Brain Tumor Severity Prediction Using MSENet

Abdul Afrid Mohammed
Data Science and Analytics Department
Georgia State University
Atlanta, GA 30324
amohammed34@student.gsu.edu

Preethi Reddy Matta
Data Science and Analytics Department
Georgia State University
Atlanta, GA 30324
pmatta1@student.gsu.edu

Abstract

Classification of brain tumors is one of the daunting tasks in medical imaging and incorrect decisions during the diagnosis process may lead to increased human fatality. Latest advances in artificial intelligence and deep learning have opened the path for the success of numerous medical image analysis tasks, including the recognition of brain tumors. In this article, we implemented a simple architecture based on deep learning that results in strong generalization without needing much preprocessing. The proposed Multi-modal Squeeze and Excitation model (MSENet) receives multiple representations of a given tumor image, learns end-to-end and effectively predicts the severity level of the tumor. Convolution feature descriptors from multiple deep pre-trained models are used to effectively describe the tumor images and are supplied as input to the MSENet. The squeeze and excitation blocks of the MSENet allow the model to prioritize tumor regions while giving less emphasis to the rest of the image, serving as an attention mechanism in the model. The model is evaluated on the benchmark brain tumor dataset that is publicly accessible from the Figshare repository. Experimental studies reveal that in terms of model parameters, the proposed approach is simple and leads to competitive performance. By increasing complexity, the model leads to generalization and achieves a state-of-the-art accuracy of 94.05% on the Figshare dataset. Compared to the existing models, the model neither uses segmentation nor augmentation techniques and without much pre-processing achieves competitive performance.

Keywords: Brain Tumor Classification (BTC); Pre-trained Convolution Neural Networks (CNNs); InceptionResNetV2, Xception, Squeeze and Excitation operations; Attention Mechanism.

1 Introduction

Tumors, abnormal masses of cells, pose a threat to human health, especially when located in the brain. Brain tumors are categorized as benign or malignant based on their impact. Benign tumors are harmless, while malignant tumors can spread rapidly, potentially leading to fatality if untreated. Malignant tumors include Meningioma, Glioma, and Pituitary tumor types. Manual diagnosis using Magnetic Resonance (MR) imaging scans is crucial but prone to misinterpretations, risking lives.

Evolution of Diagnostic Tools: Machine learning-based computer-aided diagnostic (CAD) tools have evolved to aid in brain tumor diagnosis. Early models utilized trivial features like wavelet features, progressing to multiple sets of features for improved prediction. Deep Convolutional Neural Networks (CNNs) surpassed traditional methods, demonstrating superior performance in image classification. Capsule Network architectures also showed promise but were less effective with non-segmented images.

Transfer Learning Approach: This work adopts a transfer learning paradigm, leveraging pre-trained deep CNN models like Xception and InceptionResnetV2 to obtain local feature descriptors from brain MR images. The proposed Multi-modal Squeeze and Excitation Network (MSENet) incorporates multiple squeeze and excitation (SE) blocks, acting as attention mechanisms to prioritize tumor regions. The model, requiring minimal preprocessing, accepts whole MR images and simultaneously learns from various interpretations.

Key Contributions: Implementation of a model learning from multiple local representations of brain MR images. Introduction of attention mechanisms through squeeze and excitation layers, prioritizing tumor regions. Training MSENet in an end-to-end manner with input from multiple channels.

2 Related Work

Recent advancements in pattern recognition and machine learning have led to automated brain tumor recognition from Magnetic Resonance (MR) images. This overview highlights the evolution of machine-learning models for brain tumor classification. Early approaches focused on intensity and texture-based features, while recent studies emphasize Fisher vector and Bag of Visual Word (BoW) representations for discriminating visual details. Convolutional Neural Networks (CNNs) have proven effective in medical image processing, outperforming conventional models. Deep learning models trained on augmented samples show superior performance, especially when applied to segmented tumor regions.

Transfer learning, using pre-trained CNNs like AlexNet and VGG Net, has become pivotal. Strategies involving fine-tuning and data augmentation contribute to accurate brain tumor severity prediction. Inspired by recent advancements, our model leverages multiple feature representations from various deep CNNs. The composite neural network, incorporating squeeze and excitation blocks, focuses on essential image regions, particularly tumor areas, aiming to enhance overall performance.

3 Methodology

The main objective of this work is to develop a simple and robust architecture based on Deep Neural Networks (DNN) for the classification of brain tumors. In this article, we implemented a Multi-modal Squeeze-and-Excitation network (MSENet), designed to handle and learn from two distinct representations of brain tumor images, leading to more generalized predictions. The model takes vectorized representations of tumor images from two different deep pre-trained models. To emphasize descriptors defining tumor regions, the network incorporates squeeze and excitation (SE) blocks, functioning similarly to attention mechanisms. This ensures that the model learns to focus on each local descriptor appropriately. The architectural details of MSENet are illustrated in Figure 2. This section provides a comprehensive overview of the model.

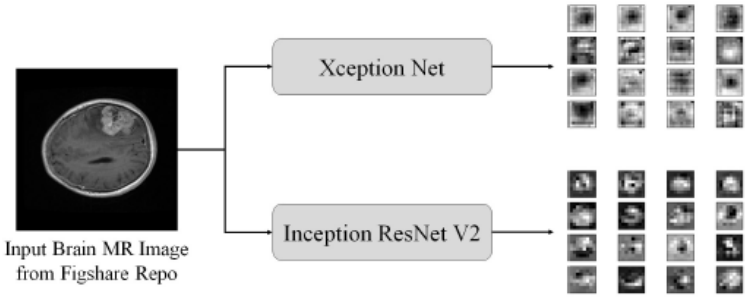

Figure 1: Local Feature representations of brain MR image from figshare repository extracted using Xception and InceptionResNetV2 pre-trained deep networks.

4 Multi-modal Feature Preparation

In recent years, it has become very common to obtain image feature representations from the fully connected layers of the deep pre-trained networks. As the pre-trained models are trained using the ImageNet dataset and the images of this dataset are completely different from those of the medical images, these deep features may not be appropriate for representing medical images. Moreover, the neurons present in the final fully connected layers of deep pre-trained models learn the discriminating information, and the classes in our application are entirely different from those of ImageNET. Addressing these issues, we suggest using the feature representations from convolution layers of deep CNN networks to acquire local representations of brain MR images. It is noted, based on the literature, that multimodal representation is often complementary in improving the model's power of discrimination. Therefore, we suggest using two distinct pre-trained deep CNNs, XceptionNet and InceptionResNetV2 (IRV2), for local feature extraction in our work. These two networks proved to be efficient in terms of performance and entered leaderboards for ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Without loss of generality, features from any two distinct deep CNNs can be used as input to the MSENet. The multi-modal local representations of the MR images provide precise descriptions, if any, of the tumor present and help in discriminating the type of tumor more accurately. Figure 2 shows the visualization of the convolution feature maps on passing one of the figshare images through Xception and IRV2 networks. From each deep CNN model, K sets of 2-D local feature descriptors, f = f1; f2; ...; fk, are collected. Each local descriptor fi is of dimension (W x H). The model obtains one such representation from the Xception network and another from the IRV2 network and offers multi-modal feature representations of tumor images as input to the model.

