Classification of MRI Brain Tumor Images using Deep Learning Segment Anything Model for segmentation and Deep Convolution Neural Network

Mani Abedini *

Head of Data Science and AI, AW Rostamani, UAE.

World Journal of Advanced Research and Reviews, 2024, 23(02), 1153–1161

Publication history: Received on 05 July 2024; revised on 12 August 2024; accepted on 14 August 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.23.2.2469

Abstract

Brain tumors pose a significant health challenge by putting pressure on healthy parts of the brain or spreading into other areas and blocking the flow of fluid around the brain. Thus, identifying and categorizing the tumor is crucial for delivering effective treatment, especially if detected early. This means the tumor is smaller, and treatment is more effective, less invasive, and has fewer side effects.

In recent years, many researchers have developed computer vision, and more specifically, deep learning methods, to automate the analysis of brain MRI scans. These methods enable efficient processing and improve the accuracy of detecting small tumors.

This paper aims to propose a deep-learning method for classifying brain tumors. In this work, the input image goes through two subprocesses: first, object detection to identify the tumor's location. Then, a fine-tuned Segment Anything Model (SAM) was applied to extract the lesion from the background. Finally, deep learning Convolution Neural Network (CNN), is applied to the cropped image for classification. This method will help doctors and researchers detect tumors at the initial stages

Keywords: Brain tumor; Image processing; Feature extraction; Machine learning; MRI image Classification; Computer vision; Cancer classification; Convolution Neural Network; Cross-modal deep learning

1. Introduction

A brain tumor is a mass or cluster of abnormal brain cells that affect the central nervous system in the brain. Brain tumors can be life-threatening if spread to other areas or become cancerous. An estimated 1 million Americans live with a primary brain tumor, and approximately 28% of such cases are malignant (cancerous) [1]. An early diagnosis is crucial for successful treatment planning. The initial evaluation is performed by oncologists using Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans to identify the abnormal mass and detect the type of tumor. Generally, types of brain tumors are classified based on the affected area: Meningioma, Glioma, and Pituitary [2,3].

Meningiomas are from the meninges, the protective tissue layers surrounding the brain and spinal cord. Gliomas are the most common type of primary brain tumor and originate from the glial cells in the brain. Glioblastoma is the most aggressive subtype of Glioma. **Pituitary** adenomas are tumors from the pituitary gland, a small gland located at the base of the brain that aims to produce hormones. As a result, these tumors impact hormonal imbalances. See Figure 1 for some examples of MRI images.

