MRI Tumor Detection & Classification in the Brain using Machine Learning

Simeon Kolev May 23rd, 2023

INTRODUCTION

Abstract

Brain tumor detection and classification using MRI imaging is an important area of study in medical analyses. With an estimated yearly hospitalization of 25,000 adults and 5,000 children in the US, it is of upmost importance for medical professionals to prescribe accurate and efficient treatments. Recently, the conjuncture of high-resolution 3D brain imaging with advanced machine learning techniques has significantly aided medical professionals in this field. MRIs and ML image models allow doctors to provide more accurate diagnoses and more personalized treatments depending on the size, shape, and location of these tumorous growths. This integration of statistics and medicine is a large and still growing area of research. Image classification techniques imbedded in machine learning have proven to aid in the treatment and management of malignant growths.

Professional research has made significant contributions to the development of machine learning algorithms aimed at improving the accuracy and efficiency of brain tumor diagnosis and classification. The performance of existing pretrained neural networks have been found to reach a 99% classification accuracy in replicable medical trials. Without the medical experience and computation power at the disposal of clinical research, replication of these results seemed infeasible in a college setting.

However, with demanding MRI selection and various image processing techniques we were able to construct a database that satisfied medical standards for our area of research. This allowed for more accurate exploratory data analysis and subsequently provided a stronger foundation for the implementation of advanced machine learning techniques to accurately detect and classify tumorous growths in the brain. With countless hours of research and a few mathematical tricks our elementary ML models were able to reach upwards of 85.83% classification accuracy. Additionally, through the construction of a 6 layers convolutional neural network, we were able to successfully approximate clinical accuracy while reaching a peak model performance at 95.79% classification accuracy.

About the Tumors

Brain tumors are potentially malignant abnormal growths in various parts of the brain debilitating a multitude of key functions within the nervous system. Even benign tumors can become problematic as they grow and put increased pressure on vital parts of the brain.

There are over 120 different subclasses of brain tumors discovered. Due to limitations in access and quantity of data, our research aims to focus on the detection and classification of just 3. Though each tumor subclass has their own unique characteristics in terms of tumor size, shape, and location, they share similarities that allow them to be grouped into 3 main categories: glioma, meningioma, and pituitary.

Glioma tumors develop in the glial cells affecting the support and insulation of neurons. These tumors are relatively large, making up 33% of all brain tumors, and can typically be found in all cortexes of the brain as well as the spinal cord.

Meningioma tumors develop in the meninges affecting the protective lining of the brain. These tumors are of medium size, making up 40% of all brain tumors, and can typically be found around the edges where the brain meets the skull.

Pituitary tumors develop in the pituitary gland affecting the production and regulation of hormones. These tumors are relatively small, making up 17% of all brain tumors, and can only be found in the pituitary gland.

About the Data

The data originates from 4 sources, totaling around 7 thousand grayscale MRIs to form one cumulative database. With 3 datasets originating from the uploads of anonymous radiologist and neuroscientists on Kaggle, and 1 dataset manually constructed of hundreds of individual MRIs from figshare.

The 3 Kaggle datasets consisted of T1 MRIs sorted into 4 different categories based on the tumor class present: no-tumor, glioma, meningioma, and pituitary. However, the MRIs from figshare were sorted into 132 classes with 44 tumor subclasses each with T1, T1+, and T2

images. Only the T1 MRIs were extracted for consistency, and images from any of the subclasses that compose the glioma category (astrocytoma, ependymoma, glioblastoma, and oligodendroglioma) were appended to the glioma database. The T1 MRIs labeled no-tumor, meningioma, and pituitary were appended to the existing databases, respectively.

The final database was composed of exactly 7,023 T1 MRIs split 80/20 into a training set of 5,712 images, and test set of 1,311 images. As well as 2,000, 1,621, 1,645, and 1,757 images categorized as no-tumor, glioma, meningioma, and pituitary, respectively.

	Glioma	*Meningioma*	*Pituitary*	*Non-Tumor*	Total
Training	1321	1339	1457	1595	5712
Testing	300	306	300	405	1311
Total	1621	1645	1757	2000	7023

Additionally, the images were manually sorted into 1 of 9 categories based on the angle and depth of each MRI: Left, Left (Eyes), Right, Right (Eyes), Front, Front (Eyes), Top, Top (Eyes), and Rear. Lastly, the images were diligently inspected and discarded if they did not meet the strict quality criteria.

Since many MRIs varied drastically in angle, depth, size, rotation, and quality, around half of the dataset would have been discarded. To partially avoid this problem, the exploratory data analysis, model development, and result validation processes were split into 2 parts, differing in preliminary data processing, model results and performance, and objective.

PART I

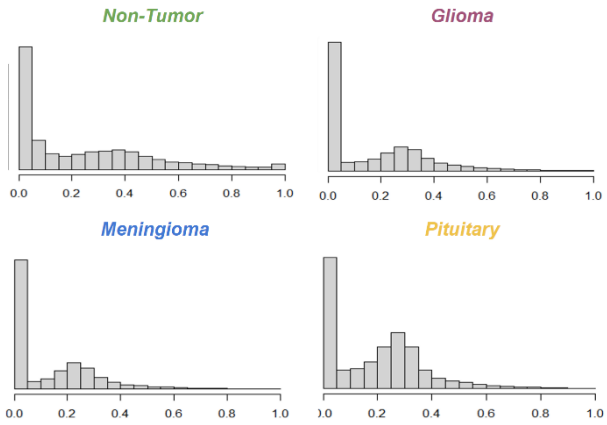
Objective

The objective of Part I was to develop a model that utilized all 7,023 MRIs available to classify images into 1 of 2 classes: no-tumor and tumor. Though model performance would heavily depend on the characteristics of the training set, (heterogenous images, unprocessed images, unbalanced image distribution, etc.) it would have the advantage of a much larger training set to develop a more versatile predictive model.

