MRI Classification Project P08

Oliver Fowler
North Carolina State University
Email: orfowler@ncsu.edu

Shubham Mankar
North Carolina State University
Email: smankar@ncsu.edu

Jonathan Wood
North Carolina State University
Email: jwood9@ncsu.edu

Jesse Wood
North Carolina State University
Email: jlwood23@ncsu.edu

1 Background

A tumor in the brain is a life threatening condition. Glioblastoma is a common form of brain cancer in adults, with a median overall survival period of 7.7 months [1]. Among patients with brain tumors, a biomarker named MGMT methylation has been shown to be an indicator of favorable prognosis and response to chemotherapy [1]. However, conclusive methylation detection requires surgical removal of a tissue sample and may take several weeks [1]. As a result, classification models using magnetic resonance imaging (MRI) scans have been developed as a less invasive alternative for identifying patients with the marker [2]. However, there are various types of MRI scans that contain information vital to detect the presence of MGMT promoter methylation. In this project, we aim to develop classification models using concepts of machine learning and neural networks to classify the presence or absence of methylation among cancer patients. We further compare the performance of our models by varying the type of MRI scans used. In doing so, we will identify which image type is most suitable for the task.

2 Literature Review

Several studies have show that the information encoded in MRI data may be useful for the prediction of a medical status. For instance, Kanas et. al [2] used a semi-automated process to obtain measurements for quantitative features, such as edema and tumor volume, and achieve a 73.6% accuracy when predicting the presence of MGMT promoter methylation using K-nearest neighbors. Another study achieved 62% accuracy for predicting MGMT promoter methylation when applying a convolutional recurrent neural network to 3-dimensional MRI data for patients with glioblastoma, indicating that spatial features from the images may complement other patient data for classifying the presence of the biomarker [3]. Additionally, 2014 research [4] has explored how T1w brain MRI scans can be used to individually diagnose patients with Parkinson's disease (PD) and Progressive Supranuclear Palsy (PSP). In the research, Principal Component Analysis (PCA) and Support Vector Machine (SVM) concepts were used to train a model that could correctly classify a given input with accuracy, specificity and sensitivity of over 90%.

3 Method

We used Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Convolutional Neural Networks - Long Short-Term Memory (CNN - LSTM), and Residual Neural Networks (ResNet) to develop four machine learning models. Using different machine learning models on the MRI scans enabled us to examine the effect of model complexity on classification performance while comparing the effect of using each type of MRI scan as input into the model. Each model is a binary classifier that predicts the presence or absence of MGMT promoter methylation in the patient based

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

on the MRI scan. Further, hyper-parameters for the models were selected using grid search with k-fold cross validation on the training dataset. Following hyper-parameter selection, the performance of each model was evaluated using the testing dataset. Since both false positives and false negatives for methylation are expensive due to the implications to patients, we will use accuracy and F1-score to measure the performance of our models.

4 Experiment Setup

4.1 Understanding the data

The project uses the RSNA-MICCAI Brain Tumor Radiogenomic Classification dataset [5]. The attributes available in the dataset are – an Identification Number (BraTS21ID) for each patient and multiple MRI scans in DICOM format. Each patient has four types of multi-parametric MRI (mpMRI) scans. The type of mpMRI scans present in the dataset are Fluid Attenuated Inversion Recovery (FLAIR), T1-weighted pre-contrast (T1w), T1-weighted post-contrast (T1wCE), and T2-weighted (T2) scans. Using the given attribute values, we built a model to predict the MGMT Value, which is 1 if tumor is present and 0 otherwise.

The dataset contains data for a total of 585 patients. However, because of some unexpected issues, the cases corresponding to three patients (BraTS21ID equal to 00109, 00123, 00709) are inaccurate. Therefore, only data of the remaining 582 patients is considered for training, validation and testing. The available data is split into two cohorts — one for training and validation, and other for testing. We use 85% of the data for training and validation, and the remaining 15% for testing. Further information about the data split is as shown in the Table 1.