* Corresponding author: Mani Abedini *

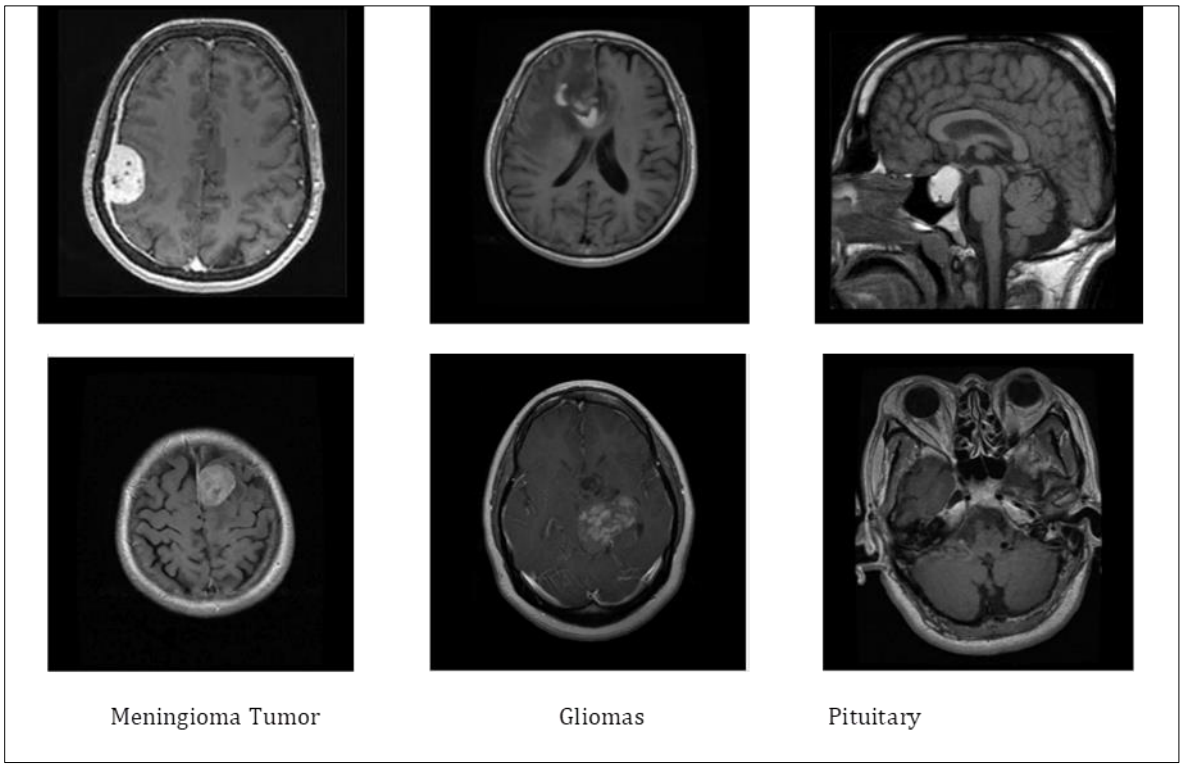

Figure 1 Some examples of MRI images for different types of Brian Tumors

Modern imaging methods such as MRI and CT scan can be used to identify the tumor location as well as assess the anatomy of the lesion, physiology, and metabolic activity. Early detection of cancer cases is vital for more effective treatment and increasing survival rates. However, early detection mostly depends on the radiologist's expertise [4]. Machine Learning has been explored to assist medical expert for more accurate diagnosis, for example common classification methods for diabetes cases [5]. However, in recent years, computer-aided diagnosis (CAD) systems leveraging computer vision techniques, especially Deep Learning, have shown promising capabilities to automate analyzing medical images and provide critical insight to medical professionals [6,7]. Much research has been conducted to develop tools for identifying tumors using such deep-learning methods. In the next section, we review some of these works.

In the next section, the related work is reviewed. The proposed method is presented in Section 3. The experimental settings and results are shown in Section 4, it follows with conclusion section.

2. Related Works

In recent years, many papers have been published using computer vision methods, specifically Deep Neural Networks, to classify brain tumors. Mzoughi et al. have used a volumetric Convolutional Neural Network (CNN) to classify glioma using T1 contrast-enhanced images [8]. Pereira et al. used CNN to classify and identify the tumor grade [9]. Ali et al. also used CNN; however, they proposed using combined cropped tumor lesions and uncropped brain images as the input to CNN. They also proposed a novel CNN architecture rather than using a popular transfer learning approach and fine-tuning the pre-trained CNN. Their results suggest that using cropped and uncropped images together produces higher accuracy, higher sensitivity, and higher specificity than each image module individually [10]. Pei et al. have used segmentation and classification methods for glioma classification. In his approach, semantic segmentation was applied in the first step to extract the region of interest (tumor location), followed by the classification method to classify the tumor cancer [11].

Also, there are many published works that exist in the literature that leverage out-of-the-shelf deep neural network architectures for classifying Brain MRI images. For example, Mobile Net [12], VGG-16 NADE Network [13], Alex Net [14], Dense Net [15], Dense Efficient-Net [16, 17, 18]. Rehman et al. have compared three different CNN models in their work: Alex Net, GoogLe Net, and VGG Net. Their paper suggests that fine-tuning CNN and data preparation (such as augmentation) is critical to achieving higher accuracy. They reported VGG 16 with carefully tuned parameters achieved

up to 98.69 accuracy. In summary, all of these studies depict that general-purpose CNNs can be used particularly for medical imaging classification, such as MRI Brain Tumor classification.

Kang et al., in their work, focused on feature engineering by combining the top 3 feature vectors extracted from 13 different pre-trained deep convolution neural networks, feeding to SVM with RBF kernel, and reporting better results than using each network individually [19]. Heba et al. combined deep learning features and discrete wavelet transform (DWT) features as part of feature extraction. They used fuzzy c-mean clustering to segment the brain tumor and extract the lesion from the background image [20]. Widhiarso et al., on the other hand, used CNN for classification; they proposed using four different classic computer vision feature extraction methods: Energy, Correlation, Contract, and Homogeneity from four different angles (0°, 45°, 90°, and 135°). All features are stacked and given to CNN for classification [21].

Khawaldeh et al. proposed a modified version of Alex Net CNN [22]. They used the entire MRI image for classification without any segmentation or localizing of the tumor. In their experiments they have reported 91% accuracy. On the other hand, Mittal et al. in their paper applied deep learning-based segmentation using a combination of Stationary Wavelet Transform (SWT) and new Growing Convolution Neural Network (GCNN) [23]. Using segmentation before classification has been shown to improve the accuracy of the classification. Sajjad, Muhammad et al. proposed extensive data augmentation can be critical for improving the accuracy of CNN for brain tumor classification. They used pre-trained VGG-19 CNN architecture for classification with an average of 87.38% accuracy. In the same experiment, they used extensive data augmentation, which resulted in a significantly increased overall accuracy of 90.67% [24]. Saeedi et al. developed a new 2D Convolutional Neural Network (CNN) and a convolutional auto-encoder network. The customized 2D CNN has eight convolution layers and a hierarchical network using a 2*2 kernel. It also has four pooling layers. The modified auto-encoder network proposed in this paper consists of a convolutional auto-encoder network and a classification layer. They have reported an accuracy of 96% using the proposed model [25].

3. Methodology

3.1. Dataset

This study used two Kaggle datasets containing Brain Tumor MRI images. The first data set has 5249 images with four categories: glioma, meningioma, pituitary and no tumor. The data set is divided into training and validation sets [26]. In this set, each image has a label and a bounding box associated with the tumor location in the MRI image. As such we used this data set to train our segmentation algorithm to identify tomur location. The detailed stats of the first data are shown in Table 1.

The second data set contains 7023 human brain MRI images, classified into the same four classes. However, the second data set does not have bounding box annotation [27]. The images of this set are gathered from three datasets: figshare [28], SARTAJ dataset [29], and Br35H [30]. The data set is split into training and test sets with 5712 and 1311 images, respectively. We used the second data set for training the classification layer. The distribution of images per training and test set for this batch is in reported in Table 2.

Table 1 The statistics of the first data set, with label and bounding box coordination

Training Set	Glioma: 1,153 images
	Meningioma: 1,449 images
	No Tumor: 711 images
	Pituitary: 1,424 images
Validation Set	Glioma: 136 images
	Meningioma: 140 images
	No Tumor: 100 images
	Pituitary: 136 images

Table 2 The statistics of the second data set with label only

Training Set	Glioma: 1,321 images
	Meningioma: 1,339 images
	No Tumor: 1,595 images
	Pituitary: 1,457 images
Validation Set	Glioma: 300 images
	Meningioma: 306 images
	No Tumor: 405 images
	Pituitary: 300 images

3.2. Pre-processing

For data preparation, grayscale transformation and noise removal using a median filter were applied to MRI images. Noise reduction is a quite common approach in computer vision methods. We also used typical resizing, cropping, rotation, zooming, flipping and color adjustments while image augmentation was applied. All images are resized to 224 × 224 to be consistent with the input layer of our CNN models. The image augmentation requires generating a good amount of annotated data so we can retrain the deep neural networks. Since annotating medical images requires to be conducted by experienced medical professionals, data acquisition takes time and very expensive. Generating more images from existing annotated data is the best cost-effective way to overcome the situation.

In our experiments, all images were scaled with 1/255. We also allow random rotation between 0 and 45 degrees. The zoom level was between 0.5 and 2; numbers below 1.0 result in zooming out, and numbers bigger than 1.0 will magnify. We also allow random adjustment of brightness. The random noise in brightness will help the network be less sensitive to specific image brightness and try to learn the underlying patterns associated with cancer lesions.

3.3. Proposed model

In this paper, we proposed a multi-step process for Brain tumor classification. The first step is to identify the location of the lesion. For this step, we have explored various states of the art deep learning object detection methods. All models have been fine-tuned using the first Brain Tumor data sets. The data set that has bounding box annotation representing the coordinate of Brain Tumors in each image. Among all models, YOLOv8 (You Only Look Once), a highly popular deep-learning object detection model, have demonstrated better performance than others.

