HIGH SCHOOL EDITION
Journal of Student Research

Volume 12 Issue 4 (2023)

Diversified AI Techniques for Augmenting Brain Tumor Diagnosis

Dhruv Mandalik[1] and Odysseas Drosis[#]

[1]Ridge High School, USA
[#]Advisor

ABSTRACT

Brain tumors affect thousands of people worldwide each year and can be extremely fatal if not diagnosed early. They are challenging to diagnose due to their complexity and the overlapping features of different tumor types. This research explores the application of AI technology to expedite the diagnosis of brain tumors. The proposed AI-based approaches involve using deep learning algorithms to analyze medical imaging data, specifically MRI scans. The goal was to build a robust and accurate model that could overcome distribution shifts. Some of the models used include classical machine learning models and a convolutional neural network. The results demonstrate that AI-based approaches can significantly improve the accuracy and expedite the process of brain tumor diagnosis. The performances of the models were evaluated by using cross-validation and measuring accuracy, using a completely different dataset of MRI scans, to assess how the models performed when dealing with distribution shifts. The logistic regression model achieved a testing accuracy of 78.56%. The multi-layer perceptron (MLP) model achieved a testing accuracy of 74.89%. The multiplicative weight update method combined two models (MLP and logistic regression) with dynamically adjusted weights and achieved a testing accuracy of 83.83%. An approach where multiple aggregating models were used collaboratively achieved a testing accuracy of 86.46%. The best performing model, a convolutional neural network (CNN), yielded a testing accuracy of 98.20%.

Introduction

An intracranial tumor, known as brain tumor, is a lethal illness that affects hundreds of thousands of people worldwide. Brain tumors are the rapid growth of abnormal cells in any part of the brain, and there are over 150 variations of these tumors, depending on where they originate. Like all other types of tumors, they can be classified as benign (noncancerous) or malignant (cancerous). Brain tumors can be further classified as primary (tumors arising from brain cells) and metastatic (tumors that spread to the brain from other places in the body).

Imaging techniques, such as magnetic resonance imaging (MRI) scans and computed tomography (CT) scans, are noninvasive and harmless methods commonly employed to capture images of the brain and diagnose brain tumors. Medical professionals tend to favor the use of MRI scans due to the detailed and high-resolution images they provide. The detection and diagnosis of brain tumors is complex as the symptoms may be vague and similar to those of other medical conditions.

Due to the similarity in appearance when viewed under a microscope, variations of brain and spinal cord tumors can be misdiagnosed. Even with skilled pathologists analyzing tissue samples, it is estimated that up to 10% of people with a brain or spinal cord tumor initially receive an incorrect diagnosis. An incorrect diagnosis can have an impact on the outcome of treatment, as tumors that appear similar at a cellular level may require vastly different treatment approaches. Accurate diagnosis is crucial in determining the treatment options. The application of AI technology to analyze large amounts of MRI images can aid in identifying underlying patterns and features that brain tumors share and can lead to more accurate diagnosis and appropriate treatment options.

Previous Works

Various methods have been previously used to diagnose these tumors, where machine learning imaging models, such as CNNs, were used to classify MRI scans.

A study conducted by Saeedi et al. previously utilized a modified 2D CNN 2*2 kernel functions, consisting of eight convolutional and four pooling layers with batch-normalization. It was trained for 100 epochs with a batch size of 16 images. They also modified a convolutional auto-encoder neural network, which consisted of layers with different filter lengths and 2*2 kernel functions for each convolutional layer. They used a dataset of 3264 MRI scans, achieving a testing accuracy of 95% for the 2D CNN and 96% for the auto-encoder network.

Another study conducted by Badža and Barjaktarovic´ proposed a new, simpler CNN architecture designed for distinguishing three types of brain tumors using 3064 MRI scans. The CNN architecture included an input layer, two main blocks, a classification block, and an output layer. 10-fold cross validation was implemented to evaluate the CNN, and it yielded an accuracy of 96.56%.

Aamir et al. proposed a method of using deep learning models to extract features from MRI scans in order to create a hybrid feature vector. Agglomerative clustering, a method of hierarchical clustering to group objects based on similarity, is then implemented to identify where tumors most commonly arise. They then feed the proposals to the head network for classification.

Methods

Datasets Utilized

Three datasets of brain MRI images were obtained from Kaggle. The first dataset consisted of 253 scans that were pre-labeled as either 0 (no brain tumor) or 1 (brain tumor).

The second dataset comprised of 5,712 images and was categorized into four different labels: 'glioma tumors', 'meningioma tumors', 'pituitary tumors', and 'no tumors'. The images portraying tumors were grouped into a broader category of images displaying tumors and images not displaying tumors. This dataset presented a challenge due to the inclusion of MRI scans from varying planes. Certain images were acquired from the axial plane, depicting the brain from top to bottom; others were captured from the coronal plane, representing the brain from front to back, or from the sagittal plane, portraying the brain from left to right. There were 4,117 scans that showed brain tumors while the remaining 1,595 did not have tumors. Because of the drastic imbalance in data, 3,000 scans in this dataset were used.