Table 1: Dataset split

	Positive	Negative
Training and CV Cohort	260	234
Testing Cohort	46	42

The RSNA-MICCAI Brain Tumor Radiogenomic Classification dataset has a few inconsistencies that may incorrectly induce bias in our model. An example of such an inconsistency is seen in the number of DICOM format images present for the mpMRI scans for each case. During the data exploration phase, we discovered that data was not evenly distributed, some patients' FLAIR MRI only had 15 images, while other patients had as high as 514 FLAIR MRI images. Similar patterns were seen in the T1w, T1wCE and T2w MRI scans. Some more inconsistencies were observed in the dimensions and aspect ratios of the DICOM images. Finally, some example mpMRI scans with different MGMT values are shown in Figure 1.

4.2 Research Question

Given the vastness and complexity of the available dataset and the limited availability of resources, it was vital to clearly structure and define the research questions that we intended to investigate. Finding the effectiveness of each MRI scan type would guide future research and help burgeon our understanding on the detection of MGMT methylation. Therefore, in this experiment, we aim at investigating the performance of each image type along with comparing the performance of the mentioned machine learning models on the given image data.

4.3 Data Pre-Processing

Due to the inconsistencies observed in the number of images available per patient, aspect ratios of the images, and slice thickness used when the MRI images were taken, we have implemented image pre-processing steps to standardize the dimensions of the input across all patients. Our pre-processing includes re-sampling the images for each patient so that the slice thickness is at 2 mm, only retaining the middle 60 images available (all images with zero padding if less than 60 are available). After re-sampling, we further transform the dimensions of each image to either 64 x 64 (SVM) or 192 x

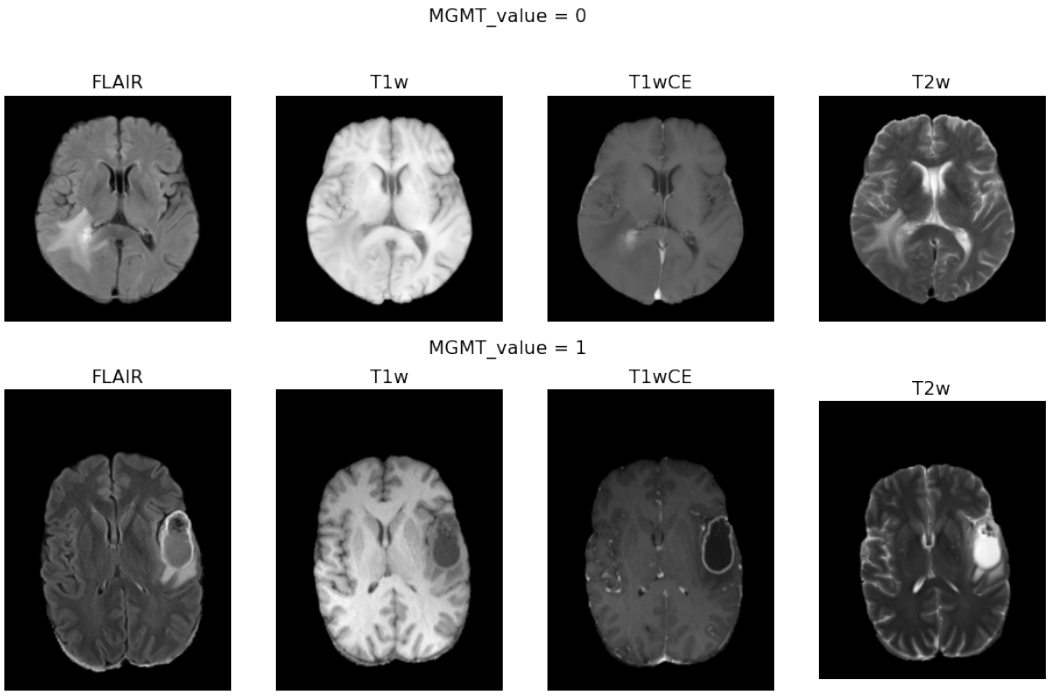

Figure 1: Example mpMRI scans

192 (all other models). In case of our Support Vector Machine (SVM) Model, we further combined the 64 images by averaging the pixel values across all images to create a representative image which was then transformed into a one dimensional array.