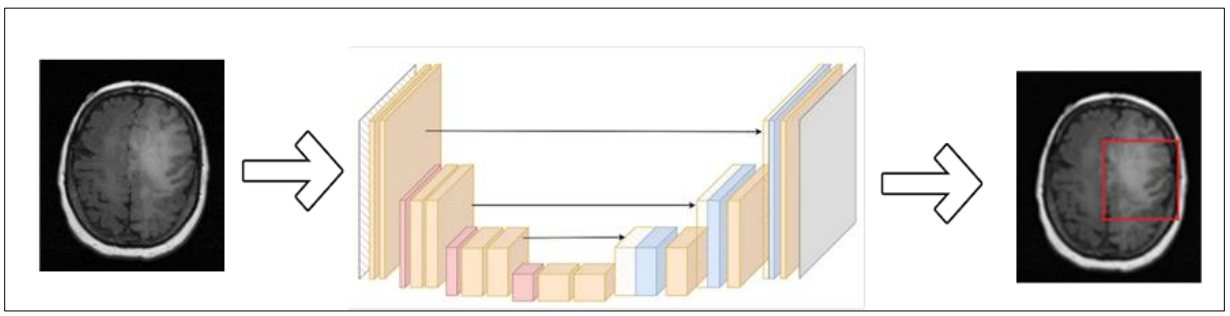

Figure 1 Identifying the location of Brain Tumor, using Semantic Segmentation network

In the second step, we used Meta's pre-trained Segment Anything Model (SAM) to segment the border of lesions. The model has a very high accuracy and has been widely used in many segmentation applications. However, the model requires the bounding box around the object as input. For this reason, we have used Deep Learning Object Detection model in the first step to identify the Tumor and the bounding box coordination fit around the Tumor. The image along with the bounding box coordinates are given to SAM. The output is the

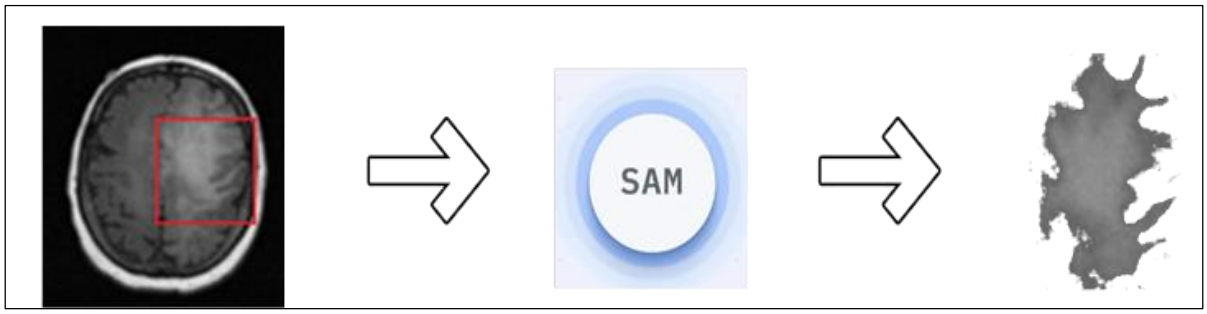

Figure 2 Lesion Segmentation using Segment Anything Model (SAM) and the lesion bonding box coordinates

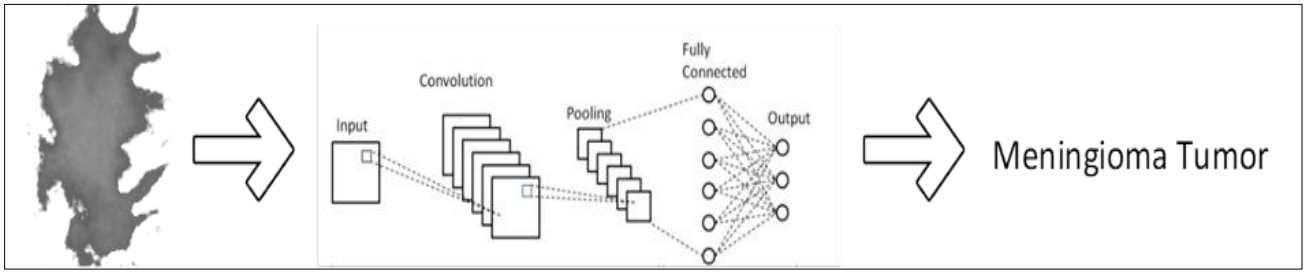

Figure 3 Tumor Classification using CNN on the cropped image of the lesion without background

3.4. Evaluation Metrics

To evaluate the performance of the proposed methods, four widely used metrics have been measured in our experiments: Accuracy (1), Precision (2), Recall (3), and F1-Score (4). Please see the formula for each metric below:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \text{ ..... (1)}$$

$$Precision = \frac{TP}{TP+FP} \text{ .....(2)}$$

$$Recall = \frac{TP}{TP+FN} \text{ .....(3)}$$

$$F1-Score = \frac{2 \times Recall \times Precision}{Recall+Precision} \text{ ..... (4)}$$

- TP (True Positive) represents the number of correctly predicted positive cases.
- TN (True Negative) represents the number of correctly predicted negative cases.
- FP (True Positive) represents the number of incorrectly predicted positive cases.
- FN (True Negative) represents the number of incorrectly predicted negative cases.

