Brain Tumor Classifications Using MRI Scans with Deep Learning

Anastasija Stefanovska and Hristijan Gjoreski
Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University in Skopje, Rugjer Boshkovik 18, 1000 Skopje, N.Macedonia
anastasijast.94@gmail.com, hristijang@feit.ukim.edu.mk

Keywords: Brain Tumor, Magnetic Resonance Imaging, Deep Learning, Convolutional Neural Networks, Medical Image Analysis.

Abstract: This paper focuses on the detection and classification of brain tumors using MRI images and Deep Learning methods. The dataset consists of 7,023 MRI scans representing healthy brains and brains with glioma, meningioma, or pituitary tumors. We developed and evaluated Convolutional Neural Network (CNN) models of varying depth, combined with preprocessing techniques such as image resizing, limited augmentation, and validation set allocation to prevent overfitting. The study demonstrates that CNN-based approaches can achieve high accuracy in tumor classification, with the best model reaching an overall accuracy of 98.5%. Results show that the greatest misclassification occurred between glioma and meningioma, reflecting their similar MRI appearance. Further analysis confirms that these tumor types share overlapping visual characteristics, making them more challenging to separate. Overall, the findings highlight that well-designed architectures and carefully controlled preprocessing can significantly enhance automated brain tumor detection, offering valuable support for radiologists and contributing to more efficient, consistent, and reliable diagnostic workflows.

1 INTRODUCTION

Brain tumors are pathological cell growths within or around brain tissue, which can be benign (non-cancerous) or malignant (cancerous). Although benign tumors typically grow more slowly and rarely metastasize, their location can still cause severe neurological impairments due to pressure on brain structures. Brain cancer ranks among the top ten leading causes of mortality worldwide, with specific risk factors including age, radiation exposure, and ethnic or racial background [1]. The World Health Organization (WHO) classifies brain tumors into four grades, based on histological characteristics and their potential to infiltrate surrounding tissue [2]. The most common tumor types in adults are gliomas, meningiomas, and pituitary tumors, while in children, medulloblastomas and other embryonal tumors predominate [3]. The symptoms of brain tumors vary depending on their size and location and may include headaches, seizures, visual and
speech disturbances, paralysis, and hormonal imbalance. Diagnosis typically involves a neurological examination, MRI imaging, and biopsy, with liquid biopsies increasingly being applied to detect tumor markers in recent years [4]. In recent years, machine learning has emerged as an effective approach for tumor detection, enabling the development of models capable not only of identifying tumors but also of classifying them by type. This research focuses on developing a CNN-based model designed for accurate detection and classification of brain tumors.

2 RELATED WORK

In recent years, deep learning (DL) has become a dominant approach for brain tumor classification using MRI images, complementing and often surpassing traditional machine learning (ML) methods. Early ML-based strategies relied on handcrafted features such as intensity, texture, and shape, combined with classifiers like support vector machines (SVM), k-nearest neighbors, or random forests [5]. While these approaches achieved reasonable accuracy, their dependence on manually designed features limited scalability and

generalization. To overcome these limitations, hybrid strategies emerged, where features extracted by convolutional neural networks (CNNs) were fed into ML classifiers, achieving classification accuracies exceeding 90% [6].

Fully CNN-based architectures were also developed for classification and segmentation tasks, reporting accuracies between 90% and 97% under controlled training conditions [7]. Recent work has shifted toward more sophisticated DL frameworks for both classification and segmentation. Multi-scale 3D CNN models, such as DeepMedic, incorporate local and contextual information from MRI volumes, improving segmentation performance [8]. Similarly, multimodal approaches that integrate T1, T2, and FLAIR sequences have enhanced the detection and classification of tumor subregions [9]. Transfer learning using pre-trained networks such as VGG, ResNet, and GoogleNet has proven especially valuable for smaller datasets, enabling robust feature extraction without extensive training [10].