4.4 Model Development

This stage goes over the design and implementation of the four machine learning models that were developed and used to compare the aforementioned MRI scans.

1. **Support Vector Machine (SVM):**

   Support vector machine is effective at learning linear and nonlinear decision boundaries in attribute space to separate classes [6]. We chose to implement a Support Vector Machine model because of its unique ability to innately regularize its learning and not suffer from overfitting, as discussed in [6]. Due to resource constraints, a slice of 64 input images were first combined to form a representative image which was then flattened. Images for all patients created a 2-dimensional input space. We fed this input to the initial SVM model build using Radial basis function and default regularization parameter values as a starting point. The model architecture is shown in Figure 2.

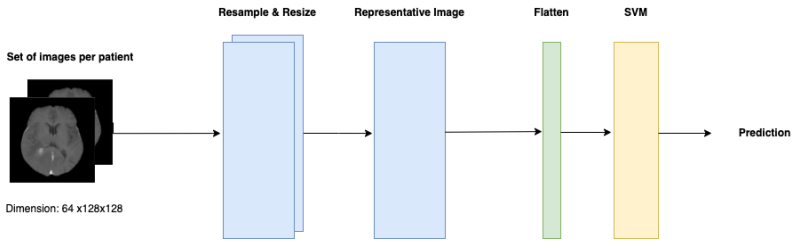

Figure 2: Architecture for SVM Model

2. **Convolutional Neural Network:**

The Convolutional Neural Network (CNN) has proven capable of extracting spatial features for images that can be used for classification [7]. As a result, we developed a simple CNN architecture with one convolutional layer and and two fully-connected layers to extract features from the images and predict the MGMT methylation status. The model architecture is shown in Figure 3.

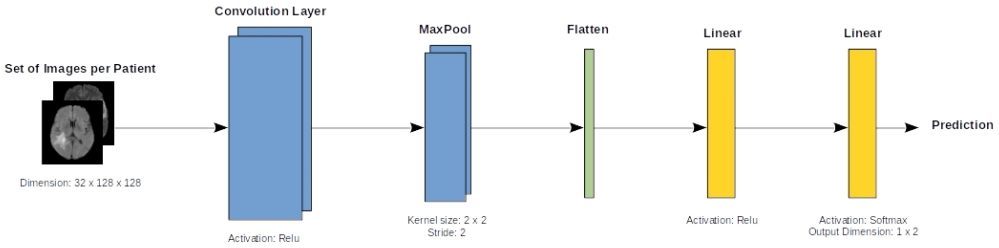

Figure 3: Architecture for CNN Model

3. **Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM):**

Another technique, in which the model is comprised of a CNN followed by a Long Short-Term Memory (LSTM) network, was found to outperform more advanced models (i.e. AlexNet and ResNet) when used to classify 3D brain tumor MRIs of patients with glioma (a common type of brain tumor) based on the type of glioma (High Grade or Low Grade) [8]. Therefore, we created another model for our task by adding a bi-directional LSTM network in between the CNN and fully-connected layers of our CNN model. Since the LSTM is a recurrent network, it's addition may help extract features that consider differences between images for the same patient. The CNN-LSTM model architecture is shown in Figure 4.

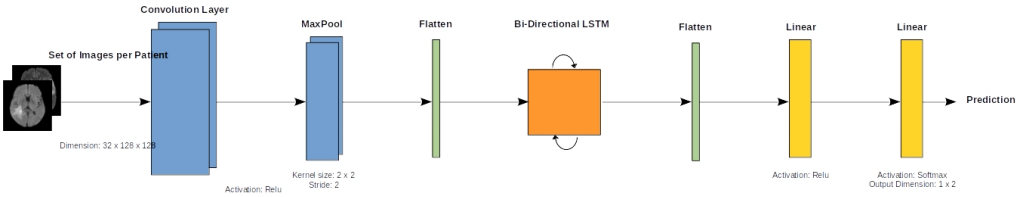

Figure 4: Architecture for CNN-LSTM Model

4. **Residual Neural Network:**

Many state of the art models in image classification are based on the Residual Neural Network (ResNet). ResNet is a deep convolutional neural network (many convolutional layers stacked together) that also has connections that skip layers to address an issue with deep networks, called degradation, in which accuracy peaks and begins declining rapidly [9]. We include a ResNet model to assess the performance of a more advanced approach for image classification on our task. The model architecture for ResNet is shown in Figure 5. Note that the diagram is for ResNet-34 while we used ResNet-101, with the only difference being the number of convolutional layers and skip connections.