However, its illusory high performance in training accuracy would likely lead the model to struggle in real world application. A large proportion of Part I's predictive accuracy can be attributed to the model's ability to notice trends in distribution of available data, instead of optimizing feature selection in correlation with tumor classes.

Exploratory Data Analysis

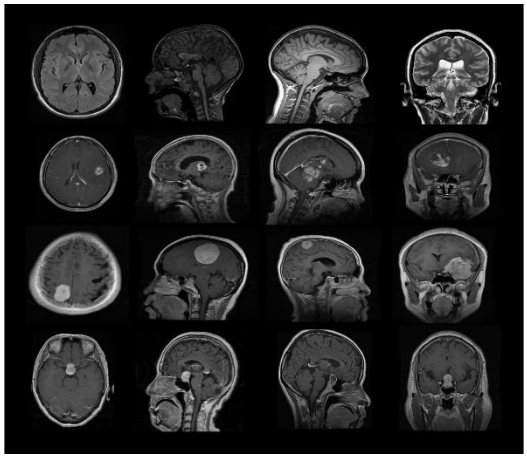
An initial investigation of the characteristics indicated that our MRIs varied significantly in size and angle, as well as unnoticeable parameters such as image dimensions and color scale. Image dimensions ranged from 512x512 px to 128x128 px, while color scales varied between grayscale and RGB. To reduce potential complications and for consistency, all images were resized to 256x256 px, converted to grayscale, and normalized in respect to pixel intensity on a 0-1 scale.



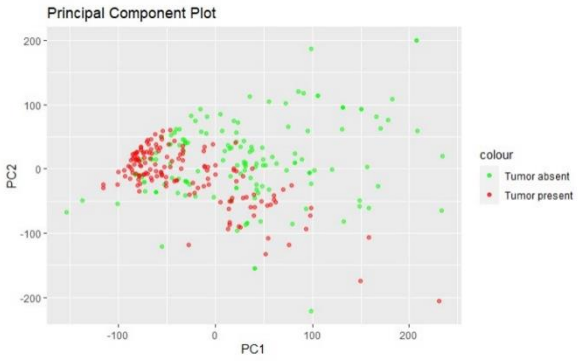
The above histograms of pixel intensity for each tumor class show a slightly right skewed distribution centered at 0.3 (gray) with a long right tail in the 0.6-1.0 range (light gray – white). This indicates that a large portion of the MRI pixel intensities are normally distributed around 0.3 with various types of brain tissues occupying the 0.1-0.5 range, and tumor growths occupying the 0.8-1.0 range. Additionally, a mode at 0 (black) is visible, indicating that a significant portion of the MRI pixels background noise.

Though the issue of background noise was not addressed in Part I, the initial pixel distribution indicates that pixel intensities alone

could reveal patterns with which our model can utilize to correctly detect tumors.



This finding signaled that elementary machine learning techniques such as principal component reduction (PCA) would be surprisingly efficient at reducing image dimensionality while maximizing within image percent variance explained (PVE).



PCA revealed that a simple re-projection of our images onto a subspace composed of the first two principal components (PC) could yield relatively high predictive accuracy. The visible clusters indicate that PCA alone may not be sufficient as a predictive model. However, it does indicate that integration without more advanced techniques would allow for high model performance at a significantly reduced computation cost.

Support Vector Machines

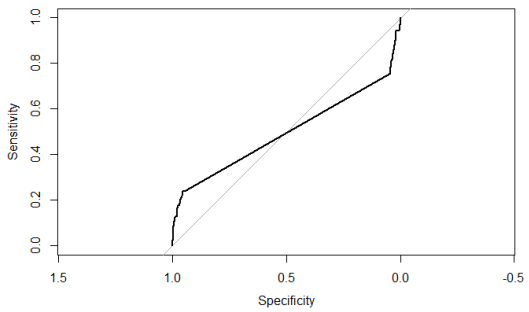
The first model developed was a support vector machine (SVM). This machine learning algorithm works by creating a hyperplane to separate our 7,023 observations in a 65,536-dimension space into just 2 classes: no-tumor and tumor. The hyperplane is constructed to maximize the margin between the tumor classes. This margin defines the distance between the hyperplane and its closest observations.

As indicated by PCA, designing a linear hyperplane to classify MRIs in just 2-D yields a low prediction accuracy. Subsequently it was logical to assume a linear classifier would perform significantly worse than a non-linear hyperplane in 65,536 dimensions.

PCA analysis was used to select the optimal number of PCs that efficiently balanced dimension reduction without sacrificing too much loss in PVE. The optimal number of PCs was 145 PCs with an 80% cumulative PVE. All images were then re-projected onto the optimal 145-dimension feature and classified by SVM.