Accuracy is the most commonly used measurement of the accurate model, defined as the portion of actual positive and negative cases over all measured cases.

Precision defines the portion of positive predictions that have been correctly identified.

Recall, also known as sensitivity or true positive rate, on the other hand, measures the portion of actual positive area that has been correctly predicted as positive.

F1-score is the harmonic mean of the precision and recall.

3.5. Experiments

In our proposed approach, the first step is developing an accurate Deep Learning Object detection network. For this purpose, we have explored various state-of-the-art Object detection models: YOLO (You Only Look Once) [30], EfficientDet [31], RetinaNet [32], and Faster R-CNN [33]. For training data, we used the first data set explained in Section 3.1. All methods have been trained and fine-tuned. However, YOLOv8 demonstrated better performance than all other methods.

The second phase is classification, after lesion segmentation and removing background from cropped images. Many recently introduced CNN models have demonstrated high accuracy in various benchmark data sets. We have selected a couple of widely used models in our experiments. A combination of newly developed models and well-established CNN architecture: InceptionV3 [34], DenseNet [35], Xception [36], VGG16 [37], EfficientNet [38], Coca-ViT [39]. Contrastive Captioner (CoCa) is one of the recently developed classification models that has the highest score on the ImageNet leaderboard. This network has a pretrain image-text encoder-decoder foundation model. CoCa is unifying natural language supervision for representation learning, applying a contrastive loss between unimodal image and text embeddings, and combining multimodal learning such as image and text classification.

4. Results and discussion

In this section, we present the results of the experiment and discuss the effectiveness of the method for each phase of the process.

For Object Detection phase, we conducted comprehensive experiments using 4 popular deep learning object detection models (YOLO, EfficientDet, RetinaNet, and Faster R-CNN). The results are presented in Table 3. The results suggest that YOLOv8 is performing better than others for object detection. However, the performance of most models are very close, suggesting the object detection is effective and regardless of which model, definitely can help identifying the location of suspicious lesion.

Table 3 The performance of Various different Object Detection Models to identify the location of Brain Tumor

Deep Learning Object Detection Model	Accuracy	Precision	Recall	F1-Score
YOLOv8	95.27%	89.12%	78.69%	83.58%
EfficientDet	93.88%	88.56%	72.48%	79.71%
RetinaNet	89.55%	87.34%	78.33%	82.58%
Faster R-CNN	80.63%	85.45%	70.62%	77.33%

In the classification phase of our experiments, we used mostly common Deep learning CNNs for classification: InceptionV3, DenseNet, Xception, VGG16, EfficientNetV2 B3, and CoCa. Most of these models are already implemented in Keras, so we used those implementations. In the case of CoCa, we used the CoCa implementation code in OpenCLIP [40]. The results of classification have been presented in Table 4.

Table 4 The Performance of various different Deep Learning Convolution Neural Network (CNN) models for image classification of Brain Tumor

Deep Learning Object Detection Model	Accuracy	Precision	Recall	F1-Score
InceptionV3	96.43%	97.32%	93.47%	95.80%
DenseNet	94.83%	98.21%	94.56%	96.35%
Xception	96.74%	96.85%	92.76%	94.76%
VGG16	94.38%	95.42%	90.10%	92.68%
EfficientNetV2 B3	96.87%	98.97%	96.50%	97.71%
CoCa	97.60%	99.04%	97.26%	98.14%

The results of classification phase, suggest that most states-of-the-art deep learning classification models are capable of classifying Brain Tumor images with above 90% accuracy. However, CoCa present higher accuracy, F1-Score among all other models, which suggest that the cross modal transfer learning approach such as CoCa which has been trained on image and text labels, has more robust pretrained network and can be fine-tuned to a very specific medical image classification tasks such as MRI Brain Tumor classification

5. Conclusion

In this paper, we utilize a three-step deep learning-based process to classify MRI Brain Tumor images. The first step uses Deep Object Detection models to identify the location of the lesion. We have explored various Deep Learning Object Detection models, fine-tuned them on our MRI Brain Tumor data set. The newly enhanced YOLO model (v8) has outperform other models in this task. In the next step, the Bounding Box surrounding the detected Lesion will be passed to the Semantic Anything Model (SAM) to identify the border of the Lesion. By extracting the background image from the lesion itself, we allow the image classification model to focus on the property of the lesion and its patterns. For classification task, we have fine-tuned various widely used Deep Convolution Neural Network models on our second acquired data set. The results shown CoCa performance on image classification is better than other CNN models for MRI Tumor classification.