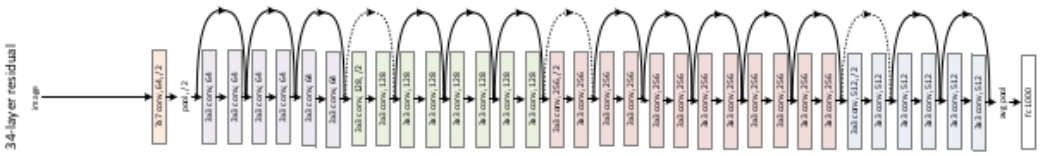

Figure 5: Architecture for ResNet-34 Model [7]

4.5 Hyper-parameter Tuning and Selection

For our SVM, CNN and CNN-LSTM models, we defined a range of hyperparameters based on manual exploration of performance. The optimal values selected Using this as a starting point, we used grid search with 3-fold cross validation to find the optimal combination of hyperparameters for the models. We selected the hyperparameter values with the best 3-fold cross-validation accuracy using the training and cross-validation cohort data. This method of hyperparameter selection was implemented for every type of MRI scan.

The hyperparameters for SVM were kernel type (linear, polynomial, RBF, or sigmoid), C value (0.1, 0.2, 0.3, 1, 5, 10, 20, 100, 200, or 1000), degree (1, 2, 3, 4, 5), coef0 (0.0001, 0.001, 0.002, 0.01, 0.02, 0.1, 0.2, 0.3, 1, 2, 5, or 10), and gamma (0.0001, 0.002, 0.01, 0.02, 0.03, 0.1, 0.2, 1, 2, or 3). The optimal hyperparameters for each image type according to the grid search for best 3-fold cross-validation accuracy across these parameters for SVM are listed in Table 2. For the CNN model,

Table 2: Selected Hyperparameter Values for SVM.

	Input Image Type			
Hyperparameter	FLAIR	T1w	T1wCE	T2
Kernel Type	Polynomial	RBF	Polynomial	Linear
C	200	0.3	100	0.3
degree	1	-	4	-
coef0	0.0001	-	0.0001	-
gamma	-	0.002	-	-

we used stochastic gradient descent for training with a learning rate of 0.0005, a batch size of 50, and the number of training epochs set to 50. We then optimized the hyperparameter values for CNN filter size (1, 15, or 30), CNN kernel size (4, 8, 12, or 16), and the number of hidden units in the first fully connected layer (100, 200, 300, or 400). The final number of training epochs was based on the number of training epochs required to achieve the best 3-fold cross-validation accuracy for the model with optimal hyperparameter values specified. The optimal hyperparameters for each image type on the CNN model are listed in Table 3.

Table 3: Selected Hyperparameter Values for CNN Model.

	Input Image Type			
Hyperparameter	FLAIR	T1w	T1wCE	T2
CNN Filters	1	1	30	1
CNN Kernel Size	12	12	12	16
Hidden Units for Linear Layer 1	300	300	300	100

For the CNN-LSTM model, we again used stochastic gradient descent with a learning rate of 0.0005, a batch size of 50, and the number of training epochs set to 50. We then optimized the hyperparameter values for CNN filter size (1 or 50), CNN kernel size (4, 8, or 16), the number of hidden units in the LSTM (400, 600, or 800), and the number of hidden units in the first fully connected layer (200, 300, or 400). The final number of training epochs was based on the number of training epochs required to achieve the best 3-fold cross-validation accuracy for the model with optimal hyperparameter values specified. The optimal hyperparameters with each image type as

input for the CNN-LSTM model are listed in Table 4.

