Comparative Analysis of Ensemble Learning Techniques for Brain Tumor Classification

Rehman Sharif [1], Muhammad Azam[2], Arshad Ali[3*], Muhammad Usman Hashmi [4], Muhammad Uzair[5]
[1]Department of Mathematics & Physics the Superior University Lahore, 54000, Pakistan
[2]Department of Computer Science & Information Technology, The Superior University Lahore, 54000, Pakistan
[3]Faculty of Computer and Information Systems, Islamic University of Madinah, Al Madinah Al Munawarah, Saudi Arabia
[4]Department of Mathematics & Physics, The Superior University Lahore, 54000, Pakistan
[5]Faculty of Engineering, Islamic University of Madinah, Al Madinah Al Munawarah, Saudi Arabia
E-mail: rehman.sharif@superior.edu.pk, pmit@superior.edu.pk, a.ali@iu.edu.sa, usman.hashmi@superior.edu.pk, muzair@iu.edu.sa
[*]Corresponding author

Keywords: classification, brain tumor classification, random forest, voting, stacking, ensemble learning

Received: July 18, 2024

For the evaluation of our work, we used a brain tumor MRI dataset obtained from Kaggle for the experimental analysis. Our study investigates the performance of ensemble learning techniques for brain tumor classification. This study focuses on comparing the efficacy of homogenous ensemble classifiers, exemplified by the Random Forest (RF) algorithm, against heterogeneous ensemble classifiers like Voting and Stacking, this study embarks on a thorough evaluation journey. Our evaluation is not limited to accuracy measures only; instead, it surrounds recall, ROC AUC, precision, and F1-score, for better assessment of classifiers' performance. Our findings demonstrate that the Random Forest classifier achieved an accuracy of 99%, an F1 score of 0.99%, and AUC of 0.99%, outperforming other ensemble techniques like Voting and Stacking classifiers. Building upon an observed assessment performed on an appropriately selected brain tumor dataset, we provide solid empirical support demonstrating that RF not only performs better than base classifiers but also outperforms the heterogeneous ensemble methods in terms of many different performance measures. Furthermore, we discuss the specific reason that makes RF outperform other algorithms in this dataset and discuss the robustness and flexibility of this method. By unscrambling these insights, this paper addresses to fill gaps in the existing knowledge regarding the ensemble learning techniques in medical imaging, particularly for brain tumor classification diagnostics.

Povzetek: Predstavljena je primerjalna analiza tehnik ansambelskega učenja za klasifikacijo možganskih tumorjev z uporabo podatkov MRI, posebej naključni gozdovi z dodatnimi tehnikami glasovanja. Najboljše rezultate so dosegli naključni gozdovi.

1 Introduction

Cerebral tumors or brain tumors present a great concern and complexity in the diagnosis as well as treatment of diseases since they are complex structures that are very difficult to diagnose and treatment. Classifying of brain tumors is a critical step in the diagnosis of the specific type of tumor and in formulating the right treatment plan and prognosis for the patient. Previously, another conventional and most common classification of brain tumors was histological, which is invasive, time-consuming, and based on judgment made based on tissue samples.

The application of the latest technology in computer-aided medical imaging as well as artificial intelligence in the form of the application of machine learning algorithms makes computer-aided diagnostic systems very efficient in a way that they give precise classification of brain tumors. These methods rely on feature extraction from tumor exams including Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) images to assist in tumor classification [1]. Out of the various kinds of machine learning, ensemble learning is one type that enhances the performance and credibility of a classification model.

The time for which a given patient is likely to survive after obtaining a brain tumor is limited especially if the condition is in the last stage. Therefore, it may help facilitate the identification of the most suitable course of action for dealing with tumors thus saving many lives. Neurologic examination and imaging remain the key diagnoses for brain tumors, particularly MRI and CT [2]. It must be noted that there is hardly any other modality that describes the structure of the brain with the accuracy that MRI imaging does. However, medical image segmentation is not an easy task, and it involves several problems like noise, ill-defined edges, low spatial resolution, varying intensity, low contrast, partial volume effect and variations in object shape, acquisition artifacts in the data, and unavailability of comprehensive atlases to incorporate the variations in the anatomy and the internal structures [3].