References

[1] Porter KR, McCarthy BJ, Freels S, Kim Y, Davis FG. Prevalence Estimates for Primary Brain Tumors in the United States by Age, Gender, Behavior, and Histology. Neuro Oncol. 2010.; 12(60):520-527.

[2] International Agency for Research on Cancer. Available online: https://gco.iarc.fr/ (accessed on 1 July 2019).

[3] Brain Tumor Basics. Available online: https://www.thebraintumourcharity.org/ (accessed on 1 July 2019).

[4] Afshar, P.; Plataniotis, K.N.; Mohammadi, A. Capsule Networks for Brain Tumor Classification Based on MRI Images and Coarse Tumor Boundaries. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1368–1372.

[5] Abedini, Mani & Bijari, Anita & Banirostam, Touraj. (2020). "Classification of Pima Indian Diabetes Dataset using Ensemble of Decision Tree, Logistic Regression and Neural Network". IJARCCE. 9. 10.17148/IJARCCE.2020.9701.

[6] Mani Abedini, Noel CF Codella, Jonathan H Connell, Rahil Garnavi, Michele Merler, Sharath Pankanti, John R Smith, Tanveer Syeda-Mahmood, "A generalized framework for medical image classification and recognition". IBM Journal of Research and Development, 2015.

[7] Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R. (2015). Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images. In: Zhou, L., Wang, L., Wang, Q., Shi, Y. (eds) Machine Learning in Medical Imaging. MLMI 2015. Lecture Notes in Computer Science(), vol 9352. Springer, Cham. https://doi.org/10.1007/978-3-319-24888-2_15

[8] Mzoughi, H. et al. Deep multi-scale 3d convolutional neural network (cnn) for mri gliomas brain tumor classification. J. Digit. Imaging 33, 903–915 (2020).

[9] Pereira, Sérgio, et al. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE transactions on medical imaging, Vol 35, pp 1240-1251, 2016. DOI: 10.1109/TMI.2016.2538465.

[10] Ali Mohammad Alqudah, Hiam Alquraan, Isam Abu Qasmieh, Amin Alqudah, Wafaa Al-Sharu, Brain Tumor Classification Using Deep Learning Technique - A Comparison between Cropped, Uncropped, and Segmented Lesion Images with Different. International Journal of Advanced Trends in Computer Science and Engineering. Vol 8, No.6.

[11] Pei, L. et al., "Brain tumor classification using 3d convolutional neural network" . In International MICCAI Brain lesion Workshop, 335–342 (2019).

[12] Si-Yuan Lu, Shui-Hua Wang, Yu-Dong Zhang, "A classification method for brain MRI via MobileNet and feedforward network with random weights", Pattern Recognition Letters, Volume 140, 2020, Pages 252-260, ISSN 0167-8655

[13] C Mukesh Kumar, J Sree Sankar, "Comparative Analysis of Convolutional Neural Networks for Brain Tumor Detection: A Study of VGG16, ResNet, Inception, and DenseNet Models", 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp.41-46, 2024.

[14] Rajat Mehrotra, M.A. Ansari, Rajeev Agrawal, R.S. Anand, “A Transfer Learning approach for AI-based classification of brain tumors” , Machine Learning with Applications, Volume 2, 2020, 100003, ISSN 2666-8270, https://doi.org/10.1016/j.mlwa.2020.100003.

[15] Özkaraca O, Bağrıaçık Oİ, Gürüler H, Khan F, Hussain J, Khan J, Laila UE. “Multiple Brain Tumor Classification with Dense CNN Architecture Using Brain MRI Images”. Life (Basel). 2023 Jan 28;13(2):349. doi: 10.3390/life13020349. PMID: 36836705; PMCID: PMC9964555.

