Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2024) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Computer Science and Computational Intelligence 2024 (ICCSCI 2024)

Identification of Biomarkers for Brain Cancer: Integrating Bioinformatics and Deep Learning

Jonathan Alvindo Fernandi[a], Vinson Luckianto[a], Ghinaa Zain Nabiilah[a,*], Jurike V. Moniaga[a]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia*

Abstract

Brain tumors present a substantial global health challenge, and accurately diagnosing them early is critical for effective treatment planning and better patient outcomes. This study uses a convolutional neural network (CNN) framework called Xception that can precisely identify and categorize brain tumor types from MRI images. This study also incorporates data augmentation techniques to maintain the model's generalization. The model achieves an impressive accuracy of 97.87% in distinguishing between gliomas, meningiomas, pituitary tumors, and non-tumor cases, with high precision and recall scores for all classes. The model's predictions are visually presented, and detailed descriptions of the tumors are provided to aid in clinical decision-making. While the study acknowledges the limitations of using a single imaging modality and suggests integrating multi-modal data and multi-omics information for further improvement, this research represents a significant step forward in non-invasive and accurate brain tumor diagnosis. By combining bioinformatics and deep learning, this approach has the potential to contribute to personalized treatment planning and better patient outcomes in neuro-oncology.

Peer-review under responsibility of the scientific committee of the 9th International Conference on Computer Science and Computational Intelligence 2024
Keywords: bioinformatics; brain cancer biomarkers; magnetic resonance imaging (MRI); deep learning; convolutional neural network (CNN)

* Ghinaa Zain Nabiilah.
E-mail address: ghinaa.nabiilah@binus.edu

Peer-review under responsibility of the scientific committee of the 9th International Conference on Computer Science and Computational Intelligence 2024

1. Introduction

Brain cancer is a significant global health burden. To plan optimal treatment and improve patient prognosis, it is crucial to diagnose them precisely and early. Brain cancer assessment often uses magnetic resonance imaging (MRI) to identify brain tumor types. However, manual MRI scan analysis is time-consuming, prone to subjectivity, and can vary between radiologists. This has led to an increasing interest in developing automated AI models to identify and categorize brain tumor types from MRI images.

Studies have proved that convolutional neural networks (CNNs) are the most successful approach in achieving high accuracy in tasks such as tumor segmentation, classification, and grading. Pei et al. (2020) introduced a context-aware 3D CNN named CANet that leverages global context features for robust tumor subtypes classification, such as glioma, glioblastoma, and oligodendroglioma, which was ranked 2nd on the BraTS 2018 dataset classification [1]. Hammad et al. (2023) developed an efficient 8-layer CNN that attained 96.86% accuracy in categorizing glioma, meningioma, and pituitary tumors using 3064 MRI images, and it has the potential for real-time clinical deployment [2].

Researchers have also focused on optimizing CNN architectures. Nassar et al. (2023) proposed BCM-CNN, a hybrid approach combining a fine-tuned Inception-ResNetV2 with a custom 12-layer CNN branch, achieving 99.98% classification accuracy on BraTS 2021 [3]. Rethemiotaki (2023) compared CNN variants and found that MobileNetV2 had the highest accuracy (99%) and GoogleNet had the best F1 score (97%) for tumor type classification [4]. Shin et al. (2021) developed a deep learning model to categorize glioblastoma and brain metastasis using conventional MRI images, achieving an AUROC of 0.94 [5]. Tariciotti et al. (2022) developed a 3D CNN model to classify glioblastoma, lymphoma, and brain metastasis, achieving 92.5% accuracy on 120 patients [6]. Saeedi et al. (2023) compared ML models to identify brain tumors using MRI images, with ResNet achieving 98.6% accuracy, outperforming traditional approaches [7].