To avoid issues regarding low performance classifiers, a radial kernel was used to construct the SVM hyperplane. The parameters of this hyperplane were chosen by evaluating 10-fold cross validation (CV) accuracy across various combinations of C (margin size) and G (decision boundary complexity).

A grid search was performed across a total of 35 different combinations of C and G values, selecting from 1 of 7 margin sizes (0.001, 0.01, 0.1, 1, 5, 10, 100) and 5 levels of decision boundary complexity (0.5, 1, 2, 3, 4). The optimal SVM model was found using 10-fold CV, C = 5, and G = 0.5. Following this cost/gamma selection, a correlated receiver operating characteristic (ROC) curve was utilized to find the optimal classification threshold. For this SVM model, the optimal threshold value was found at 0.7065 which maximized the prediction accuracy as well as sensitivity and specificity.



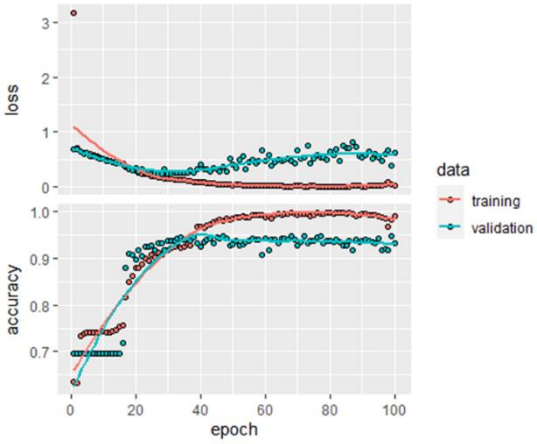
MRI Tumor Detection & Classification in the Brain using Machine Learning

Unfortunately, SVM was not able to perform very well, reaching maximum performance with specificity, sensitivity, and accuracy of 89.85%, 34.32%, and 51.26%, respectively. Poor SVM performance indicated that more complex models were required.

Sequential Neural Network

To improve model performance after SVM, a drastic jump in model complexity was made with the development of a sequential neural network (SNN). An 80/20 training/test split with a 256-image batch size was used to train the 4-layer SNN across 100 epochs. The SNN was designed with a 30,000-node input layer, 2 hidden layers at 500-nodes and 30-nodes, and a 2-node output layer. The network performance was evaluated by the Relu activation function at each node in the first 3 layers and a SoftMax activation function in the final layer. Additionally, a random 20% node dropout parameter was set within each layer to avoid overfitting to the training data.

The model was set to optimize validation accuracy and minimize categorical cross-entropy using the stochastic-gradient descent (SDG) optimizer through back propagation. While tracking model performance across epochs, it was observed that the optimal validation accuracy and loss were found at the 20-epoch level with 94.57% classification accuracy. From epoch 20-100 the validation accuracy stayed relatively stable.

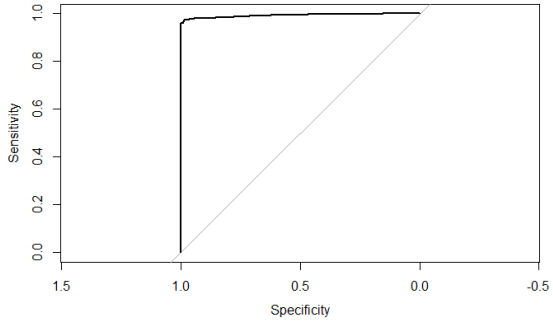


This indicates that the model reached optimal performance in the first 20 epochs, while the random 20% node dropout stabilized the validation accuracy and prevented the model from overfitting to the training data for the last 80 epochs.

Random Forests

Though the SNN achieved more than satisfactory predictive accuracy, an alternative machine learning algorithm, random forests (RF), was explored to test the reproducibility of these results. An 80/20 training/test split was used to train a RF. The parameters of this model were chosen by evaluating 10-fold CV accuracy across various combinations of ntree (number of trees) and mtry (number of features to be randomly selected at each tree split).

A grid search was performed across a total of 9 different combinations of ntree and mtry values, selecting from 1 of 3 ntree (100, 200, 500) and 3 levels of mtry (30, 150, 300). The optimal RF model was found using 10-fold CV, ntree = 500, and mtry = 30. Following this ntree/mtry selection, an ROC curve was utilized to find the optimal classification threshold. For this RF model, the optimal threshold value was found at 0.5057 which maximized the prediction accuracy as well as sensitivity and specificity.



Fortunately, RF proved to be a better predictive model than SNN, reaching maximum performance with specificity, sensitivity, and accuracy of 100%, 94.92%, and 96.49%, respectively.

Overview

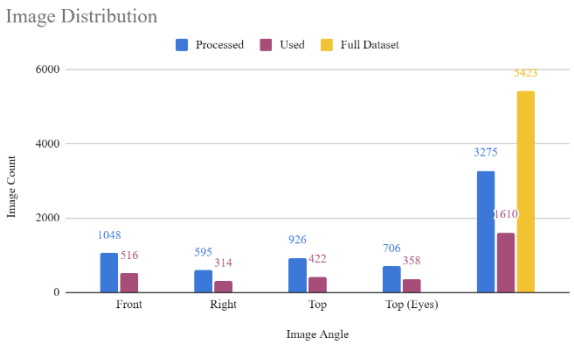
Of the 3 models the RF, SNN, and SVM models produced prediction accuracies of 96.49%, 94.57%, and 51.26%, respectively. The RF model predicted tumor class at the highest accuracy, with a1.92 percentage point advantage of SNN and a 45.23 percentage point advantage over SVM. In tumor detection, false negative

rate (sensitivity) is a much more crucial statistics to emphasize. The RF model optimally balanced sensitivity 94.92% and accuracy 96.49% and was therefore chosen as the best performing ML model for MRI tumor detection.