[16] Pallavi Priyadarshini, Priyadarshi Kanungo, Tejaswini Kar, “Multigrade brain tumor classification in MRI images using Fine tuned efficientnet”, e-Prime - Advances in Electrical Engineering, Electronics and Energy, Volume 8, 2024, 100498, ISSN 2772-6711, https://doi.org/10.1016/j.prime.2024.100498.

[17] Nayak, Dillip Ranjan; Padhy, Neelamadhab; Mallick, Pradeep Kumar; Zymbler, Mikhail; Kumar, Sachin, “Brain Tumor Classification Using Dense Efficient-Net” , Axioms (2075-1680), 2022, Vol 11, Issue 1, p34, ISSN 2075-1680, DOI 10.3390/axioms11010034.

[18] Pacal, I., Celik, O., Bayram, B. et al. “Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification”. Cluster Comput (2024). https://doi.org/10.1007/s10586-024-04532-1

[19] Rehman A, Naz S, Razzak MI, Akram F, Imran M. A deep learning-based framework for automatic brain tumors classifcation using transfer learning. Circuits Syst Signal Process. 2020;39(2):757–75.

[20] Kang, J.; Ullah, Z.; Gwak, J. MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers. Sensors 2021, 21, 2222. https://doi.org/10.3390/s21062222.

[21] Mohsen, Heba, et al. “Classification using deep learning neural networks for brain tumors”. Future Computing and Informatics Journal, Vol 3, pp 68-71, 2018. DOI: 10.1016/j.fcij.2017.12.001.

[22] Khawaldeh, Saed, et al. Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks. Applied Sciences, Vol 8,2017. DOI: 10.3390/app8010027.

[23] Mittal M, Goyal LM, Kaur S, Kaur I, Verma A, Jude HD. “Deep learning based enhanced tumor segmentation approach for MR brain images”. Appl Soft Comput. 2019;78:346–54.

[24] Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW. “Multi-grade brain tumor classification using deep CNN with extensive data augmentation”. Journal of computational science, Vol 30, pp174-82, 2019. DOI: 10. 1016/j.jocs.2018.12.003.

[25] Saeedi S, Rezayi S, Keshavarz H, R Niakan Kalhori S. “MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques”. BMC Med Inform Decis Mak. 2023 Jan 23;23(1):16. doi: 10.1186/s12911-023-02114-6. PMID: 36691030; PMCID: PMC9872362.

[26] https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

[27] Cheng, Jun (2017). brain tumor dataset. figshare. Dataset. https://doi.org/10.6084/m9.figshare.1512427.v5

[28] https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri

[29] Br35H :: Brain Tumor Detection 2020 Dataset, Open Source Dataset, author HASHIRA, https://universe.roboflow.com/hashira-fhxpj/br35h-::-brain-tumor-detection-2020

[30] Glenn Jocher and Ayush Chaurasia and Jing Qiu, “Ultralytics YOLOv8”, 2023, https://github.com/ultralytics/ultralytics

[31] M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10778-10787, doi: 10.1109/CVPR42600.2020.01079.

[32] R. Del Prete, M. D. Graziano and A. Renga, "RetinaNet: A deep learning architecture to achieve a robust wake detector in SAR images," 2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI), Naples, Italy, 2021, pp. 171-176, doi: 10.1109/RTSI50628.2021.9597297.

[33] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[34] Szegedy, Christian & Vanhoucke, Vincent & Ioffe, Sergey & Shlens, Jon & Wojna, ZB. (2016). “Rethinking the Inception Architecture for Computer Vision”. 0.1109/CVPR.2016.308.

[35] Huang, Gao & Liu, Zhuang & van der Maaten, Laurens & Weinberger, Kilian. (2017). "Densely Connected Convolutional Networks". 10.1109/CVPR.2017.243.

[36] Chollet, Francois. (2017). "Xception: Deep Learning with Depthwise Separable Convolutions". 1800-1807. 10.1109/CVPR.2017.195.

[37] Simonyan, Karen & Zisserman, Andrew. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". arXiv 1409.1556.

[38] Tan, Mingxing & Le, Quoc. (2019). "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks".

[39] Yu, Jiahui & Wang, Zirui & Vasudevan, Vijay & Yeung, Legg & Seyedhosseini, Mojtaba. (2022). "CoCa: Contrastive Captioners are Image-Text Foundation Models".

[40] CoCa model in OpenCLIP. https://colab.research.google.com/github/mlfoundations/open_clip/blob/master/docs/Interacting_with_open_coca.ipynb