residual MLP. *PLOS ONE*, 19(7):e0298102, 2024.

[8] Abeer Fayez Al Bataineh, Khalid M. O. Nahar, Hayel Khafajeh, Ghassan Samara, Raed Alazaidah, Ahmad Nasayreh, Ayah Bashkami, Hasan Gharaibeh, and Waed Dawaghreh. Enhanced magnetic resonance imaging-based brain tumor classification with a hybrid swin transformer and ResNet50V2 model. *Applied Sciences*, 14(22):10154, 2024.

[9] Mohamed Musthafa M, Mahesh T. R, Vinoth Kumar V, and Suresh Guluwadi. Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with ResNet-50. *BMC Medical Imaging*, 24(1):107, 2024.

[10] Novsheena Rasool, Niyaz Ahmad Wani, Javaid Iqbal Bhat, Sandeep Saharan, Vishal Kumar Sharma, Bassma Saleh Alsulami, Hind Alsharif, and Miltiadis D. Lytras. CNN-TumorNet: Leveraging explainability in deep learning for precise brain tumor diagnosis on MRI images. *Frontiers in Oncology*, 15:1554559, 2025.

[11] Sarmad Maqsood, Robertas Damaševičius, and Rytis Maskeliūnas. Multi-modal brain tumor detection using deep neural network and multiclass SVM. *Medicina*, 58(8):1090, 2022.

[12] A. M. El-Assy, Hanan M. Amer, H. M. Ibrahim, and M. A. Mohamed. A novel CNN architecture for accurate early detection and classification of alzheimer's disease using MRI data. *Scientific Reports*, 14(1):3463, 2024.

[13] Muhammad Umair Ali, Kwang Su Kim, Majdi Khalid, Majed Farrash, Amad Zafar, and Seung Won Lee. Enhancing alzheimer's disease diagnosis and staging: A multistage CNN framework using MRI. *Frontiers in Psychiatry*, 15:1395563, 2024.

[14] Junhao Wen, Elina Thibeau-Sutre, Mauricio Diaz-Melo, Jorge Samper-González, Alexandre Routier, Simona Bottani, Didier Dormont, Stanley Durrleman, Ninon Burgos, Olivier Colliot, Alzheimer's Disease Neuroimaging Initiative, and Australian Imaging Biomarkers and Lifestyle flagship study of ageing. Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis*, 63:101694, 2020.

[15] Masoud Nickparvar. Brain tumor MRI dataset. Kaggle, 2021. Updated version containing 7,200 images; accessed 15 July 2026.