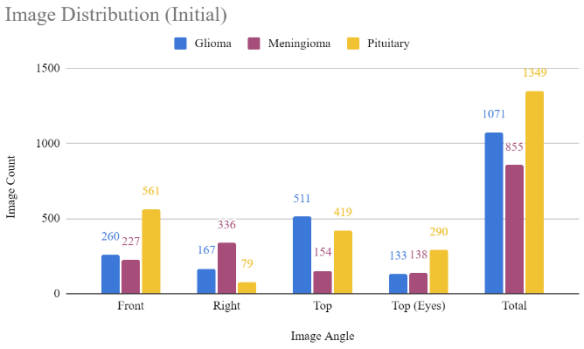
PART II

Objective

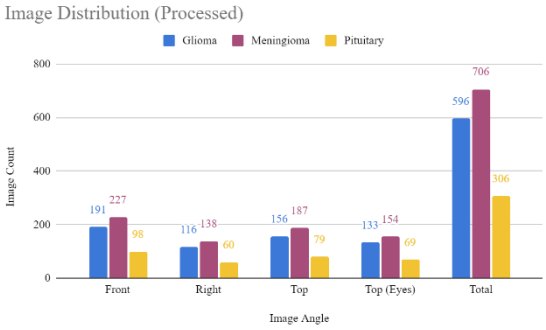
The objective of Part II was to develop a model that only utilized data that satisfies the strict quality criteria to classify images into 1 of 3 classes: glioma, meningioma, and pituitary. Though model performance would heavily depend on the characteristics of the training set, the dataset was significantly reduced, cleaned, and processed to remove these dependencies. Excluding MRIs from the no-tumor class, the dataset was reduced from 5,423 to 3,275 hand-picked images.



The data was thinned to ensure that each image angle had relatively similar distributions in the dataset. Many of MRI angles lacked a substantial image count and failed to qualify onto further analysis. Only MRI angles Front, Right, Top, and Top (Eyes) proceeded.



The remaining MRIs were thinned from 3,275 to 1,610 images to ensure IID and match the approximate 33%, 40%, 17% real world brain tumor distribution.



This process eliminated the concerns of Part I at the expense of data quantity as well as variance in MRI angle, depth, and rotation. However, the advantage of using heavily cleaned data for training ML models outweighed the loss in data. Through significant dimension reduction techniques such as PCA, LDA, and NCA, as well as heavy data cleaning and processing, we developed ML models with higher classification accuracy than in Part I.

Exploratory Data Analysis

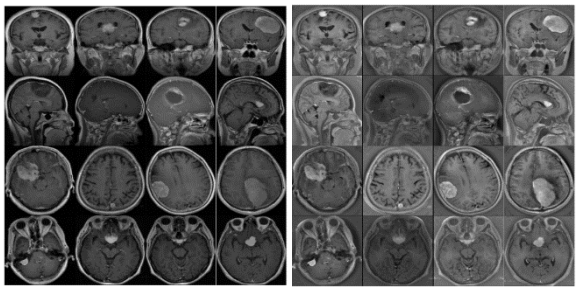
To surpass the predictive accuracy achieved by SVM, SNN, and RF in Part I with only 23% of the original data was a challenge. To gain maximum utility from the remaining data, a series of erosions, dilations, contouring, and cropping was performed on the MRIs to increase image quality.

First, the images were converted to grayscale, standardized on a 0-1 pixel intensity scale, and resized to 256x256 px. Second, the images were passed through a series of dilations and erosions. This process effectively brightened all pixel densities identified as brain matter and darkened all remaining pixels identified as background noise. Then, the images were passed through a series of contouring and cropping techniques. This effectively outlined the circular/oval shape of the highlighted brain and cropped out the background noise.

Finally, the images underwent feature wise and image wise standardization. This was done to highlight pixels considered within feature outliers, while standardizing pixel intensities with respect to within image outliers.

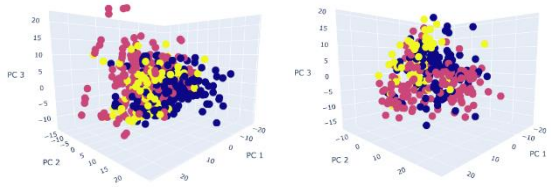
MRI Tumor Detection & Classification in the Brain using Machine Learning

Processed Standardized

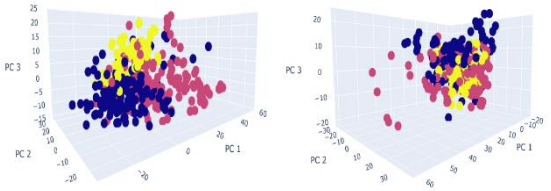


The processed and further standardized images were used to construct preliminary PCA tri-plots which were analyzed to estimate how PCA would balance computation cost and model performance. All MRI images were re-projected onto the first 3 PCs, producing 4 PCA tri-plots, one for each MRI angle.