Efforts have been made to integrate segmentation with downstream tasks in unified frameworks. Sun et al. (2019) used an ensemble of 3D CNNs for segmentation, extracted radiomic features, and trained a random forest for survival prediction on BraTS 2018 [8]. Pei et al. (2020) extended CANet to jointly perform segmentation, classification, and survival prediction [1]. Attention mechanisms and multi-scale feature fusion have been incorporated into CNNs to enhance performance. Abdusalomov et al. (2023) enhanced YOLOv7 with attention modules and components for multi-scale feature aggregation, achieving 99.5% accuracy in localizing tumors on over 10,000 MRI scans [9]. Liu et al. (2023) highlighted CNN approaches, such as U-Net, data augmentation, transfer learning, and challenges like tumor heterogeneity and ground truth annotations, to categorize brain tumors. [10]. Gao et al. (2022) developed a deep learning model to classify brain tumors from MRI images, achieving an AUROC of 0.99 for tumor detection and 0.97 for tumor typing on over 10,000 scans [11]. Jeong et al. (2019) trained a random forest using delta radiomic techniques, that classified glioblastomas into high and low grades with 90% accuracy [12]. Tabassum et al. (2023) reviewed applications of radiomics and ML to classify brain tumors, highlighting CNNs and integrating radiomics with other omics data [13]. Kumar et al. (2023) used radiomics and ML models like random forest and XGBoost to classify glioma grades from MRI, achieving 92% accuracy for low vs high grades [14]. Lam et al. (2022) developed a ML model to predict glioma tumor mutation from MRI images, achieving an AUROC of 0.84 for high vs low TMB classification [15]. Qureshi et al. (2023) developed a radiogenomic model fusing mpMRI and genomic data to predict MGMT promoter methylation status in glioblastoma, achieving an AUROC of 0.92 [16].

Despite these advances, the existing methods still face challenges, such as handling class imbalance, effectively integrating multi-modal MRI data, and balancing high performance with computationally efficient architectures suitable for real-time use. Moreover, most approaches focus on either segmentation or classification separately, without a unified framework that models the interplay between these tasks.

The main objective of this study is to incorporate a deep learning method that combines brain tumor segmentation, classification, and grading in a unified end-to-end system. This will involve using a dual CNN architecture with attention mechanisms and transfer learning to achieve high accuracy and robustness while keeping computational efficiency. The proposed model aims to offer insights into the biological basis of learned radiomic features and has the potential to be used in clinical settings as an AI-powered tool to support decision-making in diagnosing and treating brain cancer.

2. Methods

The process of research methodology involves several stages as shown in Fig. 1. It starts with data preprocessing and proceeds through the development of a model, training, evaluation, and ultimately, testing and deployment.

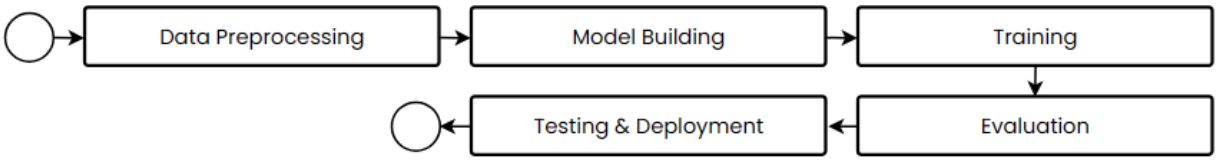


Fig. 1. Research Methodology Flowchart

2.1. Data Preprocessing

The dataset consisting of 7023 brain MRI scans from Kaggle is divided into training and testing subsets, with the latter further split into validation and test groups using stratified sampling [17]. Pandas DataFrames are created containing file paths and class labels. Data visualization using Seaborn countplots is performed to analyze class distributions. The Keras ImageDataGenerator is utilized for preprocessing, with data augmentation techniques like rescaling and brightness adjustment applied to the training and validation sets, while the testing set is only rescaled. Countplots were used to visualize the class distributions in the training and testing datasets, which include glioma, meningioma, pituitary, and no tumor. These plots were created using the Seaborn library and are shown in Fig. 3 and 4, respectively.