The third and final dataset utilized consisted of 3,000 images, evenly split into images that did have tumors and those that did not. Each scan in this dataset was one taken from the axial plane. All the images in this dataset would be used to evaluate the model.

The use of three varying datasets was implemented to create a robust model that could generalize to a variety of real-world scenarios. By training and testing models on a range of datasets that have different distributions, it becomes possible to overcome distribution shift, an issue which arises when a model performs poorly when presented with data that differs from the data it was trained on. Distribution shift is a common issue in medical image analysis, as it is challenging to acquire a large and diverse set of images, leading to the potential for bias in a model trained on a limited dataset. Testing on contrasting datasets allows realistic testing of the model's ability to generalize to previously unseen images, which is crucial in real-world applications.

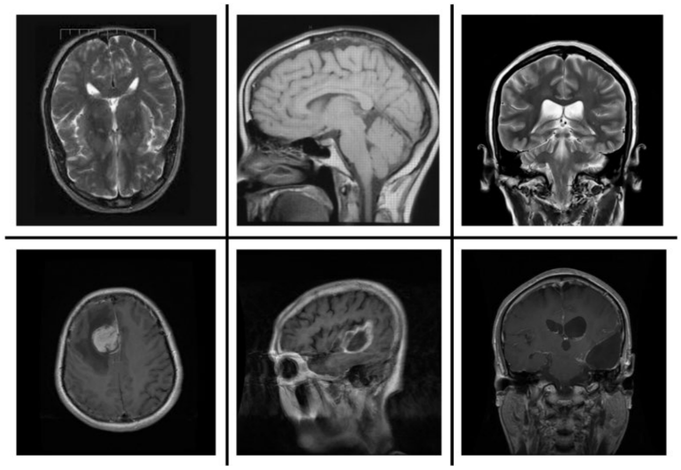


Figure 1. Sample raw brain MRI scans

Data Processing

Each raw image was a different size, and to prepare the images for analysis, an algorithm was used to resize each image to (150 x 150 x 3) pixels. The relatively smaller size of the scans ensured that the models would not become overly complex and overfit when trained on the images. Furthermore, each pixel value of each of the datasets was scaled down to a float number between 0 and 1. The images and their corresponding labels were then organized into lists and split to form training, testing, and validations sets.

Table 1. Distribution of data (images) among training, testing, validation, and unseen sets

Training	Testing	Validation	Unseen	Total
2,605	326	326	3,000	6,257

Logistic Regression (LR) Model

A preliminary logistic regression model was employed as an initial approach to address the issue at hand. Logistic regression models predict the probability of a point of data belonging to a specific category and are commonly used for binary classification. Hyperparameters such as the *maximum number of iterations*, (max_iter), and the *penalt*y value were adjusted in order to enhance the accuracy of the model. The L2 regularization method was implemented to avoid overfitting in the training phase. The L2 penalty penalizes the higher weights more harshly, generally resulting in weights that aren't fitted to a specific dataset. The maximum number of iterations was set to 50, to ensure that the model would not overfit to the training data, but would have sufficient iterations to establish a discernible pattern.

Multi-Layer Perceptron Model (MLP)

A multi-layer perceptron (MLP) classifier model, obtained from the sci-kit learn library, was applied to the dataset. MLP models are a type of simple neural networks that consist of input, output, and hidden layers. The *maximum number of iterations* hyperparameter, (max_iter), as well as the *hidden_layer_sizes* parameter, were fine-tuned to optimize the model's performance.

Multiplicative Weight Update (MWU)

An additional approach was pursued where two aggregating models were utilized together. Two models, (an MLP model and a logistic regression model), were trained on the training dataset, and their weights were initially set to 0.5. The training accuracies of the models were 97.19% and 98.85%. Both models were evaluated using the training dataset, and for each wrong classification, the error value of each model increased by 0.9^the number of iterations that had already passed. This approach ensured that recent mistakes would have a higher weight than previous ones. The weights of the models were divided by their respective errors, and at the end of each iteration, the weights were added together and stored in a variable named 'sum.' To ensure that the sum of the weights would be equal to 1, each weight was divided by the sum. After this process was repeated for the entire training dataset and the weights for each model were established, the method was evaluated on the validation and unseen datasets.