Front Right



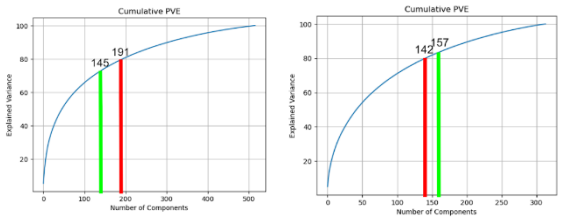
Top Top (Eyes)



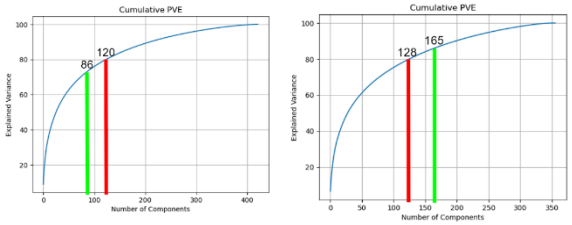
The PCA tri-plots indicate the SVM and KNN predictive models will perform better on the Front and Top angle MRIs. From the tri-plots it is visible that the Front and Top angle MRIs can almost be visually categorized into 3 distinct tumor clusters. This coincides with cumulative PVE values of 23.71% and 33.43% for the Front and Top MRIs, respectively. The Right and Top (Eyes) images display less visible tumor clusters as well as lower cumulative PVE values at 18.38% and 18.77%, respectively.

The 3-D subspace composed of the first 3 PCs explains between 18.38% - 33.45% of the variance in the data. This implies the heavily processed MRIs could likely be re-projected onto a significantly reduced subspace while maintaining high model performance. Like in Part I, the optimal number of PCs that define our image subspace was chosen around the 80% cumulative PVE. The optimal number of PCs was found around the 70% level for the Front and Top MRIs, while optimal number of PCs was determined around the 90% level for the Right and Top (Eyes) MRIs.

Front Right



Top Top (Eyes)



The optimal PC values were tested and verified to produce better SVM and KNN results as the MRI angles that had larger initial PVE values (Front and Top) could afford a larger dimension reduction without sacrificing model performance. While the MRI angles with lower initial PVE values (Right and Top (Eyes)) could not afford a larger dimension reduction and therefore developed less accurate models.

Support Vector Machines

An initial re-projection of each standardized MRIs onto their optimal number of PCs as found in the EDA was conducted. These dimensions reduced re-projected images were split using an 80/20 training/test split and used to train the SVM model and find the optimal non-linear hyperplane. The SVM model with a radial kernel model was developed using a grid search that evaluated validation accuracy under 10-fold CV at every combination of C and G between 0-100 and 0-10, respectively.

Front

	precision	recall	f1-score	support
Glioma	0.90	0.92	0.91	38
Meningioma	0.86	0.96	0.91	46
Pituitary	0.93	0.65	0.76	20
accuracy			0.88	104
macro avg	0.90	0.84	0.86	104
weighted avg	0.89	0.88	0.88	104

Right

	precision	recall	f1-score	support
Glioma	0.74	0.87	0.80	23
Meningioma	0.83	0.89	0.86	28
Pituitary	1.00	0.50	0.67	12
accuracy			0.81	63
macro avg	0.86	0.75	0.78	63
weighted avg	0.83	0.81	0.80	63

Top

	precision	recall	f1-score	support
Glioma	0.93	0.84	0.88	31
Meningioma	0.86	0.97	0.91	38
Pituitary	1.00	0.88	0.93	16
accuracy			0.91	85
macro avg	0.93	0.90	0.91	85
weighted avg	0.91	0.91	0.91	85

Top (Eyes)

	precision	recall	f1-score	support
Glioma	0.81	0.81	0.81	27
Meningioma	0.84	0.87	0.86	31
Pituitary	0.92	0.86	0.89	14
accuracy			0.85	72
macro avg	0.86	0.85	0.85	72
weighted avg	0.85	0.85	0.85	72

SVM performance greatly exceeded expectations and achieved 88%, 91%, 81%, and 85% tumor classification accuracy for Front, Top, Right, Top (Eyes) respectively. These results coincide with the 24%, 34%, 18%, and 19% PVE distribution as found by PCA for Front, Top, Right, Top (Eyes) respectively. These results also coincide with the 70%-90% PVE value criteria for selecting optimal PCs.

The model results show that classification accuracy was consistent highest across pituitary tumors. This is likely because pituitary tumors do not vary in size or location. While glioma and meningioma tumors are not as apparent nor distinct in tumor size or location. This explains why a common misclassification error across all models is between glioma tumors predicted as meningioma and vice versa.