The most common application of the head and neck MRI is to determine the existence/progression of the tumor. This information is mainly applied for tumor recognition and treatment. An MRI scan also offers more detail than an MRI or ultrasound image does a CT scan. MRI plays a vital role in providing detailed information about the location and structural integrity of the regions in the brain throughout the identification of irregularities in the tissue composition. Since medical pictures can be easily scanned and transferred to computers for analysis, scholars have suggested the usage of automated methods for the identification of brain tumors from brain MRI data. In contrast, used more recently, Neural Networks (NN) and Support Vector Machines (SVM) have been reported to have high performance [4].

Based on this, a new automated approach is introduced that can recognize the following many types of brain tumors. It creates a challenge as opposed to other methods. The following are some of the important points in our case:

- A new classification system is outlined here to distinguish three classes of brain tumors.
- The study focuses on the impacts of using full MRI images and leads to enhanced performance on multi-class categorization.
- The proposed technique marks brain tumors into several classes, which is different from the costly binary classification methods used earlier [5].

1.1 Modalities of imaging brain tumor

Various imaging technologies are used in studying brain tumors and they include SPECT imaging, CT imaging, MRI, and PET imaging. However, CT and MRI are the most common techniques employed overall because CT is highly available and MRI offers clear images of both healthy tissues and diseases [6]. These brain tumors are categorized into four categories according to the grade that the tumor is given.

a. **Tumors of Grade I**
These tumors are not invasive and are typically characterized by slow growth and progression. These factors are also crucial and related to a high probability of enhanced organization and are amenable to surgical resection. An example of such a tumor is pilocytic astrocytoma.

b. **Tumors of Grade II**
The tumors of Grade II can proliferate to the nearby tissues and advance to advanced grades. Additionally, these tumors exhibit gradual growth over time. These tumors may still be discovered despite the patient undergoing therapy. An oligodendroglioma tumor is a kind of neoplasm that grows gradually over time.

c. **Tumors of Grade III**
These tumors have a more rapid development rate compared to grade II malignancies and have the potential to invade nearby tissues. These malignancies need adjuvant chemotherapy or radiation since surgery alone would be inadequate for their treatment. Aden squamous astrocytoma is a specific kind of tumor.

d. **Tumors of Grade IV**
The tumors described in this category are the most dangerous and have a high chance of proliferating in their aggression. They may employ blood arteries to enhance their expansion rates. One of these cancers is glioblastoma multiforme [7], [8].

2 Literature review

There has been development of so many applications of machine learning in the field of diagnosis of medical conditions and healthcare. In the context of the current research, it has also been evidenced that regardless of the applicability of multi-fractal analysis in identifying brain cancers, especially using MRI, the method lacks frequent use in this area. Primarily, the MRI data is employed to train and evaluate the normal machine learning algorithms and methods. Over the last few years, there have been several techniques established based on deep learning for the identification of brain cancer and the classification of the varied types as highlighted by Neelum Noreen Et al [9]. However, binary classification of classes in this case is not very difficult since the definition of tumor form and texture does not present much difficulty. Another work relating to the detection of brain tumors is provided by Sharif et al. [10] through fuzzy logic and the U-NET CNN.

The combination of the techniques of contrast enhancement and fuzzy logic-based edge detection with the U-NET CNN classification methodology. The source pictures also get pre-processed mainly through the contrast enhancement procedure before passing through the system. After this, the contrast-enhanced images go through a fuzzy logic method for edge detection to identify the edge present in the work. Finally, at both coarse and detail levels, the employed method is a dual tree-complex wavelet transform. These features are derived from the reconstructed and deconstructed sub-band images which in turn are classified using the U-NET CNN classification algorithm which is uniquely capable of distinguishing a meningioma and a non-meningioma in brain imaging. This method of gender prediction was then compared to new algorithms and found to be 98.59% accurate.

Sobhaninia et al. [11] have listed some of the methods with distinctive techniques of diagnosis of the brain tumor through MR imaging as MT imaging, DW imaging, BOLD imaging, SW imaging, and FA imaging. To smoothly subdivide it even further, they used 3D CNNs, SVMs, and multi-class SVMs. The proposed DL technique generated very high performance that was capable of presenting a strong approach to classify brain tumors and segment them from the normal tissue in contrast to the other usual classifiers of ML. This has been propounded from another study done by HASNAIN ALI SHAH et al. [12]; they created five clinical data sets for numerous categories of clinical diagnoses. Additionally, MRI data were pre-classified to enhance the identification of various kinds of Gliomas, using a Convolutional Neural Network (CCN) that was built using transfer learning. The proposed CNN model was compared with six further classifiers and they were defined below; Decision tree, Naïve Bayes, Linear Discriminant, K-nearest neighbor, and Support Vector Machine classifiers. The study is carried out on five various datasets of brain tumors for multiclass classification and leads to the comparison of the performance indicators of the introduced CNN-based (DL) model approach with six other machine learning model techniques.

It was seen from the experimental results of the deep CNN that the deep convolutional neural network model designed based on CNN achieved, on average, 87.14 %, 93.74 %, 95.97 %, 96.65 %, and 100 % accuracy in five classes. These results were obtained using three different cross-validation procedures: For the other coefficients, K2 for the second constant, and K5 and K10 for the fifth and tenth constants, correspondingly. Other related studies have also been conducted by Hanaa Zain Eldin et al [13]. Regarding the employment of artificial neural networks to link neurons and obtain information. Likewise in the study by Raza et al [14] performed the identification of brain cancers with the aid of TK; this is a template-based K-mean clustering real-time MRI data. The primary aim or objective of this system is the identification of brain tumors in a short duration of time and with a high degree of accuracy. First, such MRI is examined using super pixel and PCA action to obtain its important features. After that, the TK method is used again to split the area to enhance the realization of the area of the brain tumor zone, with minimal time and error in the regional identification of the area. For this reason, Sarah Zuhair Kurdi et al [15]. Considers the assumption that a multi-cascade convolutional neural network should also, engage with the local pixel dependency and the discriminative multi-scale aspects that are present intrinsically in 3D MRI images. CRFs are used to increase the certainty of the outlined tumor segmentation enhancing the edges of the outlined segmentation to eliminate multiple false positive results.

The U-net architecture is formed of the encoder stage as well as the decoder section. The encoder is an ordinary FCNN and a contracting path that slowly decodes input data and extracts features with high levels of abstraction. The decoder, on the other hand, uses an expanding path to up-sample the data by a factor at each layer and as a result, increases the size of the pictures. The author used the images processing techniques to compress the image and retain the quality of images [16]. In [17,18], authors proposed the framework and used the image processing to locate urban area. This architecture enables the U-Net to perform localization and classification as explained by Ravendra Singh, in his work [19]. An outline that was developed to represent the brains of infants to map out formations that hold neurons was provided. By designing a richly enhanced CNN, the is intense regions could be well segmented. Marin Wozniak et al. in their recent work [20].

Table 1: Performance compression of brain tumor

Study/Method	Technique Used	Accuracy	F1-Score	AUC	Dataset Used
Neelum Noreen Et al [9]	Deep Learning	93.74%	0.94	0.95	Private MRI Dataset
HASNAIN ALI SHAH et al. [12]	Efficient Learning(Transfer Learning)	96.65%	0.97	0.98	Glioma MRI Dataset
Zuhair Kurdi et al [15].	Meta-Heiristic Optimized CNN	98.59%	0.98	0.99	Public Brain MRI Dataset
Our Study	Random Forest,Stacking	99%	0.99	0.99	Kaggle Brain Tumor Dataset

Table 1 provides a comparative summary of the performance metrics reported in the existing literature. Our study demonstrates superior accuracy, F1-score, and AUC suggesting that the proposed ensemble learning methods, specifically Random Forest and Stacking, Outperform the existing SOTA methods in brain tumor classification.

3 Methodology

3.1 Dataset

We used a brain tumor dataset in this study, that was publically accessible at the website known as Kaggle. The Dataset comprises 1000 MRI images of Brain tumors which label indicating whether the tumor are malignant or benign. The images are in JPG formats. The dataset consists of brain tumors MRI images accompanied by labels specifying the kind of tumor such as malignant or benign. Following are image of different brain tumor.

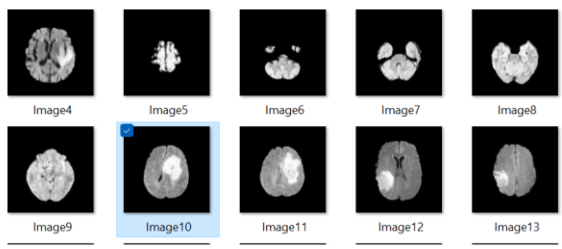

Figure 1: Following images are different brain tumors

3.2 Train-test split

The dataset was further divided into the training data with 80% of the data and the testing data with the rest 20%. To enhance the generalization capability of the developed model and avoid over fitting, we used 5-fold cross-validation. This entailed partitioning of data into 5 folds, 4 of which were used in training process with the remaining one being used in validation; the position of this validation fold was changed consecutively among the five folds every time the program was run. Performance of the models was evaluated by calculating the average of results from all the folds.

3.3 Feature extraction

These numerical attributes were extracted from MRI images of brain tumor.

These features include

- Variance of pixel intensity
- Variation of Pixel Intensity
- Standard Deviation of Pixel Intensity
- Entropy of Pixel Intensity
- Asymmetry of Pixel Intensity
- Texture measure, homogeneity, and correlation

All of them make up distinct properties of the tumor image that have helped in the classification of the tumor image.

Table 2: Various statistical features of tumors

Mean	Variance	Standard Deviation	Entropy	Skewness	Kurtosis	Contrast	Energy	ASM	Homogeneity	Dissimilarity	Correlation
4.6515	217.4596	14.74651	0.2078	3.581798	13.24406	34.6925	0.414382	0.117173	0.0661271	2.284559	0.977233
14.5713	1350.994	36.7545	0.1536	3.809051	10.033716	137.06535	0.352376	0.124169	0.564759	4.501021	0.980314
9.799197	449.9507	21.012	0.076	2.623878	7.174027	72.324409	0.240917	0.059041	0.530637	3.679742	0.95502
20.22126	1499.889	38.7284	0.30617	2.646499	7.840707	101.96901	0.150702	0.022711	0.42626	4.752189	0.976663
22.436655	1725.224	41.536	0.04659	2.391048	6.140147	186.86085	0.187357	0.035103	0.481396	4.952098	0.962109

The columns in Table 2 represent various statistical features extracted from the images. Here is a description of the features:

- Mean value of pixel intensity in the tumor image.
- Variance of pixel intensity in the tumor image
- Standard deviation of pixel intensity in the tumor image.
- The entropy of the distribution of intensities of the pixel in the images of the tumor.
- The measure of asymmetry in the distribution of intensity of the pixel in the image of the tumor.
- The measure of the Kurtosis i.e., the peak prominence or flatness of the distribution of intensity of the pixel in the images of the tumor.
- Measure of local variations in pixel intensity.
- Measure of image texture, indicating how uniform or smooth the pixel intensities are.
- Measure of image texture, indicating local homogeneity of pixel intensities.
- Measure of uniformity of pixel intensity distribution.
- Measure the difference in intensities of the pixel between adjacent pixels.
- Correlation is a metric that quantifies the extent to which intensities of the pixel at various positions determines the probability of the input data point being of a particular class.
- This is another type of learning algorithm in a group whereby an initial model is built and then followed by other models that are developed to compensate for the errors of the previous one.
- It is an algorithm based on supervised learning that divides the dataset into various Classifications by drawing a set of hyper planes in multiple-dimensional spaces in an image are linearly related.

3.4 Base classifiers

3.4.1 Naive Bayes

It is a probabilistic classifier that makes use of the Bayesian approach where the attributes are assumed to be independent.

3.4.2 Random Forest (RF)

It can be utilized as the first-level classifier in boosting architecture for developing ensembles. Collectively it is a technique in machine learning that builds multiple trees in the training process and returns the most prevalent class (for classification) or the average value of the individual trees (for regression).

3.4.3 K-Nearest Neighbors

KNN is an approach that is based on classification, and it is non-parametric and works by identifying the class in the given data point that dominates the k-nearest neighbors.

3.4.4 Logistic Regression

It is a probabilistic model applied mainly in binary classification, where it C.

3.5 Heterogeneous ensemble classifiers

3.5.1 Voting classifier

A model that incorporates the multiple base estimator's predictions and predicts the class with the most votes or the average probability for classification problems.

3.5.2 Stacking classifier

A meta-estimator trains a model to combine the predictions of multiple base classifiers, typically using a different algorithm as the final estimator.

3.6 Homogenous ensemble classifier

3.6.1 The Random Forest classifier

It works during the training where it builds many decision trees and combines the results by predicting. It is convenient to define the set of decision trees as {T1, T2, …, Tn]; where n is the entire quantity of decision trees present in the Random Forest model.

Random Forest ensemble by:

$$\hat{Y}(RF(x)) = \text{mode}\{T1(x), T2(x), \ldots, Tn(x)\}$$

The term "mode" in this context refers to the classification that appears most often and is consequently the most probable choice for the majority of individual decision trees.

Algorithm-1: Random Forest training

1. Randomly select a subset of the training dataset.
2. Grow a decision tree
   - Select a random subset of features
   - Choose the best split point among the selected features
   - Split the node into child nodes based on the best split
   - Repeat steps recursively until the tree is fully grown or a stopping criterion is met.
3. Repeat steps 1-2 to create multiple decision trees (ensemble)
4. for a new instance
   - Collect predictions from each decision tree.
   - Aggregate the predictions (e.g., take the mode for classification).
5. Output the final prediction.

3.7 Heterogeneous classification

3.7.1 Voting classifier

The Voting Classifier combines predictions from multiple base classifiers using a majority voting scheme. Let's denote the set of base classifiers as {C1, C2,..., Cn}, where n represents the quantity of the base classifiers in total. The prediction of a new instance x by the Voting Classifier is given by:

$$\hat{Y}\ (\text{Voting}(x)) = \text{mode}\{C1(x),C2(x),...,Cn(x)\}$$

Where modes represent the most frequent class predicted by the individual base classifiers.

Algorithm-2: voting classifier training

1. Train multiple base classifiers on the training dataset.
2. For a new instance:
   - Collect predictions from each base classifier.
   - Aggregate the predictions (e.g., take the mode for classification).
3. Output the final prediction.

3.7.2 Stacking classifier

This classifier acquired the ability to incorporate or merge the outputs of base classifiers using a meta-learner. Let's denote the set of base classifiers {*C*1, *C*2 ,…..... *Cn*} and the meta-learner as *M*.ss
The prediction of a new instance *x* by the Stacking Classifier is given by:

$$\hat{Y}_{\text{Stacking}}(x) = M(\hat{Y}1(x),\ \hat{Y}2(x), \ldots\ \hat{Y}n(x))$$

where $\hat{Y}i(x)$ represents the prediction of base classifier *Ci* for instance *x*, and *M* combine these predictions to make the final prediction.

Algorithm-3: stacking classifier

1. Train multiple base classifiers on the training dataset.
2. For a new instance
   - Collect predictions from each base classifier
   - To incorporate the predictions a meta-classifier can be used
3. Output the final prediction

Empirical outcomes

This part provides the results from applying numerous classifiers to the brain tumor datasets. We compare the performance of base classifiers, including Naive Bayes, Logistic Regression, Random Forest, K-Nearest Neighbors, Gradient Boosting and Support Vector Machine, against two ensemble methods: Voting Classifier and Stacking Classifier.