More recently, advanced architectures like multimodal 3D DenseNet leverage multi-sequence MRI data to grade gliomas with high precision [11]. Despite these advances, challenges remain, including the limited availability of annotated datasets, difficulty handling rare tumor types, and generalization across different imaging protocols. These issues underline the ongoing need for hybrid or novel DL architectures that can fully exploit multimodal MRI data while improving classification accuracy and robustness in clinical practice.

3 MACHINE LEARNING FOR BRAIN TUMOR DETECTION

Machine learning, and particularly deep learning, can be applied to the detection and classification of brain tumors [12]. The entire process of developing a model begins with the selection of a dataset that will be used for training. In this study, we employ a dataset consisting of 7,023 MRI images, which include both healthy brains (without tumors) and brains with tumors. The tumors in the dataset belong to three distinct categories. For image analysis, we utilize models based on Convolutional Neural Network (CNN) layers, which are particularly effective for image processing due to their ability to recognize complex spatial patterns and learn hierarchical features. Generally, the process of model development involves several steps: data collection, data preprocessing, model architecture design and training, validation, and testing. In this specific case, the following steps were undertaken:

1) Dataset selection – the aforementioned MRI dataset was selected for model training [13].
2) Data preprocessing – multiple techniques were applied to transform the images into the desired format.
3) Model architecture – three CNN-based models of different depths were constructed for training. The aim was to investigate whether deeper architectures yield improved performance without causing overfitting.
4) Model training – key hyperparameters were adjusted, and the training phase was performed, during which the model was provided with labeled data to learn correct classification.
5) Validation and testing – validation was used to monitor how the model performs on previously unseen data, while testing represented the final phase, providing accuracy results on a separate set of entirely new data (unseen by the model during training).

4 DATASET AND PREPROCESSING

4.1 Dataset

The dataset consists of 7,023 MRI images divided into four categories:

1) No tumor: 2,000 healthy brain images.
2) Glioma: 1,621 images of malignant tumors originating from glial cells.
3) Meningioma: 1,645 images of tumors arising in the meninges.
4) Pituitary tumor: 1,757 images of tumors located in the pituitary gland.

Figure 1 presents representative examples of MRI scans from each class, while Figure 2 illustrates the distribution of images across the four categories.

4.2 Data Preprocessing

Preprocessing is a crucial step in medical image analysis, as it ensures that the input data is standardized and suitable for effective model training. In this study, two preprocessing techniques were applied during data loading: resizing and augmentation. Resizing was performed to standardize all images to the same dimensions, allowing consistent transformation and use in both training and testing.

Augmentation involved applying slight transformations and rotations to the training images. Specifically, the augmentation process included random rotations up to 10°, affine translations of up to 5% along both axes, and slight brightness and contrast adjustments of ±0.05, followed by normalization to a mean of 0.5 and a standard deviation of 0.5. Because the dataset consists of medical scans, the augmentation was intentionally kept minimal to preserve anatomical accuracy. Furthermore, 20% of the training set was allocated for validation, enabling evaluation of the model's performance on unseen data and helping to prevent overfitting.

5 MODEL ARCHITECTURE

In this study, three Convolutional Neural Network (CNN) models were developed, all sharing the same core architecture but differing in depth. Each model is composed of the following layers:

- Conv2d – a two-dimensional convolutional layer commonly used for image processing. By passing the input image through convolutional filters, the layer extracts key spatial patterns that are critical for classification. Deeper networks with multiple convolutional layers can capture increasingly complex features.
- BatchNorm2d – a normalization layer applied to stabilize and accelerate the training process.
- ReLU – a nonlinear activation function that introduces nonlinearity and enhances the representational capacity of CNNs.
- MaxPool2d – a pooling layer that performs downsampling by reducing the spatial dimensions of the feature maps. This decreases the number of parameters and computations, leading to faster training while preserving essential features.
- Dropout – a regularization technique that randomly deactivates a fraction of neurons during training to reduce overfitting.

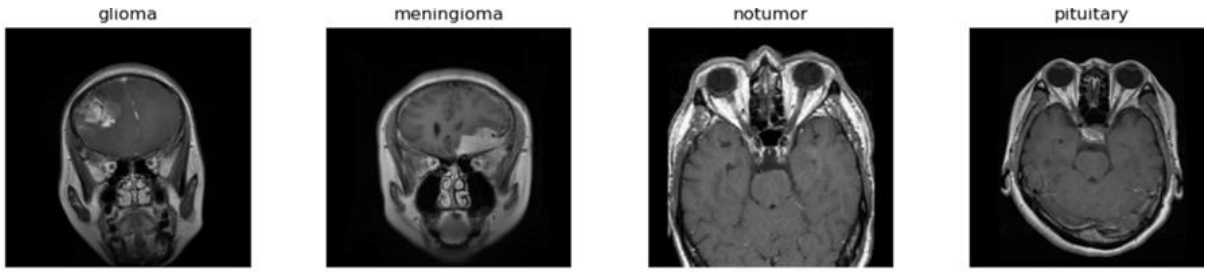

Figure 1: Example images from each class.

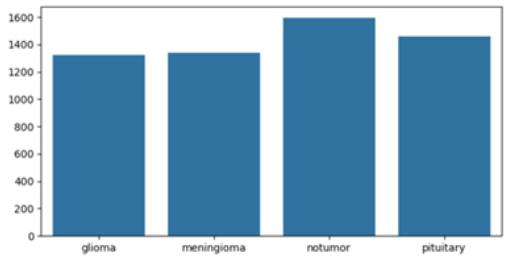

Figure 2: Class distribution.

Table 1: Comparison of model architectures and validation performance.

Model	Trainable parameters	Model size (MB)	Best validation accuracy	Best F1	Epoch of best F1
Base	8,483,588	32.36	0.9011	0.8961	26
Deeper	8,780,548	33.50	0.9466	0.9445	23
Light	16,832,388	64.21	0.8345	0.8287	8

The baseline model consists of three such convolutional blocks, the shallow model of two, and the deep model of four. After defining the architectures, the models were trained using a custom function, which takes the model class as an input. Training was performed for 30 epochs with a batch size of 32 using the Adam optimizer with a learning rate of 0.001 and CrossEntropyLoss as the classification objective, as this configuration provided a stable convergence rate and effectively balanced computational efficiency and generalization performance.

At the end of each epoch, the models were evaluated on the validation set to track accuracy and macro F1 score. To prevent overfitting, an Early Stopping mechanism with a patience of 5 epochs was implemented on the validation set.

This mechanism monitors the best macro F1 score and stops training if no improvement is observed over five consecutive iterations, indicating that continued training could lead to overfitting and a decline in validation performance.

With the established training procedure, the three models can be compared in terms of depth, representational capacity, and generalization performance (Table 1).

The Light model, despite having only two convolutional blocks, incorporates a relatively large fully connected layer, resulting in the highest parameter count (16.83M), which increases its representational capacity but also its susceptibility to overfitting. Consequently, while it trains faster, it achieves lower validation performance (best validation accuracy 0.8345, best F1 0.8287) compared to the other models. The Base model provides a balanced trade-off, with moderate depth and 8.48M parameters, capturing richer hierarchical features than Light while maintaining computational efficiency; it achieves a higher best validation accuracy of 0.9011 and F1 of 0.8961.

The Deeper model further increases representational power through an additional convolutional block and a larger dense layer, totaling 8.78M parameters, enabling it to learn more complex spatial patterns. This architecture attains the highest validation performance among the three (best validation accuracy 0.9466, best F1 0.9445), although with increased computational demand. Overall, this comparison demonstrates that optimal performance is influenced not only by network depth but also by the interaction between parameter count, model complexity, and generalization, emphasizing the importance of balancing capacity and overfitting risk.

6 RESULTS

After training, the models were evaluated using a dedicated function, which accepts the trained model as input. For each model, the confusion matrix, classification report, accuracy, and macro F1 score were computed and stored to enable comparison between models. Model comparison was performed based on the F1 score, and the deepest model achieved the best overall performance according to these metrics. The learning curves of training and validation accuracy for the three models are shown in Figure 3.