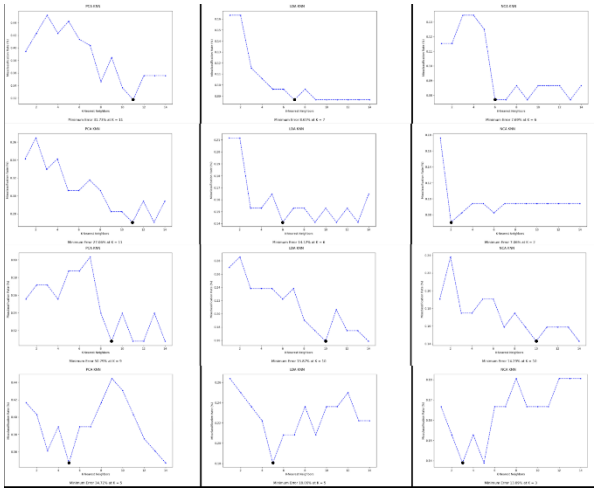
K-Nearest Neighbors

The integration of PCA with SVM created a computationally inexpensive PC subspace. This allowed for an exhaustive grid search to find optimal parameters composing the non-linear decision boundary for optimal SVM model performance. KNN, however, is a less complicated ML technique that attempts to learn very little about the data. The integration of re-projected MRIs onto a PCA subspace with the KNN algorithm would not yield as high results. Instead, various other dimension reduction techniques were explored as alternatives to PCA.

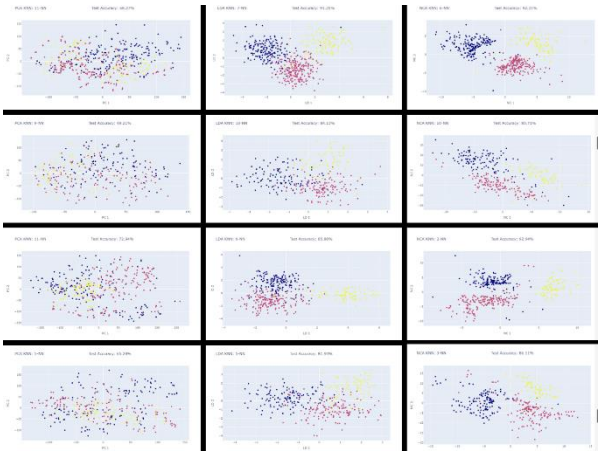
Linear discriminant analysis (LDA) is a classification technique that divides the feature space into linear combinations of features for optimal classification. LDA works by projecting the data onto a lower dimensional subspace that maximizes the ratio of between-class variance to within-class variance. Like the unsupervised PCA technique that maximizes PVE, the supervised LDA technique maximizes classification accuracy. Similarly, neighborhood component analysis (NCA) divides the feature space by evaluating leave-one-out (LOO) classification performance through a process called stochastic nearest neighbors. Like LDA, this supervised NCA technique maximizes classification accuracy.

Both LDA and NCA can be used in a similar manner to PCA. By re-projecting the standardized MRIs onto the linear discriminants that compose the LDA subspace, we can create more distinct tumor clusters and improve KNN model performance. Since both LDA and NCA derive a more performance-oriented subspace, we should expect the KNN algorithm to yield better results on the LDA/NCA subspace than the PCA subspace.

The KNN models were developed using a 80/20 training/testing split, where an exhaustive grid search was performed to find the optimal 10-fold CV accuracy across nearest neighbors (K) between 1-15. The value of K was chosen to the at the value that minimized misclassification error. A total of 12 models were developed, with optimal K values found for each subspace at each MRI angle.



From left to right, the optimal K is found in the subspace constructed by PCA, LDA, then NCA, respectively. Noticeably, as the subspace decreases in complexity and orientations toward selecting features for optimized classification accuracy, the optimal K consistently decreases at each MRI angle.



From left to right, the KNN model performance is evaluated in the subspace constructed by PCA, LDA, then NCA, respectively. Noticeably, as the subspace decreases in complexity and orientations toward selecting features for optimized classification accuracy, the model performance increases at each MRI angle. This coincides with the progressive tumor clustering and decrease in K from PCA to LDA to NCA.

PCA KNN averaged K = 9 and classification accuracy of 63.93%, and varied form K = 5-11 and classification accuracy from 49.21% - 72.94% depending on MRI angle.

LDA KNN averaged K = 7 and classification accuracy of 85.83%, and varied form K = 5-10 and classification accuracy from 81.94% - 91.35% depending on MRI angle.

NCA KNN averaged K = 5 and classification accuracy of 89.27%, and varied form K = 2-10 and classification accuracy from 85.71% - 92.94% depending on MRI angle.

KNN performance was highest in the Front and Top MRIs, with a maximum KNN classification accuracy of 92.31% and 92.94%, respectively. This coincided with the previous results found in PCA and SVM.

Convolutional Neural Network

The convolutional neural network (CNN) is the most complex model developed in our research. The optimal model constructed was a 6-layer CNN with 32, 64, 128, 128, 64, and 32 features selected within each respective layer. The feature selection was designed to prevent excessive information loss between layers. The 32-128 feature range proved to adequately balance optimal feature extraction and limit model complexity. Feature extraction was performed using a 4x4 kernel with 2x2 max pooling at every layer. Padding was to the convolution output to keep the input dimensions constant between layers. The model also included 6 dense layers that reduce the feature space to 32 features which would determine the model classification output.

The final 32 feature dense layer served as the input to a 6-layer neural network with 32, 64, 128, 128, 64, and 32 neurons at each respective layer. This framework was chosen in symmetry of the 6-layer convolution design. Through experimentation within the 5%-25% neuron dropout rate range, a 10% neuron dropout rate was chosen. This parameter served our purpose best as it decreased model complexity without risking an overfit to the training data. To fit the model, around 32 batches of 16 MRIs were used to train the 4.3 million parameter CNN across 20 epochs.