Boosting (BST)

Boosting is a machine learning technique that combines multiple weak models to create a strong model with improved accuracy. Three different logistic regression models were used for this method. The initial model was not always sufficient in predicting correctly on each MRI scan, so a subsequent model was developed specifically to make accurate predictions for the scans the initial model had predicted incorrectly. The initial model, L1, was trained on the training dataset, and its accuracy was evaluated. A second dataset was then compiled, consisting of all images and labels that L1 classified incorrectly. The second model, L2, was then trained on this dataset. Next, a third dataset was created, consisting of all images in the training dataset, but with a new set of labels. These labels were either 0 (for MRI scans that L1 correctly predicted) or 1 (for all other MRI scans). A third model, LMag, was trained on the third dataset. The purpose of the LMag model was to discern differences between the images that L1 classified correctly and incorrectly on. The *class_weight* hyperparameter was fine-tuned to a value of {1:15} in the training of LMag, as an imbalance in the number of MRI scans that were labeled 0 and 1 existed in the third dataset. The models were then tested on the testing dataset. LMag would first make predictions on which model to use based on the testing dataset's images. Based on this decision, either L1 or L2 would classify the image as having a tumor or not having a tumor.

TensorFlow CNN Model

To add onto the complexity of the models and try to overcome distribution shift, an approach using an alternative machine learning library, Keras in TensorFlow was used. A convolutional neural network (CNN) was created using the TensorFlow library. Convolutional neural networks are a subset of neural networks that are primarily used to process images. They contain three main types of layers, convolutional layers, pooling layers, and fully connected layers. Convolutional layers take in an image, represented as a 3D matrix, and use small filters to discern features in the image and generate a feature map. Pooling layers reduce the input's dimensionality by applying a filter that selects a few values to populate the output array. Pooling layers simplify and improve the efficiency of CNNs while also reducing the risk of overfitting. The fully connected layer connects every node from the previous layer to the output layer and then performs classification. The CNN created consisted of several layers: four convolution layers, four pooling layers (each with a 2x2 kernel), and a final dense layer. Each convolution layer was compiled with 'relu' activation, while the final dense layer was compiled with 'sigmoid' activation. The model was compiled with 'binary_crossentropy' as its loss function, and the optimizer was 'adam'. It was trained for 50 epochs.

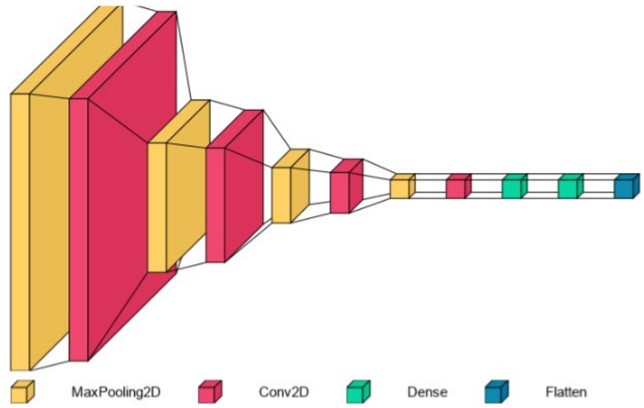


Figure 2. CNN Model Architecture

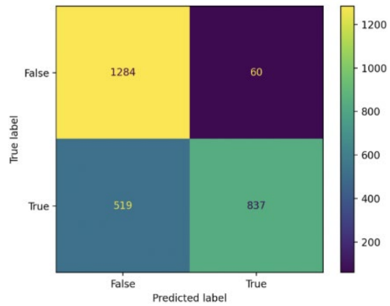
Results

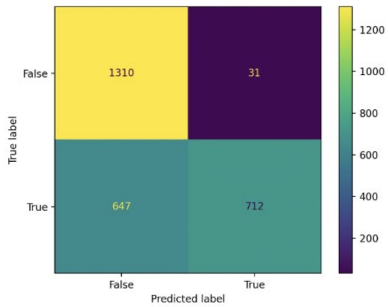
Table 2. Training accuracy, validation accuracy, and unseen dataset accuracy for each method

	Training Accuracy (%)	Validation Accuracy (%)	Unseen Data Accuracy (%)
LR	99.70	87.76	78.56
MLP	98.09	87.16	74.89
MWU	99.89	97.00	83.83
BST	98.46	93.33	86.46
CNN	95.76	98.38	98.20

Figure 4. MLP Confusion Matrix

Figure 5. MWU Confusion Ma-





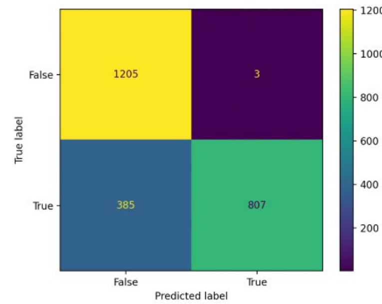


Figure 3. LR Confusion Matrix

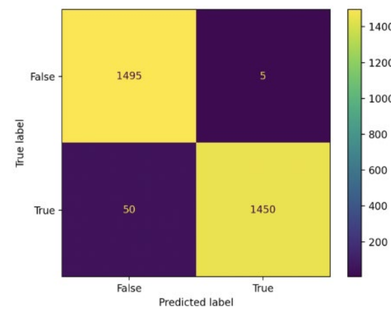


Figure 6. BST L1 Confusion Matrix

Figure 7. BST L2 Confusion Matrix

Figure 8. CNN Confusion Matrix

The logistic regression model had a training accuracy of 99.70% and a validation accuracy of 87.76%. It had an accuracy of 78.56% on the unseen dataset, demonstrating that it was not able to proficiently classify images from alternate datasets. Figure 3 displays the confusion matrix for the model. The model had high specificity and reasonable sensitivity. The model produced 519 false negatives but only 60 false positives.