Figure 3: Training and validation accuracy curves for Light, Base, and Deep models

The performance of the best-performing (deepest) model was further analyzed using a confusion matrix. The confusion matrix, shown in Figure 4, provides a detailed visualization of correct and incorrect class predictions, highlighting specific patterns of misclassification. The largest misclassification occurred between glioma and meningioma, with 43 glioma images predicted as meningioma. Additionally, 23 meningioma images were misclassified as healthy brains (no tumor). Correctly classified sample images for each tumor type are illustrated in Figure 5, highlighting representative predictions of the deepest model.

Finally, the class-wise performance metrics of the deepest model were visualized, and the per-class recall was reported as follows:

- Glioma: 0.817
- Meningioma: 0.856
- No tumor: 0.985
- Pituitary tumor: 0.993

These values are illustrated in Figure 6, providing a graphical representation of the model's per-class recall. Overall, the figure highlights that the model achieved high accuracy in distinguishing between healthy and tumor-bearing brains, with the most challenging distinction occurring between glioma and meningioma.

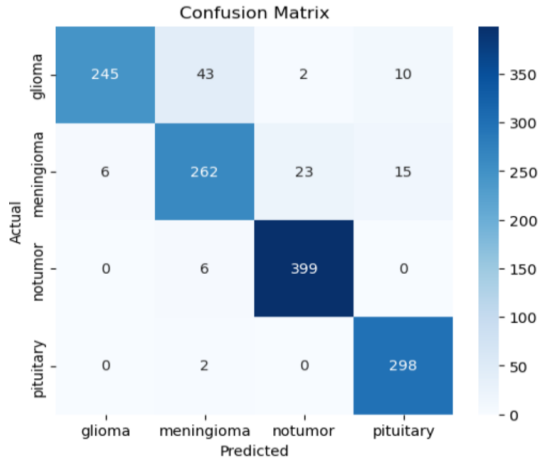

Figure 4: Confusion matrix

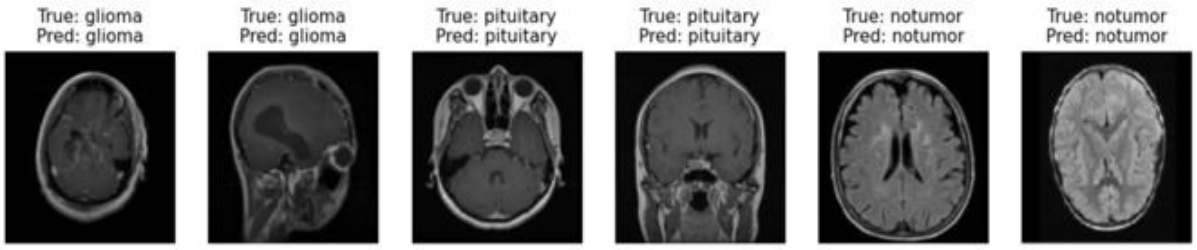

Figure 5: Correctly predicted classes.

Figure 6: Graphical representation of model per-class recall.

7 CONCLUSIONS

The results of this study demonstrate that the developed model for brain tumor detection and classification effectively distinguishes between the different classes. In particular, it achieves high accuracy in determining the presence or absence of a tumor, with an overall accuracy of 98.5%. The greatest misclassification occurred between glioma and meningioma images. This is likely because meningiomas develop exclusively in the meninges, whereas gliomas can arise anywhere in the brain, including regions adjacent to the meninges. In such cases, the model may misclassify the tumor type. Previous studies have confirmed the similarity between these two MRI image types, noting that they can be differentiated based on tumor margins: meningiomas exhibit well-defined boundaries, whereas gliomas often have irregular, blurred edges due to their malignant and infiltrative nature [14]. These findings confirm that Convolutional Neural Networks (CNNs) are a powerful tool for brain tumor classification from MRI scans. With appropriately designed architectures, carefully prepared datasets, and techniques to prevent overfitting, CNN-based models can achieve high accuracy and stability. Although distinguishing certain tumor types remains challenging, integrating additional medical parameters and employing advanced architectures may further enhance the precision of automated diagnostic support systems. For successful translation into clinical practice, the model must be validated on diverse, multi-institutional datasets to ensure robustness across different MRI protocols, scanner types, and patient populations. Integration with existing hospital workflows, including Picture Archiving and Communication Systems (PACS) and electronic health records, is crucial to provide effective decision support for radiologists. Additionally, regulatory approval, compliance with data privacy standards, and ongoing monitoring of model performance are essential to ensure safe and reliable deployment. Addressing these considerations is necessary to translate high experimental accuracy into practical, clinically useful tools, ultimately supporting more accurate, efficient, and informed brain tumor diagnosis and treatment.