The LeakyRelu activation function was utilized within the 5 input layers. This helped avoid performance plateaus in situations where the optimization gradient derived through back propagation approaches 0, leading to a stagnant model with suboptimal performance. The SoftMax activation function was utilized at the output layer as a measure to direct our model classification towards minimizing categorical cross entropy.

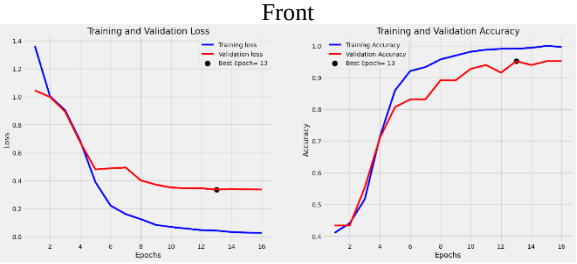
The SGD optimizer with an initial learning rate of 0.001 was used to optimize our model. Special callback functions were implemented to assign a model patience of 3 with a learning parameter adjustment multiplier of 0.5. These parameters allowed for automatic supervision of model performance across 20 epochs. The patience parameter required minimization of training and validation loss through back propagation significant performance improvements within 3 epochs. If the model did not meet this criterion, the learning rate would be adjusted by a factor of 0.5 for the remaining epochs. This mechanism slows the development of a model that completely minimizes training loss at the expense of validation accuracy. By effectively halving the learning rate, all subsequent epochs contribute less to fitting the model. This allows for optimal model development to occur at earlier epochs, and validation and testing accuracy to stabilize in later epochs.

```
Do you want model asks you to halt the training [y/n] ?
y
 Epoch     Loss   Accuracy  V_loss    V_acc     LR     Next LR  Monitor  % Improv  Duration
 1 /20     1.158   39.514   1.02879  38.554   0.00100  0.00100  accuracy     0.00    12.15
 2 /20     0.949   55.927   0.87468  71.084   0.00100  0.00100  accuracy    41.54     9.29
 3 /20     0.689   73.252   0.60871  78.313   0.00100  0.00100  accuracy    30.98     9.22
 4 /20     0.460   82.371   0.40708  84.337   0.00100  0.00100  accuracy    12.45     9.67
 5 /20     0.330   86.930   0.38811  90.361   0.00100  0.00100  accuracy     5.54     8.69
enter H to halt training or an integer for number of epochs to run then ask again
5
 training will continue until epoch 10
 Epoch     Loss   Accuracy  V_loss    V_acc     LR     Next LR  Monitor  % Improv  Duration
 6 /20     0.283   89.362   0.29786  90.361   0.00100  0.00100  accuracy     2.80     9.85
 7 /20     0.295   89.362   0.27027  90.361   0.00100  0.00050  accuracy     0.00     9.39
 8 /20     0.188   93.617   0.24212  92.771   0.00050  0.00050  val_loss    10.42    10.71
 9 /20     0.130   95.441   0.17222  92.771   0.00050  0.00050  val_loss    28.87     9.25
10 /20     0.080   97.872   0.17589  95.181   0.00050  0.00025  val_loss    -2.13     9.27
enter H to halt training or an integer for number of epochs to run then ask again
5
 training will continue until epoch 15
 Epoch     Loss   Accuracy  V_loss    V_acc     LR     Next LR  Monitor  % Improv  Duration
11 /20     0.064   98.480   0.15917  92.771   0.00025  0.00025  val_loss     7.58     9.30
12 /20     0.049   98.784   0.16270  95.181   0.00025  0.00013  val_loss    -2.22     9.39
13 /20     0.043   99.088   0.16321  95.181   0.00013  0.00006  val_loss    -2.54     9.58
14 /20     0.042   99.088   0.16188  93.976   0.00006  0.00003  val_loss    -1.70    10.72
 training has been halted at epoch 14 after 3 adjustments of learning rate with no improvement
training elapsed time was 0.0 hours,  2.0 minutes, 31.25 seconds)
```

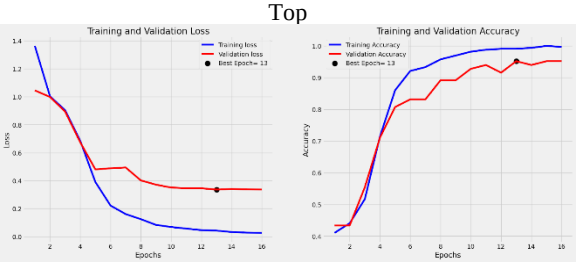
MRI Tumor Detection & Classification in the Brain using Machine Learning

The model performance across epochs was carefully monitored, and experimentation was required to decide all the parameters in developing optimal CNN. It was found that the Front and Top MRI angles achieved high validation accuracy at 92.56% and 95.79%, respectively.