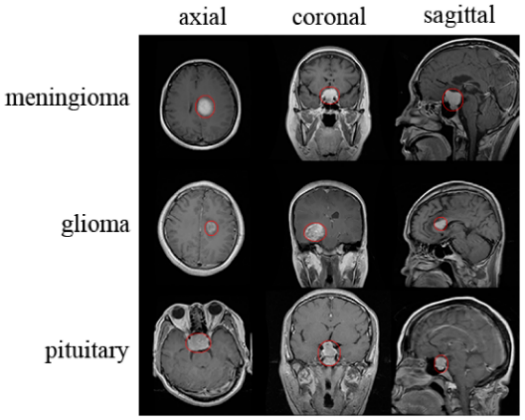


Fig. 2. MRI Images of Different Brain Tumor Types

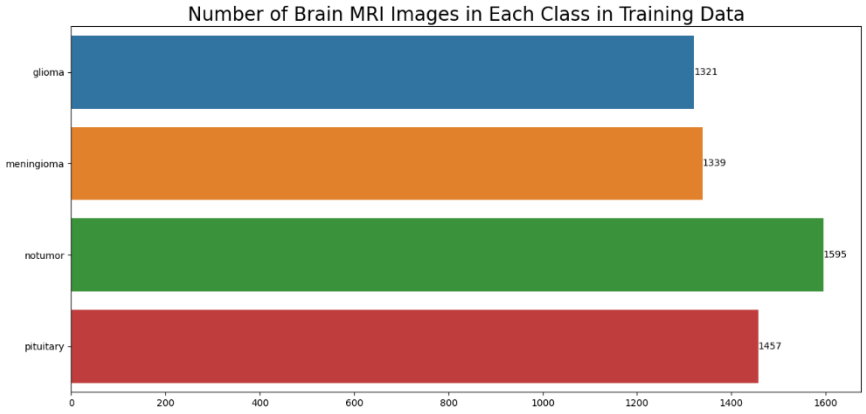


Fig. 3. Number of Brain MRI Images in Each Class in Training Data

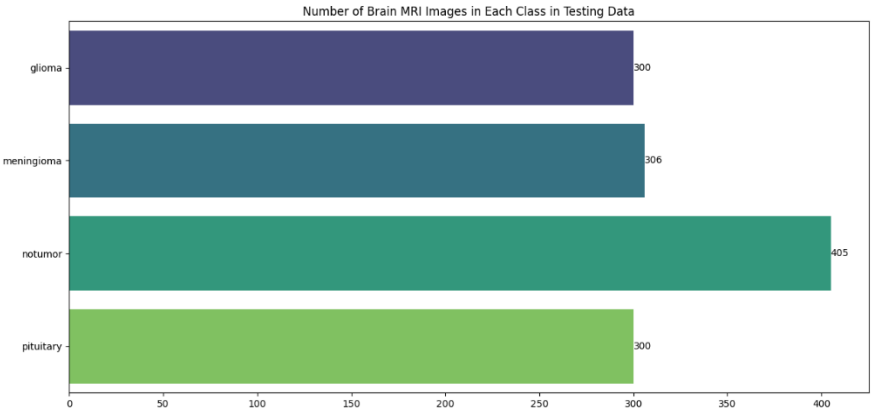


Fig. 4. Number of Brain MRI Images in Each Class in Testing Data

2.2. Model Building

The model uses pre-trained Xception on ImageNet as the base model for transfer learning, with the top layers removed and the remaining layers frozen. Flatten layer, Dropout layers, and Dense layers with ReLU and Softmax activations are also added to the base model. Adamax optimizer, categorical cross-entropy loss, and evaluation metrics like accuracy, precision, and recall, are then used to compile the model.

2.3. Training

The model was trained for 10 epochs using the Keras "fit" function, and the training and validation data generators as inputs. The model weights are adjusted based on the validation set's performance. The model also uses Matplotlib to visualize the training and validation metrics over the epochs, which includes loss, accuracy, precision, and recall, identifying the best epoch for each metric.

2.4. Evaluation

The model calculated loss and evaluation metrics (accuracy, precision, and recall) of the training, validation, and testing datasets using the Keras "evaluate" function. The "predict" function printed out its predictions on the testing

set. Seaborn was used to show the confusion matrix of the model. Then, the performance metrics of each class were generated in the form of a classification report.

2.5. Testing and Development

The trained model is saved in the HDF5 file format using the Keras "save" function, facilitating easy loading and deployment through the "load_model" function. A custom "description" function provides detailed information about the predicted tumor type, including an overview, symptoms, treatments, and relevant specialists. The "predict" function loads and preprocesses an MRI image, generates predictions using the model, displays the probability distribution for each class, and calls the "description" function for additional information. Individual predictions are showcased by invoking the "predict" function on a few MRI images from the testing set, exhibiting the image, probability distribution, and an elaborate description of the predicted tumor type.