REFERENCES

[1] D. Arulmani and R. Manickam, “Brain Tumors, ” J Stud Res, vol. 13, no. 2, May 2024, doi: 10.47611/jsrhs.v13i2.6694.

[2] D. N. Louis et al., “The 2021 WHO Classification of Tumors of the Central Nervous System: a summary,” Neuro-Oncology, vol. 23, no. 8, pp. 1231–1251, June 2021, doi: 10.1093/neuonc/noab106.

[3] P. A. Patil and P. Giridhar, “Epidemiology and Demography of Brain Tumors,” Evidence based practice in Neuro-oncology. Springer Singapore, pp. 3–7, 2021. doi: 10.1007/978-981-16-2659-3_1.

[4] M. Pichaivel, G. Anbumani, P. Theivendren, and M. Gopal, “An Overview of Brain Tumor,” Brain Tumors. IntechOpen, Apr. 20, 2022. doi: 10.5772/intechopen.100806.

[5] M. K. Abd-Ellah, A. I. Awad, A. A. M. Khalaf, and H. F. A. Hamed, “A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned,” Magnetic Resonance Imaging, vol. 61, pp. 300–318, Sept. 2019, doi: 10.1016/j.mri.2019.05.028.

[6] M. A. Khan et al., “Multimodal Brain Tumor Classification Using Deep Learning and Robust Feature Selection: A Machine Learning Application for Radiologists,” Diagnostics, vol. 10, no. 8, p. 565, Aug. 2020, doi: 10.3390/diagnostics10080565.

[7] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, “An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification,” IEEE J. Biomed. Health Inform., vol. 21, no. 1, pp. 31–40, Jan. 2017, doi: 10.1109/jbhi.2016.2635663.

[8] K. Kamnitsas et al., “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” Medical Image Analysis, vol. 36, pp. 61–78, Feb. 2017, doi: 10.1016/j.media.2016.10.004.

[9] M. Havaei et al., “Brain tumor segmentation with Deep Neural Networks,” Medical Image Analysis, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.

[10] R. Mehrotra, M. A. Ansari, R. Agrawal, and R. S. Anand, “A Transfer Learning approach for AI-based classification of brain tumors,” Machine Learning with Applications, vol. 2, p. 100003, Dec. 2020, doi: 10.1016/j.mlwa.2020.100003.

[11] S. Liang et al., “Multimodal 3D DenseNet for IDH Genotype Prediction in Gliomas,” Genes, vol. 9, no. 8, p. 382, July 2018, doi: 10.3390/genes9080382.

[12] A. Vidyarthi, R. Agarwal, D. Gupta, R. Sharma, D. Draheim, and P. Tiwari, “Machine Learning Assisted Methodology for Multiclass Classification of Malignant Brain Tumors,” IEEE Access, vol. 10, pp. 50624–50640, 2022, doi: 10.1109/access.2022.3172303.

[13] Msoud Nickparvar, “Brain Tumor MRI Dataset.” Kaggle, 2021. doi: 10.34740/KAGGLE/DSV/2645886.

[14] B. Hu, Z. Zhang, S. Chen, Q. Xu, and J. Li, “A metric for quantitative evaluation of glioma margin changes in magnetic resonance imaging,” Acta Radiol, vol. 65, no. 6, pp. 645–653, Mar. 2024, doi: 10.1177/02841851241229597.