

Investigating Deep Learning Models for the Classification of Brain Tumors to Infer Biological Significance from MRI Images

Xueshen Liu

Computer Science Courant, New York University, New York, 10012, U.S.

liurubben@gmail.com

Abstract. Brain tumors can pose a dangerous threat if misclassified. The primary way of classifying these brain tumors is through Magnetic Resonance Imaging (MRI) images, which is a highly subjective and manual process. With improved technology and machine learning, this paper proposes CapsNet model and ViT model, which differs from traditional CNN model, that aims to perform better on small image size, long-range dependencies and sequential features, which human tissue often contains. After extensive training and evaluation, the results of the CapsNet models and ViT models on the testing dataset come out to be 74.81% and 93.84%, respectively. We also find that the potential of our work will be enormous in propelling the actual application of ML in clinical diagnosis because the incredible accuracy will protect patients and doctors from the risk of misdiagnosis. Furthermore, interpreting the models also help us reinforce and gain new perspectives on clinical diagnostics.

Keywords: Brain Tumor classification, deep learning model, MRI images.

1. Introduction

According to the World Health Organization (WHO), cancer is the second most common cause of death worldwide [1]. Brain tumors can be extremely dangerous if not properly handled, and cause substantial mortality and morbidity. A tumor, unlike cancer, may be benign, pre-cancerous, or malignant. The former is non-cancerous and simply an abnormal growth in the brain, and the latter is a fast-growing cancer that can spread to other areas of the brain and spine. Unlike malignant tumors, benign ones can be surgically removed [2,3].

Brain tumors primarily are gliomas, meningiomas, and pituitary tumors. Meningiomas often have benign characteristics, whereas gliomas typically have malignant characteristics. Contrary to meningiomas, which are slow-growing tumors, pituitary tumors, even when benign, can nonetheless harm other parts of the body [4]. Due to the information provided above, it is crucial for the clinical diagnosis procedure and subsequent accurate patient assessment that these three types of tumors are precisely differentiated from one another.

To classify brain tumors, doctors often use Magnetic Resonance Imaging (MRI) images. However, brain tumor classification by hand is labor-intensive and error-prone. Moreover, a diagnostic is not fully complete without a biopsy - to determine if the tumor is benign or malignant - which is often not

collected until definitive brain surgery [5]. Therefore, it is necessary to discover a technique that can classify tumors with a high degree of accuracy.

The increase in technological innovation throughout the past few years, especially in machine learning (ML) and artificial intelligence (AI), has undoubtedly impacted the field of medical imaging. Numerous ML techniques for image classification have been tested for brain tumor classification.

Cheng et al.'s paper - which compiled a dataset used in our paper - used a mixture of feature extraction techniques such as intensity histograms, gray level co-occurrence matrix (GLCM), and bag-of-words (BoW) to classify brain tumors. This, in combination with image augmentation techniques such as image dilation and ring-form subregions, resulted in an accuracy of up to 91.28% [6].

Swati et al.'s model uses a pre-trained deep CNN model (VGG-19) and proposes a block-wise fine-tuning strategy based on transfer learning, and reaches an average accuracy of 94.82% under five-fold cross-validation [7], with a T1-weighted contrast-enhanced magnetic resonance images (CE-MRI) benchmark database [9].

Ardhendu et al.'s method retrieved the features of the brain MRI pictures through GoogLeNet, another CNN architecture. Following that, the characteristics are categorized using classifiers like K-Nearest Neighbor (K-NN), Softmax, and Support Vector Machine (SVM). Their results seem to be the best out of other brain tumor classifiers, as the accuracy of the SVM and k-NN was 98.93% and 98.63%, respectively [8].

However, state-of-the-art medical brain tumor image analysis papers often share the same flaws. Their datasets are often small, imbalanced, and too focused on glioma, the most likely to be malignant, i.e. more severe and life-threatening [13]. They also prefer a single type of ML network and aim to improve its accuracy, accepting little parallelling networks and cross-area networks, like ViT which is widely used in NLP [13]. This paper will not only expand the size and diversity of the dataset but also create several ML architectures, in order to outperform or provide extra ideas for traditional models.