3. Results

3.1. Model Training and Evaluation

The training and validation metrics (loss, accuracy, precision, and recall) over the 10 epochs are shown in Fig. 5. The plots indicate the best epoch for each metric.

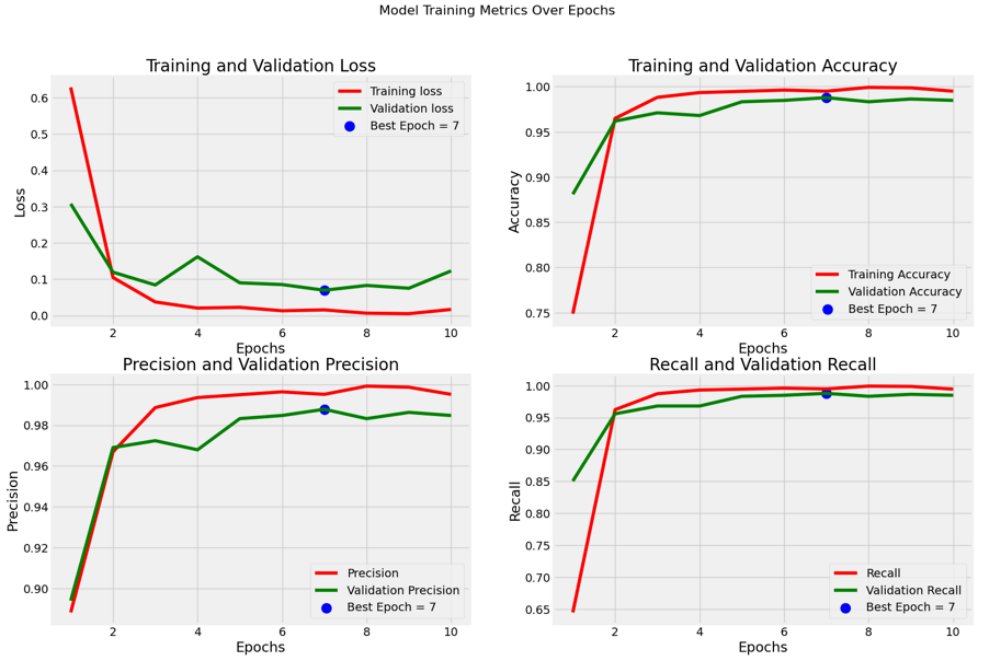


Fig. 5. Model Training Metrics Over Epochs

Table 1 displays the loss and evaluation metrics, which include the accuracy, precision, and recall of each set.

Table 1. The Loss and Evaluation Metrics

	Loss	Accuracy
Train	0.0201	99.53%
Validation	0.1201	98.47%

Test	0.0970	97.87%

3.2. Model Prediction and Evaluation

The predictions of the model on the testing set were acquired through the use of the "predict" function. To generate a confusion matrix, the predicted labels were compared with the true labels (actual tumor types). This matrix was then visualized using a heatmap from the Seaborn library, as can be seen in Fig. 6.

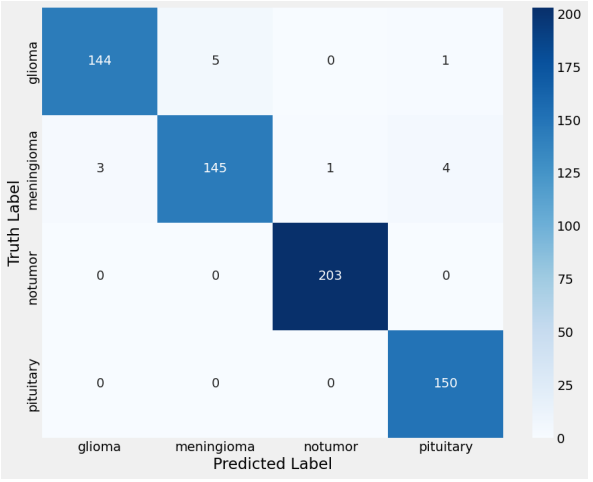


Fig. 6. Confusion Matrix for Brain Tumor Classification

Furthermore, the model's performance for each class, which includes precision, recall, and F_1 score, was shown in Table 2 in the form of a classification report.