The MLP model was able to achieve a validation accuracy of 87.16% and an accuracy of 74.89% on the unseen dataset. The 23% difference between the training accuracy and the accuracy on the unseen dataset is a clear indicator that the model was overfitting. Like the logistic regression model, it was unable to overcome distribution shifts. Figure 4 shows the model's confusion matrix. The MLP's false positive rate was 23.97%, and its false negative rate was 1.15%. The model performed very similarly to the logistic regression model and had high specificity but a lower sensitivity.

The multiplicative weight update method yielded high training and validation accuracies and satisfactory accuracy on the unseen dataset. The confusion matrix for this approach is detailed in Figure 5. This approach had high specificity but a lower sensitivity as well.

The boosting method had relatively good accuracies for all three stages. The confusion matrixes for the L1 and L2 models are displayed in Figures 6 and 7. The first logistic regression model, L1, had strong sensitivity and specificity, producing only 1 false positive and 265 false negatives. Its accuracy was 88.03%. The LMag model, whose training accuracy was 100%, predicted 178 images as ones that L2 should predict on. The L2 model was much less accurate than L1, and yielded an accuracy of 66.85%, which dropped the overall accuracy of this approach to 86.46%. Collectively, this approach produced only 35 false positives and 290 false negatives.

The negligible difference between the validation accuracy of the CNN and its accuracy on unseen data demonstrates that it was the best of the models and was able to generalize to subtly different scans. Figure 8 depicts the confusion matrix of the model. The CNN model made false positive predictions for only 0.17% of the cases and false negative predictions for only 1.7% of the MRI scans. It had both high specificity and high sensitivity.

Conclusion

Future work in the application of machine learning to expedite the diagnosis of brain tumors could involve exploring additional data sources beyond MRI scans, such as positron emission tomography (PET) and computed tomography (CT) scans. Additionally, incorporating numerical data from electronic health records and medical histories could provide additional insights into patient-specific risk factors for developing brain tumors and help to refine diagnostic algorithms. Numerical data, such as the size, shape, and location of the suspected tumor could be utilized to improve the accuracy of diagnosing tumors. Extending this research could also include tuning hyperparameters of the models as a means of increasing accuracy in the predictions.

Further investigation could also involve exploring the use of transfer learning and other techniques for addressing distribution shift. As AI-based diagnostic models are deployed in real-world clinical settings, they will encounter new data distributions and variations that may impact their accuracy and reliability. Developing approaches for adapting models to new distributions and settings could help to ensure that AI-based diagnostic tools remain effective and useful over time.

Acknowledgments

I would like to thank my mentor, Mr. Odysseas Drosis, for his guidance throughout this project. I would also like to thank the Inspirit AI Scholars Program for their support and resources.

References

Aamir, M., Rahman, Z., Dayo, Z. A., Abro, W. A., Uddin, I., Khan, I., Imran, A. S., Ishfaq, M., Guon, Y., & Hu, Z. (2022, May 26). A deep learning approach for brain tumor classification using MRI images. https://www.sciencedirect.com/science/article/abs/pii/S0045790622003603

Badža, M. M., & Barjaktarović, M. (2020, March 15). *Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network*. MDPI. https://www.mdpi.com/2076-3417/10/6/1999

Brain Tumors. AANS. (n.d.). https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors

Cancer.Net. (2022, October). Brain Tumor: Statistics. https://www.cancer.net/cancer-types/brain-tumor/statistics

Chakrabarty, N. (2019, April 14). *Brain MRI images for Brain tumor detection*. Kaggle. https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection

National Cancer Institute. (2022, July 12). Advances in Brain and Spinal Cord Tumor Research. https://www.cancer.gov/types/brain/research#diagnosis-of-brain-and-spinal-cord-tumorsnbsp

Nickparvar, M. (2021, September 24). *Brain tumor MRI dataset*. Kaggle. https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

Saeedi, S., Rezayi, S., Keshavarz, H. *et al.* MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Med Inform Decis Mak* 23, 16 (2023). https://doi.org/10.1186/s12911-023-02114-6

Sartaj, S. (2020, May 24). *Brain tumor classification (MRI)*. Kaggle. https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri