JOURNAL OF
EMERGING INVESTIGATORS

Article

A comparative analysis of machine learning approaches to predict brain tumors using MRI

Eileen Zhao[1], Hairong Wang[2], Jing Li[2]
[1] Westwood High School, Austin, Texas
[2] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia

SUMMARY

A brain tumor is an excessive development of dangerous cells in the brain and is one cause of cancer deaths worldwide. Early detection is important for the effective treatment of brain tumors. To do this, magnetic resonance imaging is used as a safe and non-invasive method. In recent years, machine learning (ML) has attracted more attention due to its achievements in the field of disease diagnosis and precision medicine. Many studies have proposed using ML to predict the onset of a brain tumor. The main objective of this analysis was to determine whether ML algorithms can predict the onset of brain tumors accurately and effectively. We hypothesized that ML algorithms would provide an accuracy of more than 90%, and an ensemble model would provide better accuracy within 30 minutes of computational time. Among various classification methods, we compared the two most popular ML algorithms, support vector machine (SVM) and random forest (RF), to provide some insight on model selection for brain tumor classification. To test the generalizability of each ML algorithm on independent datasets, we used cross-validation to evaluate model performance. To further improve the model performance, we introduced grid search to find the optimal values of hyperparameters for the classifiers. We observed the trade-off between performance improvement and time consumption of grid search. Finally, we developed an ensemble model which integrated k-fold cross-validation with the SVM and RF classifiers to achieve the highest model accuracy with less time and make the traditional ML methods more robust.

INTRODUCTION

The most complex organ in the human body is the brain (1). It contains billions of neurons, and they all must function properly to regulate different physiological processes (2). Brain tumors are abnormal cell aggregations that grow within brain tissues (3). There are two types of brain tumors: benign tumors and malignant tumors (4-6). Benign tumors are non-cancerous and grow slowly. They are small in size, feature well-defined borders, do not spread to other areas like the spinal cord, other parts of the brain, or other areas of the body, and can be removed completely by surgery (7). On the other hand, malignant brain tumors can lead to one of the deadliest types of cancers. They are fast-growing and contain an irregular structure that allows them to spread to other parts of the brain or the spinal cord. These tumors include active cancer cells, which can affect healthy brain cells, causing brain swelling and increased pressure in the skull (8).

According to the National Brain Tumor Society, an estimated 700,000 people in the United States are living with a primary brain tumor, and about 88,970 more will be diagnosed in 2022, with an estimated 25,930 of these with malignant tumors (9). A typical brain tumor can double in size within just 25 days. If not treated appropriately, the patient's survival rate is typically less than 12 months (10). Hence, early detection of brain tumors is highly important for effective treatment.

Brain tumor detection is a challenging task for biomedical imaging. Many diagnostic imaging techniques allow for early detection of brain tumors, such as computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI) (4). MRI is a safe and noninvasive imaging technique used commonly in neurology (6). This scan uses radio frequency signals to excite the target tissue under the influence of a strong magnetic field to produce an image of the interior. The method has the advantage of achieving a high soft-tissue contrast imaging with zero ionizing radiation exposure. Therefore, MRI is one of the most effective imaging methods for the detection of brain lesions.

An important task of brain tumor detection is brain tumor classification, which can assist clinicians in making an accurate diagnosis and creating treatment plans. Brain tumor classification using MRI is usually performed manually by radiologists or medical experts. Classifying tumors this way consumes lots of time, and the accuracy is also dependent on the experience of the diagnosing clinician. These limitations could potentially be overcome by developing automated systems that help in the diagnosis and prognosis of brain tumors using machine learning (ML) algorithms (11-13).

The application of information technology and ML in the medical field has expanded in modern times. Many studies have shown that ML algorithms can perform equally to or even better than human experts at various healthcare tasks, including brain tumor classification. ML is a scientific field which works on developing statistical models that can learn on their own without human intervention. This is done through training data to help make predictions for unseen testing data. One major type of ML algorithms is supervised classification, also known as predictive or directed classification. In this method of learning, the labels of training data are known in advance. Using supervised learning, we aimed at training a model that can classify MRI images with versus without a brain tumor from the Br35H dataset, a large public dataset containing 3000 brain MRI images (14). Br35H is widely recognized and popular in the field of medical image analysis, particularly

https://doi.org/10.59720/23-082

in the context of brain tumor classification, detection, and segmentation. Many research papers have leveraged Br35H to showcase the effectiveness of their proposed methods to advance medical image analysis (15-19).

Many ML techniques have been developed to detect cancer at early stages. Additionally, several learning-based classifiers have been used for the classification and detection of brain tumors, including support vector machine (SVM), artificial neural network (ANN), Naïve Bayes (NB), random forest (RF), decision tree (DT), k-nearest neighbor (KNN), etc. (20).

We aimed at improving these performance metrics by using traditional classifiers instead of deep learning classifiers. Traditional classifiers have advantages over deep learning algorithms such as small datasets for training, low computational time complexity, low cost to users, and being easy to learn (20).

The main objective of this paper was to provide insight on how to select models for brain tumor classification through a comparative study of SVM and RF classifiers. SVMs are widely used in various applications, including image classification, text categorization, and bioinformatics. They are known for their effectiveness in high-dimensional spaces and their ability to handle both linear and nonlinear classification problems. SVMs have a strong theoretical foundation and are robust against overfitting when the regularization parameter is appropriately tuned. The primary goal of a SVM is to find the best hyperplane that separates data points of different classes while maximizing the margin between them. RF is also widely used in a variety of applications, including image classification, recommendation systems, finance, and healthcare (21, 22). Its ability to handle both classification and regression tasks, handle high-dimensional data, and provide robust and interpretable results has made it a popular choice among ML practitioners. Decision trees are prone to overfitting, and RF addresses this issue by creating a collection of decision trees and combining their predictions.

Other objectives of this paper were to provide guidance on how to improve model performance by grid search and how to better evaluate models by cross-validation (CV) strategies through a comparative study of different CV methods. Grid search can help to find the best tuning parameters. Finding the right parameters (such as the C or gamma values to use in SVM) is a challenging task, but grid search allows for many combinations to be trialed to determine what works best. This analysis illustrates the trade-off between performance improvement and time cost from grid search.

Algorithm	The proposed model	Testing method	TP	FP	TN	FN	Accuracy	Precision	Recall	F1 Score
#1	SVM	70% split	423	28	422	27	0.939	0.938	0.94	0.94
#2	RF	70% split	431	32	418	19	0.943	0.931	0.958	0.94
#3	StandardScaler+ SVM	70% split	433	20	430	17	0.959	0.956	0.962	0.96
#4	StandardScaler+ RF	70% split	431	26	424	19	0.95	0.943	0.958	0.95
#5	GridSearchCV + SVM	70% split	441	34	416	9	0.952	0.928	0.98	0.95
#6	GridSearchCV + RF	70% split	433	30	420	17	0.948	0.935	0.962	0.95

Table 1: Comparative study of SVM and RF classifiers.

Finally, an ensemble model was developed using k-fold CV on the training dataset and all out-of-fold predictions from SVM and RF classifiers. This improved model achieved the highest accuracy with less time and makes the traditional ML models more robust.

Our study confirmed the hypothesis that all classification algorithms could predict brain tumor onset with an accuracy exceeding 90%. Notably, the ensemble model demonstrated advantages by achieving higher accuracy while significantly reducing computational time, making it both more efficient and precise compared to other models.

RESULTS

The Br35H (Brain Tumor Detection 2020 dataset) contains 3000 images of brain MRI images which are divided into two classes, 1500 tumorous and 1500 non-tumorous (14). Images were resized to a fixed size and flattened before inputting them into the classifier.

Comparison of SVM and RF based on random train-test split

In this study, two popular classifiers, SVM and RF, were chosen, and six algorithms were developed with the flattened data to compare the effect of applying data normalization preprocessing and grid search hyperparameter tuning.

To evaluate the performance of different classifiers, we used the confusion matrix, accuracy, precision, recall, and

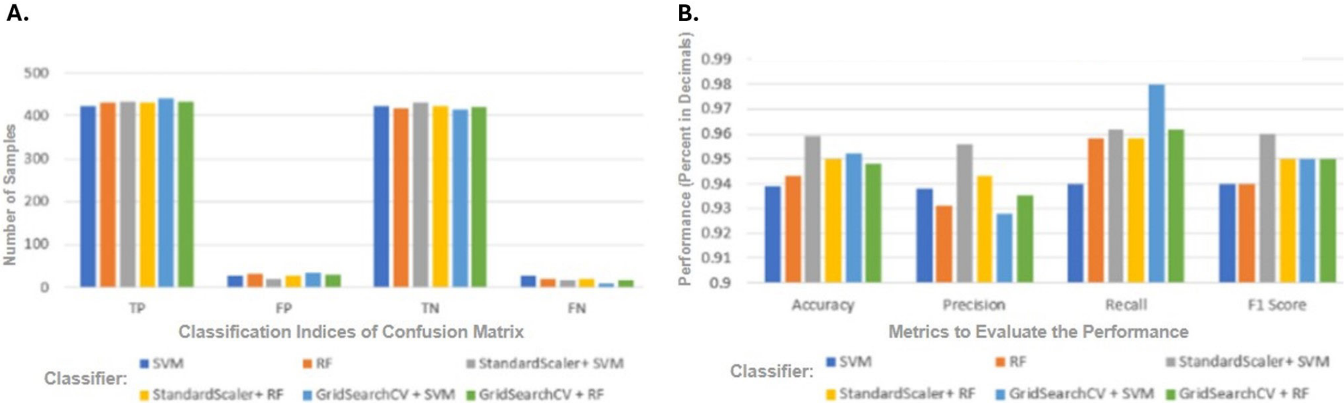

Figure 1: Confusion matrix and performance evaluation results for the six algorithms following data normalization preprocessing and grid search hyperparameter tuning. (A) Bar graph showing the number of samples for true positive (TP), false positive (FP), true negative (TN), and false negative (FN) of the confusion matrix for all algorithms. **(B)** Bar graph showing the accuracy, precision, recall, and F1 score for all algorithms. A model was built for each combination of the classifiers (SVM or RF) with StandardScaler and GridSearchCV.

F1 score (23-25). The confusion matrix was based on four classification indices: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Comparing the default SVM classifier with the RF classifier in the confusion matrix (**Table 1 and Figure 1A**), the SVM classifier returned 423 TPs, which meant that 423 patients were correctly identified as having tumors, and the RF gave TP 431. Furthermore, the SVM gave FN 27 and RF gave FN 19, which meant that the classifiers incorrectly predicted that those many patients had no tumor. The FP of the SVM was 28 and that of RF was 32, which meant that the SVM had lower false prediction of the tumor disease as compared with the RF. The TN of the SVM was 422 and RF gave TN 418, which meant that the classifiers correctly classified that these patients did not have the tumor disease.

Evaluating the performance of the SVM classifiers and RF classifiers for all six algorithms (**Table 1 and Figure 1B**) confirmed that all six SVM and RF models worked well. SVM with StandardScaler gave the best accuracy of 0.959, the best precision of 0.956, and the best F1 score of 0.96. For SVM, scaling was highly recommended. SVM with GridSearchCV gave the highest TP and lowest FN values, which made its recall the highest at 0.98. The GridSearchCV took about three hours to complete.

Comparison of different CV methods using SVM and RF

Since the train_test_split() we implemented might have introduced some randomness, we used cross-validation methods for train-test split and tested the generalizability of each ML algorithm. We then compared the different cross-validation methods, K-Fold CV, StratifiedKFold CV, nested K-Fold CV, and RepeatedStratifiedKFold.

K-Fold CV train-test split was performed for the SVM and RF classifier with splits 5 and 6 for algorithm #7 and #8, respectively. Performance was evaluated and compared for each fold of the SVM classifier (linear kernel and C=1, **Table 2**) and the RF classifier (n_estimators=500, max_depth=30, **Table 2**). The average accuracy of the SVM classifier was 0.9697 ± 0.0076 (splits 5) and 0.971 ± 0.0095 (splits 6), and that of the RF classifier was 0.96 ± 0.01 for both splits 5 and 6.

StratifiedKFold CV train-test split was performed with splits 5 for different SVM classifiers for algorithm #9. The accuracy of each fold was recorded and compared for the SVM models (**Table 3**). The average accuracy of the default SVM classifier was 0.959 ± 0.02, and that of the SVM classifier (linear kernel and C=1) was 0.961 ± 0.035. StratifiedKFold cross-validation train-test split was performed with splits 5 for different RF classifiers with n_estimators=100 and max_depth=10, 15, 20, 30, and 40 for algorithm #10, and max_depth=30 and n_estimators=30, 50, 100, 150, 250, and 500 for algorithm #11. The accuracy of each fold for the StratifiedKFold CV with different RF classifiers was recorded and their distributions were shown (**Figure 2**).

		SVM Classifier (kernel ='linear', C=1)				RF Classifier (n_estimators=500, max_depth=30)			
	Fold Number	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Kfold CV (n_splits= 5)	1	0.97	0.97	0.97	0.97	0.9667	0.9662	0.9673	0.9666
	2	0.982	0.982	0.982	0.982	0.9767	0.9772	0.9767	0.9767
	3	0.97	0.97	0.97	0.97	0.945	0.945	0.9449	0.9449
	4	0.962	0.961	0.962	0.962	0.9517	0.9545	0.9503	0.9514
	5	0.965	0.965	0.964	0.965	0.9617	0.9626	0.9608	0.9615
Kfold CV (n_splits=6)	1	0.968	0.968	0.968	0.968	0.962	0.9616	0.9626	0.9619
	2	0.984	0.984	0.984	0.984	0.97	0.9705	0.9704	0.97
	3	0.978	0.978	0.978	0.978	0.97	0.9699	0.97	0.97
	4	0.964	0.964	0.964	0.964	0.956	0.9571	0.9563	0.956
	5	0.958	0.958	0.957	0.958	0.954	0.9559	0.9523	0.9536
	6	0.974	0.974	0.974	0.974	0.968	0.9681	0.9677	0.9679

Table 2: Performance evaluation of KFold CV for SVM and RF classifiers with different splits.

		SVM Classifier default	SVM Classifier (kernel ='linear', C=1)
	Fold Number	Accuracy	Accuracy
StratifiedKFold CV (n_splits=5)	1	0.985	0.995
	2	0.9267	0.8933
	3	0.955	0.9717
	4	0.9533	0.9617
	5	0.975	0.9817

Table 3: Accuracy of StratifiedKFold CV for different SVM classifiers.

We conducted a one-way analysis of variance (ANOVA) test to determine whether there were any statistically significant differences between the mean accuracy of multiple RF classifiers with different max_depth and n_estimators. For multiple RF classifiers with different max_depth, the one-way ANOVA test reported a p-value=0.27845, which is larger than 0.05. This meant that the deviation from the null hypothesis was not statistically significant, and the null hypothesis was not rejected. For multiple RF classifiers with different n_estimators, the one-way ANOVA test reported a p-value=0.89401, which is larger than 0.05. This meant that the deviation from the null hypothesis was not statistically significant, and the null hypothesis was not rejected.

For nested K-Fold CV, outer K-Fold CV (splits 10) train-test split was performed, with hyperparameters found through GridSearch in space {'C': [0.1,1, 10, 100], 'gamma': [1,0.1,0.01,0.001]} for SVM and space {'n_estimators': [10, 100, 500], 'max_depth': [4, 10, 20]} for RF with inner K-Fold CV (splits 3) for algorithms #11 and #12, respectively. The average accuracy scores of the SVM and RF classifiers collected for all 10 outer folds were 0.969 ± 0.006 and 0.965 ± 0.004, respectively (**Table 4**).

For RepeatedStratifiedKFold CV, GridSearchCV was performed. For the SVM classifiers, algorithm #13 was done in space {'kernel': ['poly', 'rbf', 'sigmoid'], 'C': [0.1, 1, 10, 100]} with RepeatedStratifiedKFold cross-validation (splits 5 and repeats 3). For the RF classifiers, algorithm #14 was done in space {'n_estimatorsl': [10, 100, 500], 'max_depth': [4, 10, 20]} with RepeatedStratifiedKFold cross-validation (splits 10 and repeats 3). The average accuracy for all grid combinations was compared (**Figure 3**). The best parameter combination found for the SVM classifier was {'C': 10, 'kernel': 'rbf'} with an average accuracy of 0.9772 ± 0.007, and that for the RF classifier was {'max_depth': 20, 'n_estimators': 500} with an average accuracy of 0.9657 ± 0.0103.

Overall, all four methods worked well for both SVM and RF classifiers on Br35H. (**Table 4 and Figure 4**). Train-test splits with CV could validate the classifiers more fairly with less randomness than the simple train-test split. The accuracy of SVM and RF classifiers with 70% train-test split was lower

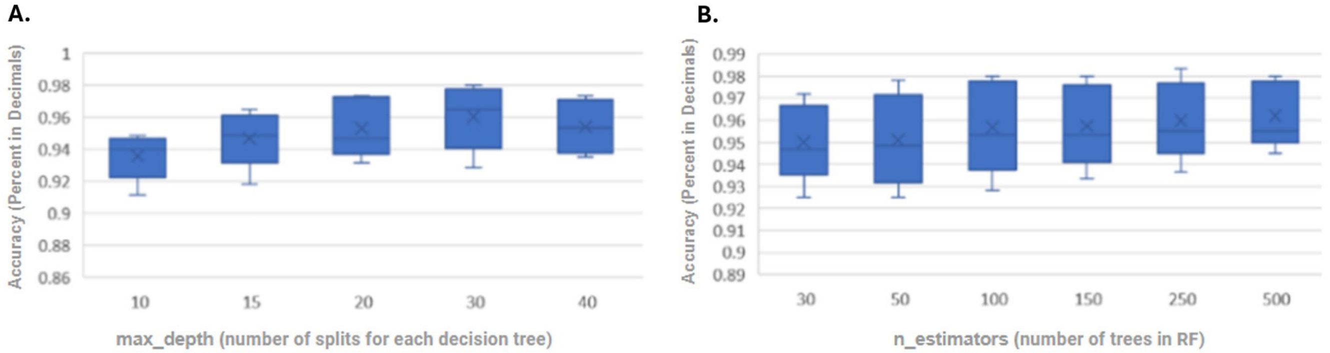

Figure 2: Accuracy of StratifiedKFold CV for RF classifiers with different combinations of parameters varied. (A) Box and whisker plot showing the accuracy of each fold for StratifiedKFold CV (splits 5) train-test split with RF classifiers where n_estimators=100 and indicated max_depth values. **(B)** Box and whisker plot showing the accuracy of each fold for StratifiedKFold CV (splits 5) train-test split with RF classifiers where max_depth=30 and indicated n_estimators values. The range is represented by the whiskers of the boxplot. For the one-way ANOVA test, the p-value>0.05 for both **(A)** and **(B)**. A model was built for each combination of StratifiedKFold CV train-test split with RF classifiers of different hyperparameters.

than those with CV. This might be caused by the randomness of the 70% train-test split.

RepeatedStratifiedKFold CV (splits 5 repeats 3) for SVM classifier demonstrated the best accuracy of 97.72% among all different CV methods combined with either the SVM or RF classifier, with the best parameter combination of kernel 'rbf' and C=10 found through GridSearchCV (**Table 4 and Figure 4**).

RepeatedStratifiedKFold CV (splits 10 repeats 3) yielded the best accuracy among the CV methods for the RF classifiers, with the best parameter combination of n_estimators=500 and max_depth=20 found through GridSearchCV (**Table 4 and Figure 4**).

Nested K-Fold CV was the most time-consuming method for both SVM and RF classifiers. K-Fold CV (splits 5) was a simple and fast method with good accuracy for the SVM classifier (linear kernel and C=1), while StratifiedKFold CV (splits 5) worked a little better than KFold (splits 5) as a simple and fast method for RF (n_estimator=500 and max_depth=30, **Table 4 and Figure 4**).

Ensemble model using out-of-fold predictions

To further improve the classification efficiency while maintaining good accuracy, we developed an ensemble model using out-of-fold predictions. train_test_split() of the Br35H dataset was performed using 70% of the samples for training and the remaining 30% for validation. SVM and RF classifiers were used as the base models and evaluated using K-Fold (splits 5) CV on the training dataset, and all out-of-fold predictions were retained. LogisticRegression with a liblinear (Library for Large Linear Classification) solver was used as a meta classifier, which took the input data for each sample and the out-of-fold predictions as inputs, showing how to best combine and correct the predictions. Performance evaluation was done using the validation set for the meta model, final SVM model, and final RF model, which were all developed using the entire training set. The accuracy of the SVM model was 0.952, that of the RF model was 0.960, and that of the meta model was 0.980. It took about 10 minutes for the ensemble model fitting to be completed.

Algorithm	Train-Test Split	The proposed model	Hyperparameters	Accuracy
#3	70% split	StandardScaler+ SVM	default	0.959
#4	70% split	StandardScaler+ RF	default	0.95
#7	K-Fold CV (splits 5)	SVM	kernel='linear', C=1	0.9697 ± 0.0076
#8	K-Fold CV (splits 5)	RF	n_estimator=500, max_depth =30	0.96 ± 0.01
#9	StraitifiedKFold CV (splits 5)	SVM	kernel='linear', C=1	0.961 ± 0.035
#10	StraitifiedKFold CV (splits 5)	RF	n_estimator=500, max_depth =30	0.962 ± 0.013
#11	Nested K-Fold CV (splits: outer 10 inner 3)	SVM	GridSearchCV	0.969 ± 0.006
#12	Nested K-Fold CV (splits: outer 10 inner 3)	RF	GridSearchCV	0.965 ± 0.004
#13	RepeatedStratifiedKFold CV (splits 5 repeats 3)	SVM	GridSearchCV, best: kernel='rbf', C=10	0.9772 ± 0.007
#14	RepeatedStratifiedKFold CV (splits 10 repeats 3)	RF	GridSearchCV, best:n_estimators=500, max_depth =20	0.9657 ± 0.0103

Table 4: Average accuracy of different train-test split methods for SVM and RF classifiers.

DISCUSSION

The diagnosis of brain tumors is essential in directing clinical treatment. An automated ML brain tumor detection approach is a non-invasive and cost-effective diagnostic tool, making detection easier.

In this study, we primarily focused on the SVM and RF classifiers. We conducted a series of experiments to understand the effect of normalization preprocessing, grid search, and different types of cross-validation techniques (CV) by using SVM and RF, which are two of the most popular and widely used ML algorithms in brain tumor classification, detection, and segmentation. For example, Vankdothu et al. (26) and Senan et al. (27) have shown that it is possible to develop models and apply ML to accurately identify and classify brain tumors.

Many existing studies reported that SVM and RF could achieve relatively high accuracy compared with other ML algorithms, especially when the model is trained based on texture features (28, 29). In addition, recent studies proposed ensemble methods that combine SVM or RF together with deep learning (DL) methods. Plenty of existing work has already proved that DL methods can work well on image data (30, 31). Then, in the existing work, researchers proposed to extract informative features by DL methods based on medical image data and use ML algorithms to train models based on features which were generated by DL methods.

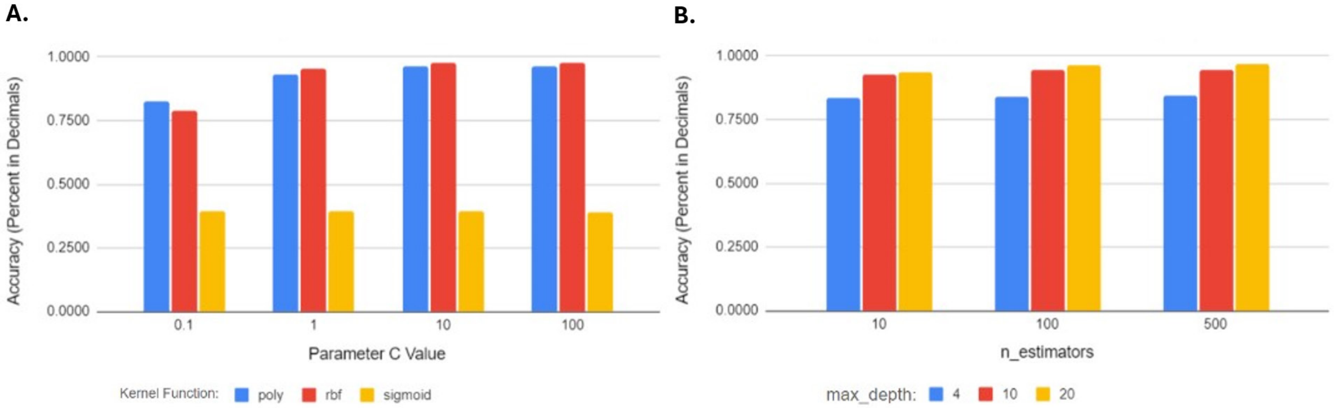

Figure 3: Average accuracy of RepeatedStratifiedKFold CV for SVM and RF classifiers with different combinations of parameters varied. (A) Bar chart showing the average accuracy of RepeatedStratifiedKFold CV (splits 5 and repeats 3) train-test split for SVM classifiers with different combinations of parameter C values and kernel. **(B)** Bar chart showing the average accuracy of RepeatedStratifiedKFold CV (splits 10 and repeats 3) train-test split for RF classifiers with different combinations of n_estimators and max_depth values. A model was built for each combination of RepeatedStratifiedKFold CV train-test split with SVM or RF classifiers of different hyperparameters.

Since the tasks and datasets used in experiments differ, it can be challenging to directly compare results across different ML papers. Among recent studies using SVM and RF, similar to the task in this study, Vankdothu et al. investigated CT-based brain tumor segmentation with preprocessing techniques to improve the image's clarity and classifiers applied to classify it as normal or abnormal (26). The accuracy reported by this study is slightly lower than that in our study. The potential reasons include the different data type used as input data (CT vs. MRI) and the different preprocessing steps (implementing a segmented step before classification) (26).

Our results demonstrate that cross-validation is a better tool for train-test split than simple train-test. One potential problem with a simple train-test split is that a portion of the data is set aside for testing, meaning the model does not use this data during training, which can lead to issues with generalization. Since the data is sampled at random, these splits may be skewed and fail to represent the whole dataset properly. Random train-test split may induce randomness to the model performance, especially when the sample size is small.

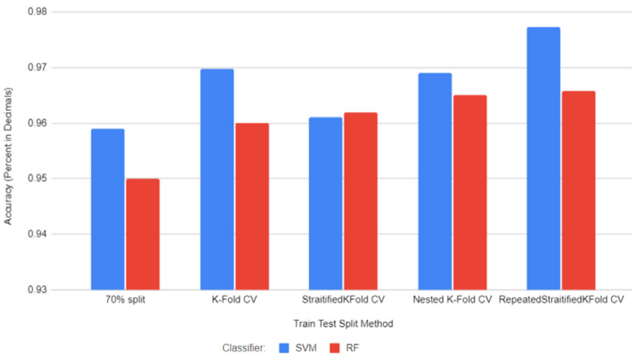

Figure 4: Summary and comparison of performance for SVM and RF classifiers with different train-test split using average accuracy. Bar graph showing the average accuracy of SVM and RF classifiers with 70% split, K-Fold CV, StratifiedKFold CV, Nested K-fold CV, and RepeatedStrafitiedFold CV train-test split methods. A model was built for each combination of the classifiers (SVM or RF) with the different train-test split methods.

This study compared the performance of SVM or RF classifiers and their combination with normalization preprocessing, grid search, and different types of cross-validation techniques methods for the detection of brain tumors using MRI images. These comparisons allowed us to gain some knowledge about different classifiers' performance, model selection, and parameter tuning for brain tumor related domains. Our experimental results illustrated that performing data preprocessing, refining tuning parameters, and using proper cross-validation were essential in building robust ML models. In addition, the proposed ensemble model combined K-Fold CV with the SVM and RF classifiers and made the traditional ML algorithms more robust. For predicting the onset of brain tumors, our study verified that all our classification algorithms had an accuracy higher than 90%. Better accuracy and lower computational cost could be achieved with the ensemble model.

This study has several limitations. First, our performance evaluation was based on CV. While CV is a valuable technique for hyperparameter tuning, model selection, and providing an initial estimate of the model's performance, it should not replace the use of an independent test dataset. Both have their roles in the ML pipeline, with CV helping make improvements during development and an independent test set providing the final assessment of a model's generalizability on unseen data. Further research is imperative to validate the generalizability of each model on an independent test set. Second, it is of great importance to compare traditional ML models like SVM and RF with DL models. DL models are known for their ability to capture complex patterns and representations in data. Comparing traditional models with DL can help establish a performance benchmark. This benchmark can provide an assessment whether the added complexity and computational requirements of deep learning are justified by improved performance. Moreover, comparing models helps determine which approach is most appropriate for the specific problem. Traditional models can perform well when the dataset is small or when the data has simple patterns, while deep learning models can excel when handling large-scale, complex data. Future studies could include trying neural networks for the detection of tumors for better accuracy. We will extend the project to predict different types of brain tumors which can be classified into more than two categories. When a variety of image modalities are available, we will use them to explore the segmentation of the tumor from normal brain tissue.

MATERIALS AND METHODS

Our classification was carried out with Jupyter Notebook 6.4.8 open-source software and a PC running a 64-bit Windows 10 operating system with 11th Gen Intel(R) Core(TM) i7-1185G7 @ 3.00GHz processor and 32.0 GB RAM.

Some of the key parameters in SVM (32-34) are gamma, C, and kernel. C controls the cost of miscalculations. Kernels are a set of mathematical functions including linear, RBF (Radial Basis Function), and polynomial kernel. A kernel trick is used to transform the input into a higher-dimensional space, making it easier to segregate the points.

Random forest (RF) (35) is an ensemble classifier formed by the fusion of multiple decision trees that are not correlated. The result is calculated based on the majority voting method. As a tree grows deep, it starts to over-fit, having low bias and high variance.

In K-Fold Cross-Validation (36), data is split into k different subsets (or folds). k-1 subsets are used to train the model, and the last subset is used as test data. Then, the model is averaged against each of the folds and finalized. During k-fold cross-validation, predictions are made on test sets composed of data not used to train the model, a type of out-of-sample prediction (37).

For Br35H, the original image shape was (316, 270,3), the resized image shape was (150, 150, 3), and the flattened data was (3000, 67500).

Support vector machine (SVM) and random forest (RF) classifiers

Train_test_split() in the scikit-learn library was used to generate training and testing sets, with 70% of the dataset used for training the models and 30% used as a testing set. Default models svm.SVC() and RandomForestClassifier() were used.

Normalization preprocessing was performed for algorithms #3 and #4 using sklearn.preprocessing.StandardScaler(). The fit method calculated the mean and standard deviation of each feature present in our training data, and the transform method performed standardization by centering and scaling independently on each feature on our training and testing data.

The idea of creating a 'grid' of parameters and trying out possible combinations is called a Gridsearch. GridSearchCV is a library function that is a member of sklearn's model_selection package. This function iterates through predefined hyperparameters and fits the parametors of the model to the training set. At the end, the best hyperparameters are selected based on model performance.

GridSearchCV was performed for algorithms #5 and #6. For algorithm #5, the parameter grid was {'C': [0.1,1, 10, 100], 'gamma': [1,0.1,0.01,0.001]}, and the SVM classifier was fit to the training data. The same loop was run with cross-validation, fitting 5 folds for each of the 16 candidates, totaling 80 fits, which took about three hours to complete. For algorithm #6, the parameter grid was {'n_estimators': [10, 100, 500], 'max_depth': [4, 10, 20]}, and the RF classifier was fit to the training data. The same loop was run with cross-validation, fitting 5 folds for each of 9 candidates, totaling 45 fits.

A confusion matrix shows the number and types of errors made by the classifier. The accuracy of the classifier is defined in Equation (1) as the ratio of the total number of correct predictions to the total number of predictions. The precision of the classifier is defined in Equation (2) as the ratio of correct positive predictions to the total positive predictions of brain tumor cases represented. Recall of the classifier is defined in Equation (3) as the ratio of correctly predicted brain tumor cases to overall correct predictions. F1 score is the measure of the harmonic mean between precision and recall of classifiers as shown in Equation (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ Score = \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Different cross-validation (CV) methods

A one-way analysis of variance (ANOVA) test was used to analyze if the difference between the averages of five or six groups with different max_depth or n_estimators, respectively, was statistically significant. To conduct the statistical test, a standard p-value of 0.05 was used and the null hypothesis and alternative hypothesis were defined as follows:

Null hypothesis: There are no significant differences among the average accuracy of RF classifiers with different max_depth for **Figure 2A** (or n_estimators for **Figure 2B**), meaning that RF classifiers with different max_depth for **Figure 2A** (or n_estimators for **Figure 2B**) have the same average accuracy.

Alternative hypothesis: The RF classifiers with different max_depth for **Figure 2A** (or n_estimators for **Figure 2B**) have different average accuracy.

For nested K-Fold CV, it took about a day to complete the SVM model fitting. For RepeatedStratifiedKFold CV, it took about three hours to complete the SVM model fitting.

Ensemble model using out-of-fold predictions

Out-of-fold predictions were used to train the meta-model. First, data was split into the training and validation datasets. The training dataset was used to fit the base models and meta-model, while the validation dataset was used to evaluate the meta-model and base models. Then, the k-fold cross-validation procedure was performed on each base model, and the out-of-fold predictions were collected. Next, a meta-model was trained on the predictions made by the base models. Finally, each base model was trained on the entire training dataset to produce the final models. The performance of this ensemble was evaluated on a separate validation dataset.

Received: April 16, 2023
Accepted: January 27, 2024
Published: April 15, 2025

REFERENCES

1. Buffalo, Elizabeth A., et al. "From basic brain research

to treating human brain disorders." Proceedings of the National Academy of Sciences, vol. 116, no. 52, 26 Dec. 2019, pp. 26167-72, https://doi.org/10.1073/pnas.1919895116.

2. Herculano-Houzel, Suzana. "The human brain in numbers: a linearly scaled-up primate brain." *Frontiers in human neuroscience*, vol 3: 31, 9 Nov. 2009, https://doi.org/10.3389/neuro.09.031.2009.
3. DeAngelis, Lisa M. "Brain Tumors." *New England Journal of Medicine*, vol. 344, no. 2, 11 Jan. 2001, pp. 114–123, https://doi.org/10.1056/NEJM200101113440207.
4. Ezhilvani, A. Mary, et al. "A Survey on classification techniques, used for detection of brain tumor." *IJCRT*, vol. 8, 2 Feb. 2020, pp. 413-416.
5. Gupta, Manali, et al. "Analysis of machine learning algorithms in brain tumour prediction." *Journal of Physics: Conference Series*, vol. 2070, no. 1, Nov. 2021, https://doi.org/10.1088/1742-6596/2070/1/012090.
6. Gu, Xiaoqing, et al. "Brain Tumor MR Image Classification Using Convolutional Dictionary Learning With Local Constraint." *Front Neurosci.*, vol. 15, 27 May 2021, https://doi.org/10.3389/fnins.2021.679847.
7. Ballester, Leomar Y., et al. "Practical Surgical Neuropathology: A Diagnostic Approach." *Elsevier Health Sciences*, edited by Arie Perry and Daniel J. Brat, 2nd ed., 2018, https://doi.org/10.1016/C2014-0-01404-9.
8. Urbańska, Kaja, et al. "Glioblastoma multiforme an overview." *Contemporary Oncology (Poznan, Poland)*, vol. 18, 4 Aug. 2014, pp. 307-12, https://doi.org/10.5114/wo.2014.40559.
9. "Brain Tumors." *National Brain Tumor Society.* https://braintumor.org/brain-tumors/. Accessed 10 Aug. 2024.
10. Ellor, Susan V., et al. "Glioblastoma: Background, standard treatment paradigms, and supportive care considerations." *J. Law Med. Ethics,* vol. *42(2)*, 2014, pp. 171–182, https://doi.org/10.1111/jlme.12133.
11. Gao, Peiyi, et al. "Development and Validation of a Deep Learning Model for Brain Tumor Diagnosis and Classification Using Magnetic Resonance Imaging." *JAMA Network Open,* 8 Aug. 2022, https://doi.org/10.1001/jamanetworkopen.2022.25608.
12. Díaz-Pernas, Francisco J., et al. "A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network." *Healthcare,* vol 9, 2 Feb. 2021, https://doi.org/10.3390/healthcare9020153.
13. Younis, Ayesha, et al. "Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches." *Applied Sciences,* vol 12, 20 July 2022, https://doi.org/10.3390/app12147282.
14. J. Cheng. Brain tumor dataset. vol. 5, April 2017, https://doi.org/10.6084/m9.figshare.1512427.v8.
15. Agrawal, Tarun, et al. "Multi-scale feature scaling in deep neural network for the brain tumour classification in MRI images." *International Journal of Imaging Systems and Technology*, 25 Aug. 2023, https://doi.org/10.1002/ima.22956.
16. Dheepak, G., et al. "A Novel Approach for Feature Extraction and Stacking with GLCM in Advanced MRI Brain Tumor Classification." *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 23-25 Aug. 2023, https://doi.org/10.1109/icaiss58487.2023.10250508.
17. Mazlin, Muhammad Syukri, et al. "Saliency-based variational active contour model for image with intensity inhomogeneity." *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 32, no. 1, Oct. 2023, pp. 206-215, http://doi.org/10.11591/ijeecs.v32.i1.pp206-215.
18. Waskita, A. A., et al. "EfficientNetV2 based for MRI brain tumor image classification." *2023 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 4-5 Oct. 2023, https://doi.org/10.1109/IC3INA60834.2023.10285782.
19. Gorade, Vandan, et al. "SynergyNet: Bridging the Gap between Discrete and Continuous Representations for Precise Medical Image Segmentation." 26 Oct. 2023, https://doi.org/10.48550/arXiv.2310.17764.
20. Garg, Ginni, et al. "Brain Tumor Detection and Classification based on Hybrid Ensemble Classifier."1 Jan. 2021, https://doi.org/10.48550/arXiv.2101.00216.
21. Bosch, Anna, et al. "Image classification using random forests and ferns." *IEEE 11th international conference on computer vision*, 14-21 Oct. 2007, https://doi.org/10.1109/ICCV.2007.4409066.
22. Ko, B. C., et al. "Cell image classification based on ensemble features and random forest." Electronics Letters, vol. 47, no. 11, June 2011, pp. 638-639, https://doi.org/10.1049/el.2011.0831.
23. Kolla, Morarjee, et al. "CNN-Based Brain Tumor Detection Model Using Local Binary Pattern and Multilayered SVM Classifier." *Computational Intelligence and Neuroscience,* 27 June 2022, https://doi.org/10.1155/2022/9015778.
24. Nanmaran, R., et al. "Investigating the Role of Image Fusion in Brain Tumor Classification Models Based on Machine Learning Algorithm for Personalized Medicine." *Computational and Mathematical Methods in Medicine,* 7 Feb. 2022, https://doi.org/10.1155/2022/7137524.
25. Faraz, Nuzhat, et al. "Data Mining Approach for Detection and Classification of Brain Tumor." *Mehran University Research Journal of Engineering and Technology,* vol. 41, no. 1, Jan. 2022, pp. 53–64, https://doi.org/10.22581/muet1982.2201.06.
26. Vankdothu, Ramdas, et al. "Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning." *Measurement: Sensors,* vol. 24, Dec. 2022, https://doi.org/10.1016/j.measen.2022.100440.
27. Senan, Ebrahim Mohammed, et al. "Early Diagnosis of Brain Tumour MRI Images Using Hybrid Techniques between Deep and Machine Learning." *Computational and Mathematical Methods in Medicine,* 18 May 2022, https://doi.org/10.1155/2022/8330833.
28. Wang, Lujia, et al. "Quantifying intra-tumoral genetic heterogeneity of glioblastoma toward precision medicine using MRI and a data-inclusive machine learning algorithm." *PloS one*, vol. 19, no. 4, 3 April 2024, https://doi.org/10.1371/journal.pone.0299267.
29. Wang, Lujia, et al. "A Novel Hybrid Ordinal Learning Model With Health Care Application." IEEE Transactions on Automation Science and Engineering, Jan. 2024, https://doi.org/10.1109/TASE.2024.3350894.
30. Wang, Hairong, et al. "Biologically-informed deep neural networks provide quantitative assessment of intratumoral heterogeneity in post-treatment glioblastoma." *Research*

Square, 27 March 2024, https://doi.org/10.21203/r3.3.rs-3891425/v1.

31. Shen, Dinggang, et al. “Deep learning in medical image analysis.” *Annual review of biomedical engineering*, vol.19, no. 1, 9 March 2017, pp. 221-248, https://doi.org/10.1146/annurev-bioeng-071516-044442.
32. Thai, Le Hoang, et al. “Image Classification using Support Vector Machine and Artificial Neural Network.” *International Journal of Information Technology and Computer Science,* vol. 4, no. 5, 8 May 2012, pp. 32–38, https://doi.org/10.5815/ijitcs.2012.05.05.
33. Latif, Ghazanfar, et al. “Glioma Tumors’ Classification Using Deep-Neural-Network-Based Features with SVM Classifier.” *Diagnostics*, vol. 12, 18 April 2022, https://doi.org/10.3390/diagnostics12041018.
34. “Support Vector Machine Algorithm.” *Javatpoint.* www.javatpoint.com/machine-learning-support-vector-machine-algorithm. Accessed 15 Feb. 2023.
35. Amiri, Samya, et al. “Deep random forest-based learning transfer to SVM for brain tumor segmentation.” *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 21-23 March 2016, https://doi.org/10.1109/ATSIP.2016.7523095.
36. Anello, Eugenia. “K-Fold Cross Validation for Machine Learning Models.” Medium, 17 Feb. 2021, pub.towardsai.net/k-fold-cross-validation-for-machine-learning-models-918f6ccfd6d. Accessed 15 Feb. 2023.
37. Raschka, Sebastian. “StackingCVClassifier: Stacking with cross-validation.” *Mlxtend, 2023,* rasbt.github.io/mlxtend/user_guide/classifier/StackingCVClassifier/. Accessed 15 Feb. 2023.