Table 2. Classification Report

	Precision	Recall	F_1 Score	Support
Glioma	0.98	0.96	0.97	150
Meningioma	0.97	0.95	0.96	153
No Tumor	1.00	1.00	1.00	203
Pituitary	0.97	1.00	0.98	150
Accuracy			0.98	656
Macro Average	0.98	0.98	0.98	656
Weighted Average	0.98	0.98	0.98	656

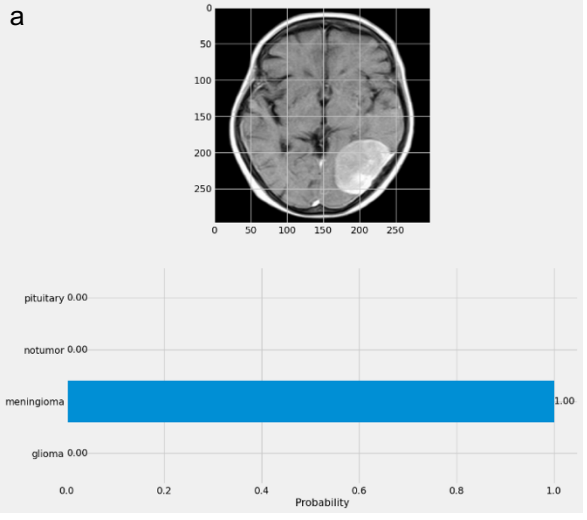
4. Discussion

The proposed model based on deep learning is highly efficient in identifying and classifying brain tumors from MRI scans, achieving 97.87% accuracy on the testing set. The model's performance for each tumor type can be analyzed through the confusion matrix (Fig. 6) and classification report (Table 2). The model exhibits high precision and recall scores across all classes, which indicates its ability to classify tumors and minimize false positives and false negatives accurately. One noteworthy observation is the model's exceptional performance in detecting the absence of tumors (no tumor) and pituitary tumors with perfect precision (recall of 1.00), glioma tumors with a

recall of 0.96, and meningioma tumors with a recall of 0.95. This is particularly significant in clinical settings as it can help reduce unnecessary invasive procedures or treatments for patients without brain tumors.

Utilizing the pre-trained Xception model's weights, derived from the ImageNet dataset, allowed for effective feature extraction from MRI images without the need for extensive training on a limited dataset. The application of data augmentation techniques, such as rescaling and brightness adjustment, helped improve the model's generalization. The strategic combination of the Xception base model, Flatten layer, Dropout layers, and Dense layers with suitable activation functions enabled the model to learn complex patterns for tumor classification. The efficient training and convergence of the model can be achieved by the usage of the Adamax optimizer and categorical cross-entropy loss function.

The model also includes the “prediction” and “description” functions, which enhance the interpretability and usability of the model's predictions. The “prediction” function is implemented to load and preprocess a given MRI image, obtain the model's predictions, and visualize the probability distribution across the classes. It allows users to input new MRI scans and receive the model's classification along with the corresponding probability scores for each tumor type. This function facilitates the practical application of the model in clinical settings, enabling healthcare professionals to evaluate new cases and make informed decisions based on the model's predictions. The “description” function provides a detailed description of the predicted tumor type, including an overview, symptoms, treatments, and relevant specialists. This function enhances the interpretability of the model's output by providing additional context and information about the predicted condition. It can assist healthcare professionals in understanding the implications of the predicted tumor type and guide them in developing appropriate treatment plans or seeking further consultations with relevant specialists.



b

```
Predicted Tumor Type:
1. Meningioma
Probability: 100.00%

a. Overview
A usually non-cancerous tumor that arises from the membranes surrounding the brain and spinal cord.

- It is not clear what causes a meningioma. Radiation therapy, female hormones, and genetics may play a role. In most cases, the condition is non-cancerous.
- Symptoms may include changes in vision, headaches, hearing loss, and seizures.
- A small, slow-growing meningioma that is not causing signs or symptoms may not require treatment. When required, treatment might involve surgery or radiation.

[!] Rare
Fewer than 150 thousand cases per year

- Treatable by a medical professional
- Requires a medical diagnosis
- Lab tests or imaging are always required
- Chronic: can last for years or be lifelong

b. Symptoms
[!] Requires a medical diagnosis
Symptoms may include changes in vision, headaches, hearing loss, and seizures.

[!] People may experience:
- Visual: double vision or vision loss
- Eyes: swelling of the optic disc or unequal pupils
- Also common: headache, impaired coordination, or seizure


c. Treatments
Note: Treatment depends on stage
Treatments include surgery, radiation therapy, chemotherapy, and targeted molecular therapy.

[!] Supportive Care
- Monitoring
Regular follow-up to evaluate for improvement.

[!] Surgery
- Craniotomy
Brain surgery in which a piece of bone is removed from the skull.
- Microsurgery
Surgery is performed on very small blood vessels and nerves using a microscope.

[!] Medical Procedure
- Radiation Therapy
Treatment that uses X-rays and other high-energy rays to kill abnormal cells.

d. Specialists
- Neurosurgery
Specialises in nervous system disorders.
- Radiation Oncology
Treats and manages cancer by prescribing radiation therapy.
- Neurology
Treats nervous system disorders.
- Oncology
Specialises in cancer.
```

Fig. 7. (a) Example of the Prediction Bar Chart; (b) Example of the Prediction Description

Fig. 7. (a) and (b) are examples of the output generated by the “prediction” and “description” functions, respectively. These figures are not obtained from the testing phase of the code but rather serve as visual representations of the functions’ outputs. Fig. 7. (a) shows an example of the prediction bar chart, which displays the model’s probability distribution across the different tumor classes for a given input MRI image. Fig. 7. (b) illustrates an example of the prediction description, providing detailed information about the predicted tumor type, such as an overview, symptoms, treatments, and relevant specialists.

While the proposed model demonstrates promising results, it is important to acknowledge potential limitations and areas for further improvement. One limitation is the dependence on a single imaging modality (MRI scans). The inclusion of additional imaging data, such as computed tomography (CT) scans or positron emission tomography (PET) images, could enhance the model's performance and provide more comprehensive information for diagnosis. Furthermore, the integration of multi-omics data, including genomic, transcriptomic, and proteomic information, could help the model in identifying molecular biomarkers, as well as provide insights into cancer biology and treatment response. This aligns with the emerging field of radiogenomics, which aims to establish relationships between imaging features and genomic profiles [18]. It is also crucial to validate the model's performance using diverse datasets of tumor types, patient demographics, as well as imaging protocols, to ensure the model's generalizability across different clinical settings [19], [20].

Despite these limitations, the proposed deep learning model represents a significant step towards non-invasive and accurate brain tumor diagnosis. By integrating bioinformatics techniques and deep learning algorithms, this approach has the potential to be used in treatment planning, and surgical decision-making, ultimately contributing to improved patient outcomes and personalized medicine in neuro-oncology.

References

[1] L. Pei, L. Vidyaratne, M. M. Rahman, and K. M. Iftekharuddin, "Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images," *Sci Rep*, vol. 10, no. 1, p. 19726, Nov. 2020, doi: 10.1038/s41598-020-74419-9.

[2] M. Hammad, M. ElAffendi, A. A. Ateya, and A. A. Abd El-Latif, "Efficient Brain Tumor Detection with Lightweight End-to-End Deep Learning Model," *Cancers (Basel)*, vol. 15, no. 10, p. 2837, May 2023, doi: 10.3390/cancers15102837.

[3] S. E. Nassar, I. Yasser, H. M. Amer, and M. A. Mohamed, "A robust MRI-based brain tumor classification via a hybrid deep learning technique," *J Supercomput*, vol. 80, no. 2, pp. 2403–2427, Jan. 2024, doi: 10.1007/s11227-023-05549-w.

[4] I. Rethemiotaki, "Brain tumour detection from magnetic resonance imaging using convolutional neural networks," *Współczesna Onkologia*, vol. 27, no. 4, pp. 230–241, 2023, doi: 10.5114/wo.2023.135320.