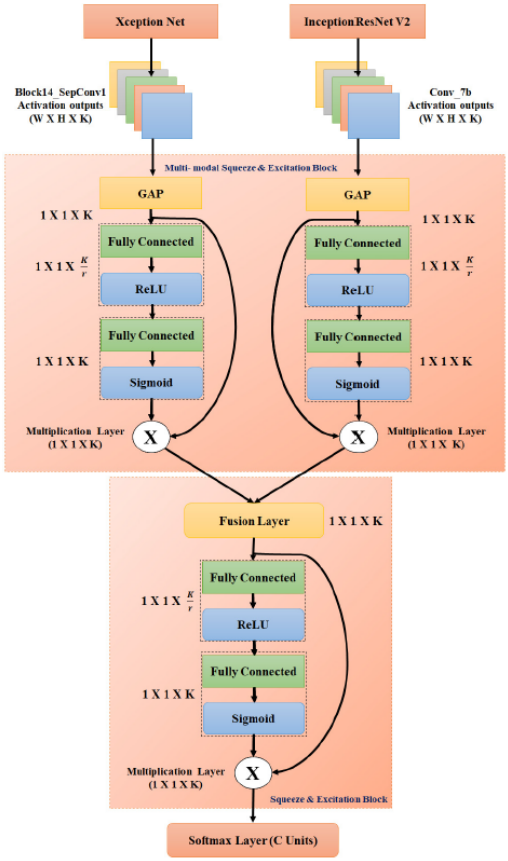

Figure 2: Architecture of Multi-modal Squeeze and Excitation Network for Brain Tumor Severity Detection

5 Squeeze and Excitation Blocks for Attention

To address the limitation of standard neural networks treating all input descriptors equally, especially in medical image processing, we introduce a Squeeze and Excitation (SE) block. This block assigns varying degrees of focus to different local descriptors of tumor images, enhancing the prioritization of features representing tumor regions in the classification process.

The SE block takes K sets of 2-D local feature descriptors, denoted as $f = f_1, f_2, ..., f_k$, as input and produces a K-dimensional vector $\hat{f} = \hat{f}_1, \hat{f}_2, ..., \hat{f}_k$. Each input local descriptor f_i is of dimension $W \times H$, and the output descriptor $\hat{f}_i$ is a scalar. The block applies a global average pooling (GAP) operation on each 2-D input feature map f_k, squeezing global spatial information into a single scalar descriptor. Formally, each $\hat{f}_k$ is computed as:

$$\widehat{f_k} = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} x_{ij}, \quad where x_{ij} \in f_k$$

To calibrate feature descriptors for varying attention levels, attention scores are computed using a two-layer feedforward network. This network acts as a gating mechanism, with ReLU non-linearity at the first layer and sigmoid non-linearity at the second. Attention scores a_i for the i-th local feature set $\hat{f}_i$ are obtained as:

$$a_i = \sigma(W_2 ReLU(W_1 \widehat{f_i})), \quad where \sigma is the sigmoid function$$

Finally, the SE block outputs the re-scaled $\hat{f}_i$ with attention weights using a multiplication layer:

$$z_i = a_i \cdot \hat{f}_i$$

Here, $a_i = [a_{i1}, a_{i2}, ..., a_{iK}]$ represents the attention scores, and a_{iK} is the attention weight associated with the k-th feature descriptor $\hat{f}_{iK}$. The excitation followed by multiplication layers acts as self-attention on feature maps, providing a weighted attention-based representation.

6 Multi-modal Squeeze and Excitation Network for Brain Tumor Classification

Our goal is to improve the representational power of a network, and hence we have considered two distinct sets of local feature maps from two different pre-trained CNNs. We implemented a deep neural network that can accept these two sets of local feature descriptors and learn from both the input representations of tumor images independently in parallel. Each of these feature sets is passed through a separate SE block independently, and then the attention scores are computed. Each SE block produces weighted feature descriptors such that the descriptors representing tumor regions are given higher attention over the rest. Now, these two weighted feature descriptors are merged using a merge layer to get a single consolidated representation out of the multi-modal input representations.

This merged representation is now passed through another SE block to ensure that the network learns to selectively emphasize informative features while suppressing less useful features. The final softmax layer of the model serves as a classification layer and produces the type of tumor present in the input MR images as output. The model with multiple channels can be trained jointly in an end-to-end manner.

The multi-modal representations provide complementary details of the tumor images, whereas the attention mechanism allows giving appropriate attention to each of the regions of the brain MR images. If properly trained, the MSENet learns attention scores such that the feature descriptors representing tumor regions get high focus, while the descriptors representing non-tumor regions get reduced focus. Thus, the MSENet with multi-modal features as input and attention mechanism provided in the SE block contributes to the improved performance of the classification model.

Algorithm 1: Proposed Algorithm that uses MSENet for Brain Tumor Severity Prediction

Input: Let D be the dataset of Contrast Enhanced-MRI brain images from Figshare repository, where $D = \{(X_i, y_i)\}_{i=1}^{3062}$ and y_i is the severity level of brain tumor associated with X_i

Output: $y_i \in \{0, 1, 2\}$ representing Meningioma, Glioma and Pituitary brain tumors.

Step-1:*Pre-Processing:* Each image X_i in the dataset D is pre-processed such that they are compatible with deep pre-trained models.

Step-2: *Feature Extraction:*

Each pre-processed image is passed through pre-trained Xception and IRV2 models to get two sets of 2-D feature representations as follows:

$F_1 \leftarrow$ Features from the first depth-wise separable convolution layer of block-14 of Xception network.

$F_2 \leftarrow$ Features from the first convolution layer of the seventh block of IRV2

Where $F_1 \in \mathcal{R}^{\{16\times16\times1536\}}$ and $F_2 \in \mathcal{R}^{\{14\times14\times1536\}}$

Now, each $x_i \in D$ has two high-dimensional representations F_1 and F_2.

Step-3: Preparing the dataset, D, for five-fold cross validation:

Split the dataset D into five partitions, $D = \{D_1, D_2, D_3, D_4, D_5\}$.

Step-4: Model training and evaluation:

Consider one of the split D_i for testing and remaining splits of the dataset $D - \{D_i\}$ for training.

Supply both the high dimensional feature representations F_1 and F_2 of the training data to the proposed MSENet and train the model. Get tumor severity predictions for all the samples, x_i in D_i.

Compute all the required performance measures by comparing the model predictions with ground truth labels

Step-5: Five-fold cross-validation approach for evaluating model performance:

Repeat step-3 for $i \in \{1, 2, 3, 4, 5\}$ and obtain the model predictions for the split D_i.

Average performance obtained for all the five splits is considered as the model performance

Figure 3: Proposed Algorithm

Note:

It's noteworthy to mention that in the algorithm's original design, we recommended a 5-fold split for training. However, owing to computational constraints, each model for every split consuming over 10 GB, we pragmatically adopted an 80-20 split. It's crucial to underscore that all results and methodology detailed in this context stem from the 80-20 split. Looking ahead, our future endeavors involve implementing the initially proposed 5-fold split, leveraging enhanced computational resources.

7 Experimental Studies

7.1 Dataset Summary

A small-scale brain tumor dataset is introduced by Jun Cheng 13 and is made available in the Figshare repository. This dataset includes 3064 images of 233 patients distributed across one of three distinct (axial, coronal, sagittal) planes. Each image in the dataset represents one of the brain

tumor (Meningioma, Glioma, and pituitary) types. Table 4 shows statistics of the dataset along with Class-wise distribution. Figure 5 shows samples of Contrast Enhanced-MRI (CE-MRI) brain images of the Figshare brain tumor dataset representing three different types of tumors in three different views.

Tumor Type	# Patients	MRI view	# Images	Total
Meningioma	82	Axial	209	708
		Coronal	268	
		Sagittal	231	
Glioma	89	Axial	494	1426
		Coronal	437	
		Sagittal	495	
Pituitary	62	Axial	291	766
		Coronal	319	
		Sagittal	320	
Total	**233**	-	**3064**	**3064**

Figure 4: Dataset Summary

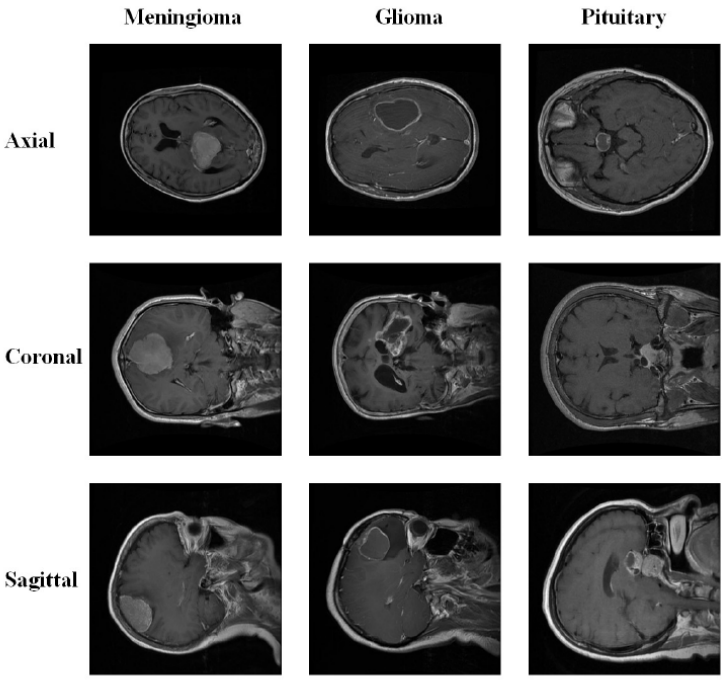

Figure 5: Samples of CE-MRI images

7.2 Model Hyper parameters

This section provides the configuration and various hyperparameters of the proposed MSE Network. The model takes 1536 (K) feature maps of 14x14 each and another 1536 (K) feature maps of dimensions 16x16 each. A gated attention pooling layer transforms the 2-D feature representations into a vectorized form. Each SE block in the proposed MSENet is designed with a ReLU layer of Kr neurons followed by a softmax layer of K neurons. For our experimental studies, different r values are considered from the set of 8, 16, 32. For the fusion layer in the MSENet architecture, we considered both average fusion and simple concatenation for our experiments. The Softmax layer at the end of the model is designed to have 3 neurons as the dataset has MR images representing one

of the three different types of tumors. Table 6 summarizes the additional hyperparameters used in designing the proposed MSENet architecture. The cumulative number of trainable parameters of the proposed MSENet with concatenation operation in the fusion layer is counted at 3.57 million, and the model has 1.79 million parameters with the average operation in the fusion layer.

Hyper parameter	Setting
Loss Function	categorical cross-entropy
Optimizer	ADAM
Initial Learning Rate	0.001
Batch Size	32
Number of Epochs	100
Regularization	$L_2(0.0001)$

Figure 6: Model Hyper parameters

7.3 MSENet for tumor classification Results

From the previous series of experiments, we conclude that hybrid features significantly enhance model performance compared to feature descriptors obtained from single deep CNNs for representing brain tumor images. The subsequent set of experiments aims to demonstrate the effectiveness of the proposed model in brain tumor classification tasks. In earlier experiments, either a single deep representation or a composite representation was employed to represent brain MR images. Learning from both these features is expected to yield more information from tumor images, as opposed to simply concatenating features from various versions.

To facilitate this, the proposed Multi-modal Squeeze-and-Excitation Network (MSENet) is implemented, enabling the learning from multiple deep feature representations and acquiring tumor details through multiple paths. Additionally, the proposed model allows for appropriate attention allocation to each tumor descriptor. MSENet utilizes two sets of tumor image descriptors derived using the process described in Section 3. The next two experiments involve MSENet, differing only in the fusion operation: the former experiment employs averaging, while the latter experiment utilizes concatenation.

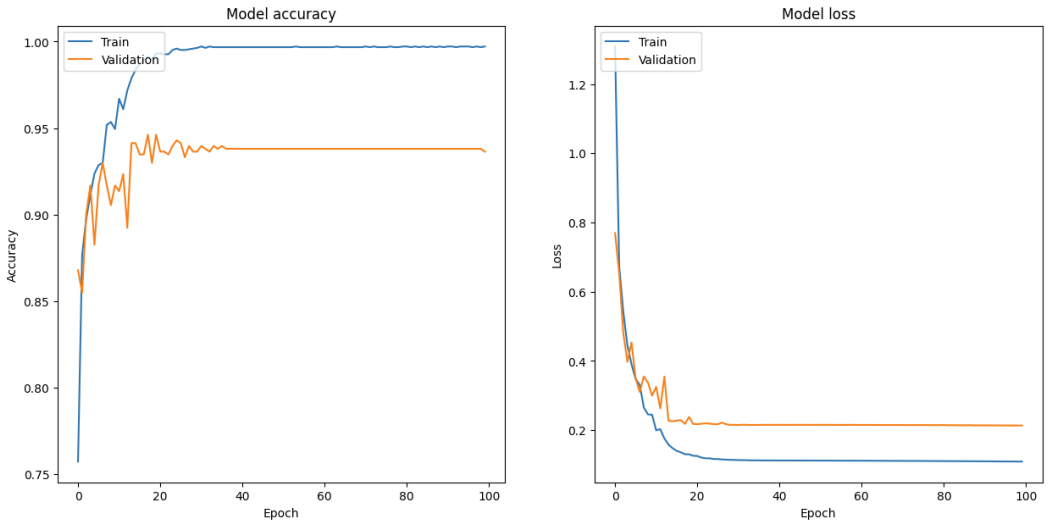

Figure 7: The trend of training and validation accuracy and loss over successive epochs.

The neural network was trained for 100 epochs on a dataset, and its performance was monitored during each epoch. The training process involved updating the model's weights to minimize the loss function. The training accuracy steadily increased from 75.72% in the first epoch to 99.71% in later epochs. The validation accuracy improved from 86.79% to 94.62%, indicating that the model generalized well to unseen data. The model's performance was saved during training based on the highest validation accuracy achieved. The learning rate was adjusted dynamically during training using techniques like ReduceLROnPlateau to fine-tune the model. Overall, the neural network demonstrated a robust learning process, achieving high accuracy on both training and validation sets, with a focus on generalization to new data.

7.4 Evaluation results

The evaluation metrics provide a comprehensive understanding of the performance of our brain tumor classification model. Precision, which measures the accuracy of positive predictions, is noteworthy across all tumor types. For Meningioma, the model demonstrated an 88.57% precision, indicating the reliability of its predictions for this class. Glioma and Pituitary classifications exhibited similarly high precision at 87.32% and 87.94%, respectively. Recall, capturing the model's ability to identify all relevant instances, was exceptional. Notably, Glioma recall reached 96.49%, showcasing the model's effectiveness in recognizing instances of this tumor type. These high recall values were complemented by strong F1-Scores, emphasizing a harmonious balance between precision and recall.

The overall accuracy of 94.62% indicates the model's proficiency in correctly classifying brain tumors. Macro average metrics, representing averages across all classes, also reflected consistent high performance, with macro precision, macro recall, and macro F1-Score at 94.11%, 93.77%, and 93.03%, respectively. Weighted averages, considering the class distribution, confirmed the robustness of the model, with weighted precision, recall, and F1-Score at 94.58%, 94.62%, and 94.59%, respectively. The Quadratic Kappa Score of 91.18 further underscores the model's ability to make predictions beyond random chance, indicating a strong agreement between predicted and actual classes. These comprehensive metrics collectively affirm the success of our brain tumor classification project, highlighting the model's accuracy and reliability in identifying different tumor types.

```
Accuracy:  94.62
Macro Precision:  94.11
Macro Recall:  93.77
Macro F1-Score:  93.93
Quadratic Kappa Score:  91.18
```

Figure 8: Evaluation results of the model

	precision	recall	f1-score	support
Meningima	0.8986	0.8732	0.8857	142
Glioma	0.9514	0.9614	0.9564	285
Pituitary	0.9733	0.9785	0.9759	186
accuracy			0.9462	613
macro avg	0.9411	0.9377	0.9393	613
weighted avg	0.9458	0.9462	0.9459	613

Figure 9: Class wise evaluation results

8 Conclusion and Future Scope

The primary goal of this study is to present a straightforward yet robust deep learning methodology for assessing the magnitude of brain tumors. We introduce a Multi-modal Squeeze and Excitation Network (MSENet) designed to accept multiple representations of tumor images, learning from

each representation by focusing ample attention on the tumor regions within MR images. The incorporated squeeze and excitation blocks operate as a gated mechanism, functioning as an attention mechanism. This allows the model to prioritize tumor regions, while de-emphasizing non-tumor areas. Our method is assessed using the Figshare dataset, and extensive experiments reveal that the proposed model, employing an averaging operation for fusion, achieves competitive performance with existing models, utilizing significantly fewer parameters. Additionally, when concatenation operation is used for fusion, our model surpasses existing ones, setting a new state-of-the-art performance on the Figshare dataset.

Given the limited computational resources, it's crucial to acknowledge that the reported findings are based on a small-scale dataset. Extensive evaluations on large-scale datasets are imperative for robust conclusions. This work utilizes Xception and InceptionResNetV2 networks for representing brain tumor images; however, for efficiency, exploring other deep networks is a potential avenue. Furthermore, considering Granular computing techniques and adopting common preprocessing methods for unbalanced datasets could be valuable extensions to this research.

9 Contributions

Task	People
1. Collecting and preprocessing data	Preethi Reddy Matta
2. Research and Implementing Pre-trained Xception Model	Preethi Reddy Matta
3. Research and Implementing Pre-Trained Inception Model	Abdul Afrid Mohammed
4. Implementing Final and Combined Model	Abdul Afrid Mohammed
5. Evaluating Metrics	Preethi and Afrid
6. Slides, demo, and Presentation	Preethi and Afrid
7. Writing report	Preethi and Afrid

Table 1: Task Assignment

10 References

[1] Seetha, J., and S. Selvakumar Raja. "Brain tumor classification using convolutional neural networks." Biomedical & Pharmacology Journal, vol. 11, no. 3, 2018, pp. 1457.

[2] Zikic, Darko, et al. "Segmentation of brain tumor tissues with convolutional neural networks." Proceedings MICCAI-BRATS, 2014, pp. 36-39.

[3] Afshar, Parnian, Arash Mohammadi, and Konstantinos N. Plataniotis. "Brain tumor type classification via capsule networks." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

[4] Chen, Sihong, Kai Ma, and Yefeng Zheng. "Med3d: Transfer learning for 3d medical image analysis." arXiv preprint arXiv:1904.00625, 2019.

[5] Amin, Javeria, et al. "A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning." Journal of Medical Systems, vol. 43, no. 11, 2019, pp. 326.

[6] Brain MRI Dataset. (2021, June 15). Figshare. `https://figshare.com/articles/dataset/Brain_MRI_Dataset/14778750`

[7] Sarkar, A. (2023, May 19). Xception: Implementing from scratch using Tensorflow. Medium. https://towardsdatascience.com/xception-from-scratch-using-tensorflow-even-better-than-inception-940fb231ced9

[8] Raj, B. (2020, July 31). A Simple Guide to the Versions of the Inception Network. Medium. https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

[9] Lipková, J., Chen, R. J., Chen, B., Lu, M., Barbieri, M., Shao, D., Vaidya, A., Chen, C., Zhuang, L., Williamson, D. F. K., Shaban, M., Chen, T., & Mahmood, F. (2022, October 1). Artificial intelligence for multimodal data integration in oncology. Cancer Cell. https://doi.org/10.1016/j.ccell.2022.09.012

[10] Srivastava, T. (2023, July 22). 12 Important Model Evaluation Metrics for Machine Learning Everyone Should Know (Updated 2023). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/

[11] Gad, A. (n.d.). Evaluating Deep Learning Models: The Confusion Matrix, Accuracy, Precision, and Recall - KDnuggets. KDnuggets. https://www.kdnuggets.com/2021/02/evaluating-deep-learning-models-confusion-matrix-accuracy-precision-recall.html

[12] Introduction to Convolution Neural Network. (2023, November 2). GeeksforGeeks. https://www.geeksforgeeks.org/introduction-convolution-neural-network/

[13] Ramadoss, V. (2022, November 1). Squeeze-and-Excitation Networks - Vinoth Ramadoss - Medium. Medium. https://medium.com/@Vinoth-Ramadoss/squeeze-and-excitation-networks-84e3db0e04e2

[14] Xie Y, Zaccagna F, Rundo L, Testa C, Agati R, Lodi R, Manners DN, Tonon C. Convolutional Neural Network Techniques for Brain Tumor Classification (from 2015 to 2022): Review, Challenges, and Future Perspectives. Diagnostics (Basel). 2022 Jul 31;12(8):1850. doi: 10.3390/diagnostics12081850. PMID: 36010200; PMCID: PMC9406354.

[15] Deep Neural Networks. (n.d.). https://www.tutorialspoint.com/python_deep_learning/python_deep_learning_deep_neural_networks.htm