Front

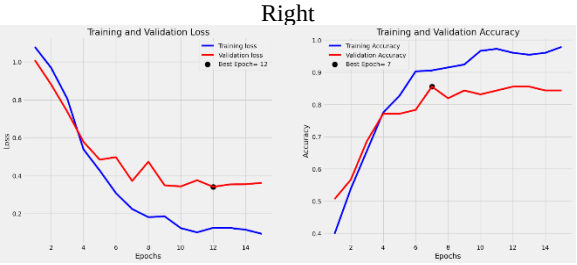


Top

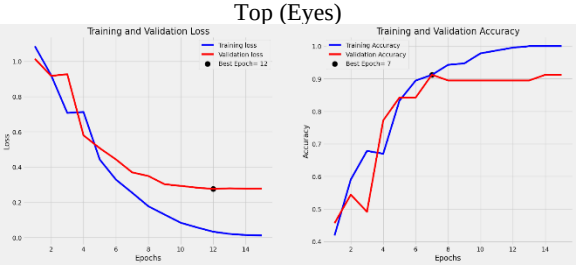


While the Right and Top (Eyes) achieved lower validation accuracy at 86.35%, and 91.67%, respectively. These results coincided with the results found in PCA, SVM, and KNN as Front and Top MRI angles consistently outperform Right and Top (Eyes).

Right



Top (Eyes)



However, the Front and Top MRI angles reach peak model performance around 15 epochs, while the Right and Top (Eyes) reached peak model performance around 7 epochs. This also coincided with the conclusion from EDA and optimal number of PCs for SVM/KNN as the Right and Top (Eyes) angles required a larger number of PCs to reach optimal PVE. This indicates that some angles may not contain as many critical features that distinguish tumor classes as effectively as others. Therefore, requiring fewer epochs to reach and stabilize suboptimal model performance when fitting to less potent MRI angles. Surprisingly, angles that reached peach performance at earlier epochs utilized later epochs to better stabilized model performance. This produced a more balanced fit between training and validation accuracy while producing more stable testing results.

Front	Right	Top	Top (Eyes)
92.56% Validation 92.11% Test	86.35% Validation 87.11 % Test	95.79 % Validation 94.26 % Test	91.67% Validation 91.78% Test

Overview

Of the 4 models produced through integration of PCA with SVM, peak model performance was found using Top angle MRIs projected onto the first 86 PCs with a 72% PVE to achieve 91% testing accuracy.

Of the 12 models produced through integration of PCA, LDA, and NCA with KNN, peak model performance was found using Top angle MRIs projected onto the NCA linear discriminant to achieve 92.94% testing accuracy.

The CNN model produced peak model performance using 6 convolution, dense, and neuron layers, 4.3 million total parameters, 10% neuron dropout rate, 0.001 learning rate, 0.5 learning rate adjustment factor, 16 batch size, and a 3-epoch patience across 20 epochs. Peak performance was found using Top angle MRIs to achieve a 95. 79% testing accuracy.

With heavy data cleaning, processing, and selection, using just 1,610 of the original MRIs, the 6-layer CNN was able to outperform the 4-layer SNN trained on 7,023 MRIs by 1.22 percentage points. For this reason, the CNN model was selected as the best performing ML model for MRI tumor classification.

Conclusion

The developed tumor detection and classification models prove that achieving results produced by pretrained neural networks design for MRI analysis, is not infeasible.

Research standards can be met with proper EDA and image processing technique, and clinical results can be approximated using advanced ML algorithms. The results in our research show the potential ML holds in developing models with accuracy upwards of 94.57% for tumor detection and upwards of 95.79% for tumor classification.

However, it is important to note the limitations of our models and the applications of our research. The undersupply of correctly classified high quality MRI scans at various depths, angles, and rotations limited the ability for our ML models to achieve better results. This limitation was circumvented but not avoided.

The highest performing model required heavy data processing MRIs from patients with tumors from the Top angle. Hence, successful diagnosis for patients with glioma, meningioma, or pituitary requires the presumption that the patient's tumor is malignant, and that Top angle MRIs are available.

As for future improvements for our research, developing a model that accounts for multiple categories of MRI depth, angle, and tumor subclass could potentially improve the versatility and medical application of our models. The optimal models in Part I and Part II can be combined for these purposes. Essentially conducting two tests, first detecting tumor presence in MRIs, and second, accurately classifying tumorous growth by tumor class.

Future models should build off the optimal RF and CNN models developed in this research. Suggested improvements include processing and fitting models to larger quantities of high-quality MRIs from all 9 MRI angles. Additionally, improvements should extend the scope of this research to ML MRI tumor detection and classification for the 120 different subclasses of brain tumors.

Bibliography

Brain Tumor MRI Dataset | Kaggle
Msoud Nickparvar. (2021).
Brain Tumor MRI Dataset.
[Data set]. Kaggle.
https://doi.org/10.34740/KAGGLE/DSV/264588

Brain Tumor Classification (MRI) | Kaggle
Sartaj Bhuvaji, Ankita Kadam, Prajakta Bhumkar, Sameer Dedge, & amp; Swati Kanchan. (2020).
Brain Tumor Classification (MRI).
[Data set]. Kaggle.
https://doi.org/10.34740/KAGGLE/DSV/118315

Brain Tumor Dataset (figshare.com)
Jun Cheng. (2017).
Brain Tumor Dataset.
[Data set]. figshare.
https://doi.org/10.6084/m9.figshare.1512427.v5

Br35H :: Brain Tumor Detection 2020 | Kaggle
Ahmed Hamada. (2020).
Br35H :: Brain Tumor Classification 2020.
[Data set]. Kaggle.

Brain Tumor MRI Images 44 Classes | Kaggle
Fernando Feltrin. (2023).
Brain Tumor MRI Images 44 Classes.
[Data set]. Kaggle.