[5] I. Shin *et al.*, "Development and Validation of a Deep Learning–Based Model to Distinguish Glioblastoma from Solitary Brain Metastasis Using Conventional MR Images," *American Journal of Neuroradiology*, vol. 42, no. 5, pp. 838–844, May 2021, doi: 10.3174/ajnr.A7003.

[6] L. Tariciotti *et al.*, "A Deep Learning Model for Preoperative Differentiation of Glioblastoma, Brain Metastasis and Primary Central Nervous System Lymphoma: A Pilot Study," *Front Oncol*, vol. 12, Feb. 2022, doi: 10.3389/fonc.2022.816638.

[7] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Med Inform Decis Mak*, vol. 23, no. 1, p. 16, Jan. 2023, doi: 10.1186/s12911-023-02114-6.

[8] L. Sun, S. Zhang, H. Chen, and L. Luo, "Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning," *Front Neurosci*, vol. 13, Aug. 2019, doi: 10.3389/fnins.2019.00810.

[9] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," *Cancers (Basel)*, vol. 15, no. 16, p. 4172, Aug. 2023, doi: 10.3390/cancers15164172.

[10] Z. Liu *et al.*, "Deep learning based brain tumor segmentation: a survey," *Complex & Intelligent Systems*, vol. 9, no. 1, pp. 1001–1026, Feb. 2023, doi: 10.1007/s40747-022-00815-5.

[11] P. Gao *et al.*, "Development and Validation of a Deep Learning Model for Brain Tumor Diagnosis and Classification Using Magnetic Resonance Imaging," *JAMA Netw Open*, vol. 5, no. 8, p. e2225608, Aug. 2022, doi: 10.1001/jamanetworkopen.2022.25608.

[12] J. Jeong *et al.*, "Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images," *Quant Imaging Med Surg*, vol. 9, no. 7, pp. 1201–1213, Jul. 2019, doi: 10.21037/qims.2019.07.01.

[13] M. Tabassum, A. Al Suman, E. Suero Molina, E. Pan, A. Di Ieva, and S. Liu, "Radiomics and Machine Learning in Brain Tumors and Their Habitat: A Systematic Review," *Cancers (Basel)*, vol. 15, no. 15, p. 3845, Jul. 2023, doi: 10.3390/cancers15153845.

[14] A. Kumar *et al.*, "Machine-Learning-Based Radiomics for Classifying Glioma Grade from Magnetic Resonance Images of the Brain," *J Pers Med*, vol. 13, no. 6, p. 920, May 2023, doi: 10.3390/jpm13060920.

[15] L. H. T. Lam, N. T. Chu, T.-O. Tran, D. T. Do, and N. Q. K. Le, "A Radiomics-Based Machine Learning Model for Prediction of Tumor Mutational Burden in Lower-Grade Gliomas," *Cancers (Basel)*, vol. 14, no. 14, p. 3492, Jul. 2022, doi: 10.3390/cancers14143492.

[16] S. A. Qureshi *et al.*, "Radiogenomic classification for MGMT promoter methylation status using multi-omics fused feature space for least invasive diagnosis through mpMRI scans," *Sci Rep*, vol. 13, no. 1, p. 3291, Feb. 2023, doi: 10.1038/s41598-023-30309-4.

[17] M. Nickparvar, "Brain Tumor MRI Dataset." Accessed: Feb. 06, 2024. [Online]. Available: https://www.kaggle.com/dsv/2645886

[18] M. R. Tomaszewski and R. J. Gillies, "The Biological Meaning of Radiomic Features," *Radiology*, vol. 298, no. 3, pp. 505–516, Mar. 2021, doi: 10.1148/radiol.2021202553.

[19] P. Palee, B. Sharp, L. Noriega, N. Sebire, and C. Platt, "Heuristic neural network approach in histological sections detection of hydatidiform mole," *Journal of Medical Imaging*, vol. 6, no. 04, p. 1, Nov. 2019, doi: 10.1117/1.JMI.6.4.044501.

[20] S. Monti, M. E. Truppa, S. Albanese, and M. Mancini, "Radiomics and Radiogenomics in Preclinical Imaging on Murine Models: A Narrative Review," *J Pers Med*, vol. 13, no. 8, p. 1204, Jul. 2023, doi: 10.3390/jpm13081204.