Table 4: Selected Hyperparameter Values for CNN-LSTM Model.

	Input Image Type			
Hyperparameter	FLAIR	T1w	T1wCE	T2
CNN Filters	50	1	50	1
CNN Kernel Size	4	4	16	8
Hidden Units for LSTM	600	600	600	400
Hidden Units for Linear Layer 1	200	300	300	400

For the ResNet model, we fine-tuned the pre-trained ResNet-101 model by adding a linear layer with 2 output units and softmax activation and training the resulting model with the training and cross-validation data for 50 epochs.

5 Results

Overall, the MRI images demonstrated poor predictive power for the MGMT promoter methylation. Interestingly, the SVM model out-performed the more complex neural network models in our study, regardless of the type of input image. In fact, the highest overall performance (59.83% accuracy) was achieved by the SVM model with FLAIR images as input. The final results for each model and input image type are shown in Table 5.

Table 5: MRI Classification Model Evaluation. Accuracy is listed first with F1 score in parenthesis.

	Model			
	FLAIR	T1w	T1wCE	T2
SVM	**59.83 (0.6803)**	**56.54 (0.6946)**	**56.41 (0.6792)**	**57.26 (0.5763)**
CNN	55.68 (0.5568)	47.73 (0.4773)	52.27 (0.5227)	48.86 (0.4886)
CNN-LSTM	53.41 (0.5341)	54.55 (0.5455)	52.27 (0.5227)	51.14 (0.5114)
ResNet	51.21 (0.4444)	52.63 (0.4565)	50.40 (0.4411)	52.63 (0.4565)

6 Conclusion

Unfortunately, the MRI images for patients with brain tumors did not prove to be individually capable predictors of MGMT promoter methylation. However, using the FLAIR images to predict the biomarker status did outperform the other images, on average. Thus, continued experimentation using FLAIR images or combining features from all or combinations of the types of images may prove beneficial. Additionally, we observed that the SVM model performed better than the more complex neural networks. We suspect that this may be due to the limited computing resources and amount of time required to tune the neural networks for optimal performance. As such, the results demonstrate the advantage of simpler machine learning algorithms, especially when resources are constrained.

GitHub Link: `https://github.ncsu.edu/jwood9/CSC_522_Project`

References

[1] G. Reifenberger et al., "Predictive impact of MGMT promoter methylation in glioblastoma of the elderly," Int. J. Cancer, vol. 131, no. 6, pp. 1342–1350, 2012, doi: 10.1002/ijc.27385.

[2] V. G. Kanas, E. I. Zacharaki, G. A. Thomas, P. O. Zinn, V. Megalooikonomou, and R. R. Colen, "Learning MRI-based classification models for MGMT methylation status prediction in glibbest," Comput. Methods Programs Biomed., vol. 140, pp. 249–257, Mar. 2017, doi: 10.1016/j.cmpb.2016.12.018.

[3] Han L, Kamdar MR. "MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks, " Pac Symp Biocomput., vol. 23, pp. 331-342, 2018.

[4] Salvatore, Christian, et al. "Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy." Journal of neuroscience methods, vol. 222, pp. 230-237, 2014.

[5] U. Baid et al., "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification," ArXiv210702314 Cs, Jul. 2021, Accessed: Sep. 08, 2021. [Online]. Available: http://arxiv.org/abs/2107.02314

[6] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. 2018. Introduction to Data Mining (2nd Edition) (2nd. ed.). Pearson.

[7] Pi. Wang, E. Fan, and Pe. Wang. "Compararative analysis of image classification algorithms based on traditional machine learning and deep learning," Pattern Recognition Letters, vol. 141, pp. 61-67, 2021, doi: 10.1016/j.patrec.2020.07.042.

[8] I. Shahzadi, T. B. Tang, F. Meriadeau and A. Quyyum, "CNN-LSTM: Cascaded Framework For Brain Tumour Classification," 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2018, pp. 633-637, doi: 10.1109/IECBES.2018.8626704.

[9] He, Kaiming, et al., "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.