2. Image Dataset

This paper utilizes a combination of 3 brain tumor datasets: Figshare, SARTAJ, and Br35H. The Figshare dataset contains 3064 T1-weighted contrast-enhanced images from 233 patients, 708 images displaying meningioma, 1426 images displaying glioma, and 930 images displaying pituitary[9]. The SARTAJ dataset contains a total of 3264 T1-weighted contrast-enhanced images, 926 images displaying glioma, 937 images displaying meningioma, 901 images displaying pituitary, and 500 images displaying no tumor[10]. The Br35H dataset contains 3060 images, 1500 images displaying brains with tumors, 1500 images without tumors, and 60 images for predictions [11]. Because the Br35H dataset doesn't classify the tumor images into glioma, meningioma, and pituitary, we'll only use the 1500 images of brains without tumors. Moreover, the SARTAJ images have some overlap with the Figshare dataset, leading to the removal of 806 images. In summary, the dataset this paper uses has a total of 7022 images, 1621 images displaying glioma, 1645 images displaying meningioma, 1757 images displaying pituitary, and 2000 images displaying no tumor [12].

3. Capsule Network (CapsNet) Model

3.1. Overview of the CapsNet (CapsNet) Model

CapsNet is a deep learning architecture proposed by Geoffrey Hinton and colleagues in 2017. It aims to overcome some limitations of traditional Convolutional Neural Networks (CNNs) and provide a more powerful representation capability. CapsNet introduces the concept of "capsules," replacing the conventional CNN's pooling layers, to better capture the spatial hierarchical structure of images [14]. By extracting features from 1 or 2 convolutional layers, the digit capsule layer is used to classify the image, which then gets interpreted by the fully connected layers as decoders. Figure 7 shows one common flowchart of the CapsNet Model.

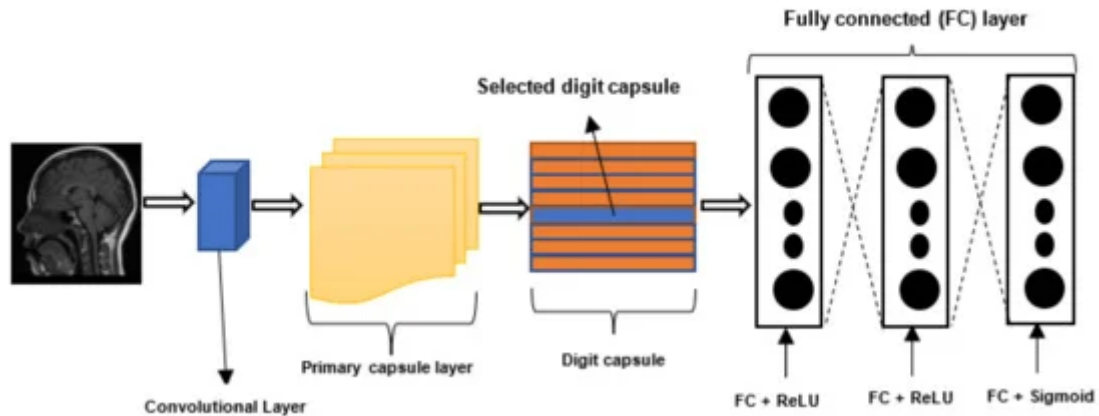


Figure 1. Overview of the CapsNet architecture

CapsNet has better resistance to simple, linear image transformations like rotation, as well as better accuracy in small and imbalanced datasets [13]. It also gives up the traditional pooling mechanism used by CNN models, aiming to capture all details in one image [13].

3.2. Image Dataset

The dataset for the CapsNet model is also not augmented. The main reason is the time limitation. As CapsNet never omits any details, its time complexity outweighs Feature Fusion Model and ViT Model significantly. While using a similar image size for the CapsNet Model, the average time for 1 epoch is 7000-8000 seconds with the University Computer Science Laboratory server, with no augmentation on the dataset. The slow learning rate of the CapsNet model also amplifies the time-consuming problem, making augmentation on the dataset less feasible. Finally, as suggested above, our data augmentation methods still belong to simple linear transformation, in which the accuracy of the model should make less difference compared to the model with no data augmentation.

We also decrease the image size using the resizing method in data preprocessing to reduce the time constraint. Dividing the image size by 2 roughly improves the time per epoch by 6-7 times, making a convergence of model accuracy more feasible.

3.3. Code

The code for the CapsNet model also gets revised. The CapsNet model was originally designed for the MNIST dataset, which typically uses a dataset of 28 by 28 images. For brain tumors, images with less than 100 by 100 make the brain tumor hardly visible while resizing a small image back also creates high pixelation, both making humans unable to read and distinguish tumors. We hence manually change the input size and adjust all other layers to fit the brain tumor's required size. With the final accuracy decrease compared with the original code, we still manage to achieve a feasible accuracy and hence the model.

4. Proposed Vision Transformer (ViT) Model

4.1. Overview of the Vision Transformation (ViT) Model

ViT is a model involving dividing an image into fixed-size patches, embedding each of them linearly, adding position embeddings, and feeding the resulting sequence of vectors to a standard Transformer encoder. To perform classification, ViT uses the standard approach of adding an extra learnable “classification token” to the sequence. In general, ViT is an architecture that uses self-attention mechanisms to process images. The architecture consists of a series of transformer blocks. Each transformer block consists of two sub-layers: a multi-head self-attention layer and a feed-forward layer. The method involves dividing an image into fixed-size patches, embedding each of them linearly, adding

position embeddings, and feeding the resulting sequence of vectors to a standard Transformer encoder. ViT models have recently emerged as a competitive alternative to convolutional neural networks [15]. Figure 6 is a diagram that illustrates the model.

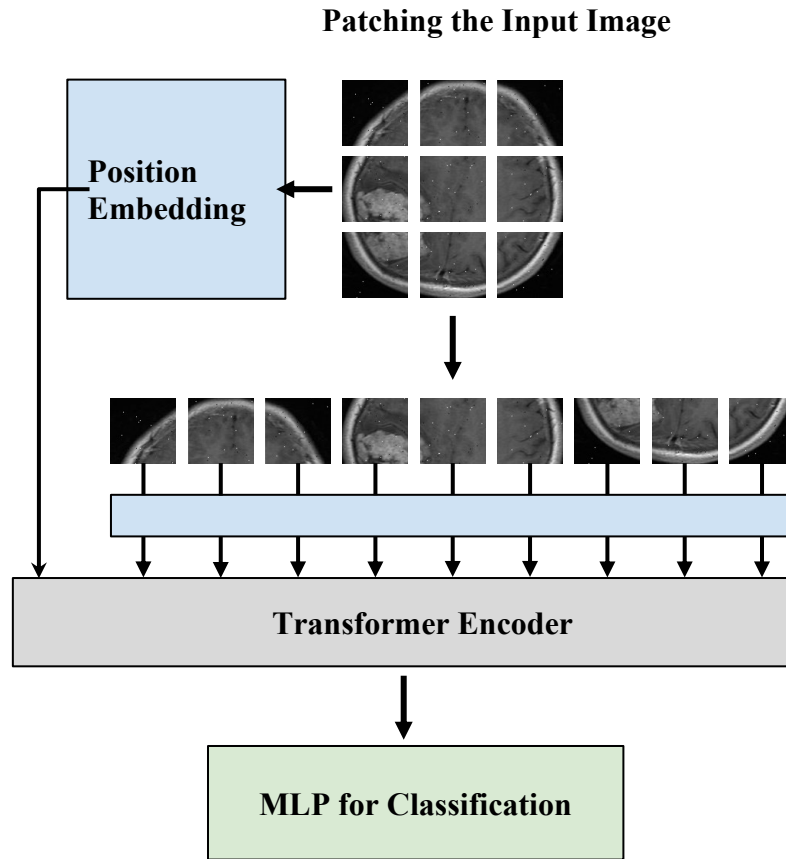


Figure 2. Overview of the ViT architecture

ViTs have become popular in medical imaging due to their ability to pick up sequential dependencies, which is extremely prevalent in human organs[16]; for example, the sequential way in which blood flows through our body - up the arteries, then down the veins.

4.2. Image Dataset

The dataset for the ViT model is not augmented. This is because of the sequential dependencies in human organs. In common cases, those sequential dependencies, such as blood flow, behave in one direction. Rotation and flipping make little sense here in training the model, as it is uncommon in real-world MRIs. We also tried to train the ViT model with the same data augmentation method and obtained approximately 55% accuracy throughout all epochs. We believe that this result confirms our assumption.

Attempting to apply data augmentation would confuse the model and decrease the precision of predicting each class. Hence, we only apply basic cropping, resizing, normalization, and filtering to the original dataset.

5. Results & Discussion

5.1. Training and Machine Specifications

The deep feature fusion model was trained for 8 epochs on the outlined augmented dataset with a batch size of 8. The vision transformer model was trained for 25 epochs on the pre-processed original data with a batch size of 16.

The training for Feature Fusion Model and ViT Model was conducted on an Apple Macbook Pro with the M2 Pro chip with a 10-core CPU and 16-core GPU. The Python libraries used for training were *tensorflow-macos* version 2.13.0, *tensorflow-metal* version 1.0.1, and *keras* version 2.13.1.

For CapsNet, as it requires higher computational power, we used the University of Wisconsin-Madison Computer Science Laboratory server to train and test it. The version of Python libraries remains the same as the other two models.

5.2. Results & Discussion

5.2.1. Main Evaluation and Inferences

After training, the models were evaluated on both the augmented and original databases to check for potential blind spots in the model. The results are below in Figures 3-4.

CapsNet Model	Training Dataset		Testing Dataset	
	Original	Augmented	Original	Augmented
Accuracy	79.11%	N/A	74.81%	N/A
Loss	0.1832	N/A	0.1947	N/A

Figure 3. Evaluation of the ViT model on the original versions of the training and testing datasets, from image size of 112*112

ViT Model	Training Dataset		Testing Dataset	
	Original	Augmented	Original	Augmented
Accuracy	99.97%	N/A	93.84%	N/A
Loss	0.0031	N/A	0.2507	N/A

Figure 4. Evaluation of the ViT model on the original versions of the training and testing datasets

From Figures 3-4, we made two key inferences. First, in terms of the ViT model, an accuracy of 93.84% is also accurate, convincing us it is suitable for brain tumor classification with MRI images. For the CapsNet model with only 74.81% accuracy, we still need time to validate the model, yet it is proved to be feasible. Second, high accuracy and low loss which are universal across the training and testing dataset evaluation indicate that our CapsNet and ViT models don't seem to be overfitted to the training dataset. This means that both two models are indeed learning from different MRI data, making them potent for future analysis.

5.2.2. Exploring Model Misclassifications

Despite the models achieving high accuracies, we still seek to identify potential weak spots within our model to note for future improvements and real-world implementations. Figure 5 compiles classification reports which identify a weakness regarding classifying meningioma.

ViT Model - Testing: Original Dataset				
Glioma	93.68%	89.00%	91.28%	300
Meningioma	87.42%	88.56%	87.99%	306
No Tumor	97.82%	99.51%	98.65%	405
Pituitary	96.38%	97.67%	97.02%	300
CapsNet Model - Testing: Original Dataset				
Glioma	72.11%	78.50%	75.17%	300
Meningioma	67.53%	51.78%	58.62%	253
No Tumor	87.78%	89.77%	88.76%	88
Pituitary	77.69%	94.00%	85.07%	300

Figure 5. Classification reports for both models on both the original and augmented versions of the training and testing datasets

Across the three classification reports, meningioma is always the worst in precision. The distance between it and the second least precise category is at a minimum of 2.75% and at a maximum of 15.75%. What this indicates is that the models often misclassify other brain tumor types to be meningioma. The confusion matrices of the model evaluations also indicate this, as shown in Figures 6-7.

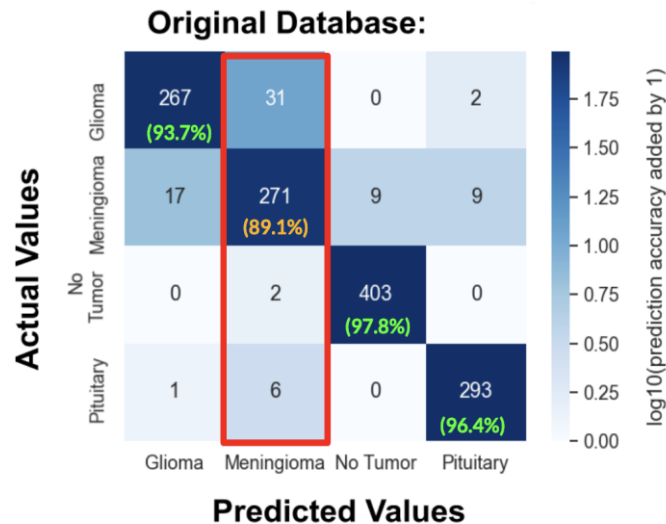


Figure 6. Confusion matrix of the ViT model on the original dataset

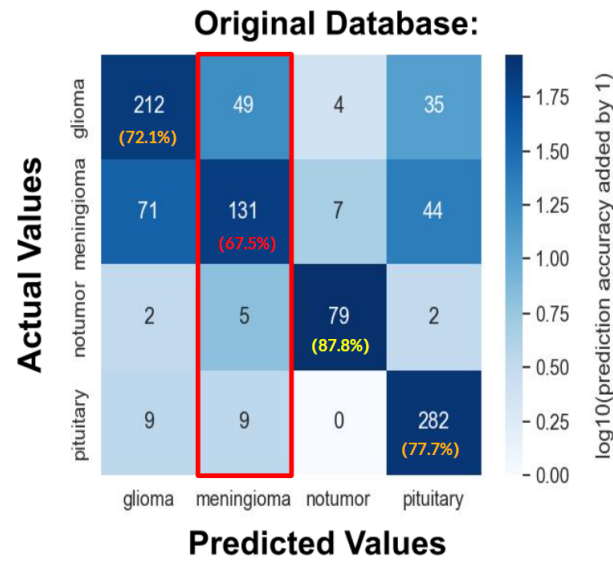


Figure 7. Confusion matrix of the CapsNet model on the original dataset

The red outlines in each confusion matrix represent the model-predicted meningioma values. In both confusion matrices, it is evident that both models misclassify meningioma more than they misclassify glioma, pituitary, or non-tumors.

5.3. Limitations and further improvements

5.3.1. Imbalanced Datasets

The number of images in the no tumor class in the dataset we used was significantly less than the other three classes. This is because the images in the no tumor category range from hundreds to thousands in width and tend to be skewed, making it difficult to identify the brain. Whereas the images in the other categories are uniform 512*512 shapes. After eliminating the images with large skew and small size, the preprocessed dataset is out of balance. We expect future papers to explore a uniform image size for no tumor dataset, in order to balance the image number for each class.

5.3.2. Time and structure limitation for CapsNet

Initially, we introduced the CapsNet model in order to solve the tumor position problem as well as the small and imbalanced dataset problem. As CapsNet is able to resist rotation, we assume it extracts fewer features from the space and position of the tumor and hope it focuses more on the size and appearance of the tumor in order to overcome such loci problems.

However, we discovered that the accuracy of CapsNet failed to achieve a satisfactory level: with smaller image sizes like 56*56, the model could achieve above 85% accuracy, while humans could hardly recognize the tumor. With larger image sizes, common people could distinguish the tumors while our CapsNet Model is still struggling at 80% accuracy.

We explain this result as two: The most important problem is the time limit. As CapsNet never drops any details in an image, and brain tumor images are typically over 100 by 100 pixels, it usually takes hours to compute 1 epoch and days for the model. This result in the figure is obtained from the dataset with image size 112 by 112, which the dataset image size already shrunk compared to Feature Fusion Model and ViT Model. We also tried to use image size 224 by 224 and obtained 73.68% accuracy in the training dataset and 70.24% in the testing dataset, which takes over 36000 seconds for pure computation and 3 days in total for the result, thus we will only discuss the result from image size of 112 by 112. Both results are obtained from 5 epochs also due to the time limit, while the original paper suggested 500 epochs to stabilize the accuracy of the model.

The other problem is the structure of CapsNet. We found that the CapsNet model often starts with 45% accuracy in the beginning, which we believe is a sign that the model is learning. When using images with sizes 56*56 and 112x112, both models demonstrated the expected output. However, when we tried image size 224x224, the accuracy of the first several epochs stayed at 18%-35%, with the loss remaining at 1.21. The confusion matrix at this point is mostly biased: the model chooses to classify all images to one class, and simply pays less attention or just ignores the other classes. It then usually takes 3-5 epochs to finally decrease its loss and improve its accuracy. Although models with all image sizes converge to approx. 80% accuracy, we think it is the signal of inefficient CapsNet structure. In other words, the original CapsNet used to treat smaller-size images may not be suitable for brain tumor classification with typically large image sizes. We hope future work could investigate this part deeply.

6. What Remains to Be Done

6.1. Human interpretation of brain tumors

While this paper used mathematical results to classify brain tumors, doctors and scientists often classify brain tumors according to their shapes and positions in the human brain. According to clinical results, glioma usually appears as an irregularly shaped lesion on MRI images, which may be scattered or relatively large masses; Meningioma usually appears as a well-defined and regular mass, which usually attached to the surface of the meninges, but may be found elsewhere; Pituitary tumor tends to be found in the pituitary gland, located at the base of the brain and is usually a well-demarcated round or oval mass, usually small. We would hence try to incorporate with doctors and scientists to evaluate if our model classification matches with human interpretation.

7. Conclusion

Overall, our team has made significant progress in developing a method for classifying brain tumors from MRI images. In the field of biology, our work has great potential in interpreting the method of a highly accurate model classifying brain tumors. Within these interpretations, we can exploit more information behind the physical structures of the brain. Furthermore, the research on the relevance between brain areas and brain functions can be propelled if we have a more comprehensive dataset with each patient's symptoms. In the field of CV & ML, our work has great potential as well. During the project, we used CapsNet and ViT model for brain tumor classification on MRI images. We extract the features from each model and compare their validation accuracy in any circumstance. It turns out that there are still questions to be resolved, for example why ViT has a worse performance in the augmented dataset than the original dataset which goes against the mainstream view. We believe that the answer - which is still unknown - will greatly propel the CV & ML research in the second it is answered.

Acknowledgement

We would acknowledge Professor Manolis Kellis and the TAs Ash and Zhu for providing original paper inspiration, data evaluation, and model suggestions for our paper.

References

- [1] World Health Organization—Cancer. Available online: <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed on 5 November 2019).
- [2] Priya, V.V. An Efficient Segmentation Approach for Brain Tumor Detection in MRI. *Indian J. Sci. Technol.* 2016, 9, 1–6. [Google Scholar]
- [3] Miller, KD, Ostrom, QT, Kruchko, C, Patil, N, Tihan, T, Cioffi, G, Fuchs, HE, Waite, KA, Jemal, A, Siegel, RL, Barnholtz-Sloan, JS. Brain and other central nervous system tumor statistics, 2021. *CA Cancer J Clin.* 2021. <https://doi.org/10.3322/caac.21693>
- [4] American Association of Neurological Surgeons—Classification of Brain Tumors. Available online: <https://www.aans.org/en/Media/Classifications-of-Brain-Tumors>

- [5] Byrne, J.; Dwivedi, R.; Minks, D. Tumors of the brain. In Nicholson T (ed) Recommendations Cross-Sectional Imaging Cancer Management, 2nd ed.; Royal College of Radiologists: London, UK, 2014; pp. 1–20. Available online: <https://www.rcr.ac.uk/publication/recommendations-cross-sectional-imaging-cancer-management-second-edition>
- [6] Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, et al. (2015) Correction: Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. PLoS ONE 10(12): e0144479. <https://doi.org/10.1371/journal.pone.0144479>
- [7] Swati, Z. N., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain Tumor Classification for MR images using transfer learning and fine-tuning. Computerized Medical Imaging and Graphics, 75, 34–46. <https://doi.org/10.1016/j.compmedimag.2019.05.001>
- [8] A. Sekhar, S. Biswas, R. Hazra, A. K. Sunaniya, A. Mukherjee and L. Yang, "Brain Tumor Classification Using Fine-Tuned GoogLeNet Features and Machine Learning Algorithms: IoMT Enabled CAD System, " in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 3, pp. 983-991, March 2022, doi: 10.1109/JBHI.2021.3100758.
- [9] Cheng, J. Brain Tumor Dataset. 2017. Available online: <https://doi.org/10.6084/m9.figshare.1512427.v5>
- [10] Kaggle Brain Tumor MRI Dataset. Available online: <https://doi.org/10.34740/KAGGLE/DSV/1183165>
- [11] Kaggle Brain Tumor MRI Dataset. Available online: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection?select=no>
- [12] Kaggle Brain Tumor MRI Dataset. Available online: <https://doi.org/10.34740/KAGGLE/DSV/2645886>
- [13] Akinyelu, A. A., Zaccagna, F., Grist, J. T., Castelli, M., & Rundo, L. (2022). Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey. Journal of Imaging, 8(8), 205. <https://doi.org/10.3390/jimaging8080205>
- [14] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. Advances in neural information processing systems, 30. <https://doi.org/10.48550/arXiv.1710.09829>
- [15] Zhao P, Li C, Rahaman MM, Xu H, Yang H, Sun H, Jiang T, Grzegorzec M. A Comparative Study of Deep Learning Classification Methods on a Small Environmental Microorganism Image Dataset (EMDS-6): From Convolutional Neural Networks to Visual Transformers. Front Microbiol. 2022 Mar 2; 13: 792166. doi: 10.3389/fmicb.2022.792166. PMID: 35308350; PMCID: PMC8924496.
- [16] Moiyadi AV, Tongaonkar HB, Bakshi GK. Symptomatic intracranial metastasis in penile carcinoma. Indian J Urol. 2010 Oct;26(4):585-6. doi: 10.4103/0970-1591.74471. PMID: 21369397; PMCID: PMC3034073.