3.8 Experimental setup

3.8.1 Model training

Each classifier was trained using the preprocessed dataset, with hyper parameters tuned using techniques such as grid search or cross-validation to optimize performance.

a. Model evaluation

The trained models were evaluated using cross-validation or a separate test set to assess their performance based on the evaluation metrics mentioned above.

b. Comparison and analysis

The performance of each classifier, including base classifiers and ensemble methods, was compared and analyzed to assess the efficacy of various methodologies for the categorization of brain tumor. Following are algorithms of different homogenous and heterogeneous ensemble classifiers used to compare the results.

3.9 Performance metrics

The evaluation of the performance of all the classifiers was done using the metrics mentioned below

a. Accuracy

The proportion of the learned model that accurately classifies instances concerning the entire population of instances

b. Precision

In predicting a model's efficacy in avoiding false positives, the true positive to total positive ratio will

c. Recall

This is the proportion of the actual positive cases that have been accurately recognized by the model. It essentially measures the capability of the model to locate all the pertinent cases that comprise the dataset

d. F1-score

It offers the compromise between the accuracy/precision and the recall

e. The receiver operating characteristic area under the curve (ROC)

It is a performance metric of the classifier on the ability to differentiate between the two classes on the basis of the True and False Positive Rate

4 Key findings & discussions

In this section, we examine the performance of numerous classifiers in brain tumor type, that specialize in key metrics such as accuracy, precision, recall, F1-rating, and ROC AUC. The findings highlight the advanced efficacy of ensemble methods, specifically Random Forest and Stacking Classifier, in achieving high precision and sturdy overall performance throughout multiple evaluation standards. These effects underscore the ability of ensemble learning in enhancing type accuracy and generalization ability, making it mainly precious for clinical programs.

Table 3:Comparison Table for distinct measures Performance

Classifiers	F1-Score	Accuracy	Recall	Precision	ROC AUC
KNN	0.98	0.98	0.97	0.98	0.99
Naive Bayes	0.95	0.96	0.92	0.98	0.98
SVM	0.98	0.98	0.97	0.99	0.99
Random Forest	0.99	0.99	0.98	1.00	0.99
Gradient Boosting	0.98	0.98	0.97	0.99	0.99
Logistic Regression	0.98	0.99	0.97	1.00	0.99
Voting Classifier	0.98	0.99	0.97	1.00	0.99
Stacking Classifier	0.99	0.99	0.98	1.00	0.99

4.1 Comparison with the other methods of SOTA

Our study demonstrate that Ensemble learning methods, the Random Forest and the Stacking classifiers provided better performance indices of accuracy, F1-score and the AUC compared to the current SOTA models. For instance, the way depicted in Table I, approaches like the Efficient-Net based way [12] provided 96%. 65% whereas the meta-heuristic optimized CNN proposed in [15] got as high as 98.59% accuracy. On the other hand, tested Random Forest classifier reached an accuracy of 99% with F1 score and AUC of 0.99 thereby performing better than these SOTA methods that were used for the comparison.

There are number of reasons which may contribute to the performance enhancements visible in the course of our study. First, ensemble techniques like Random Forest and Stacking are used to minimize the over fitting of models which are the aggregation of several weak learners. Second, our study uses a different dataset which can be considered to have better generalizability in terms of different types of brain tumors as opposed to limited generalizability of the earlier studies.

4.2 Explanation of random forest performance

The Random Forest classifier had a better outcome than other ensemble techniques due to built-in noise tolerance as well as capability to manage high-dimensional data efficiently. While building Random Forest, several decision trees are constructed and their results are combined to lower model variance. This ensemble method also gains from the randomness that is used in creating the trees thus the property of Universal Greatness since it will perform well on unseen data. In addition, Random Forest also possesses the characteristic of performing feature selection by default because the function chose the most relevant feature for the split amongst others. It is especially useful in medical image classification since not all features that may be extracted may play a part in classification.

As seen above, the Stacking classifier had comparable results with the Random Forest with both having an accuracy and AUC of 99%. This performance is probably attributed to the nature of the stack since it integrates the output from several base classifiers, say SVM, Logistic Regression or any other. One of the advantages of applying meta-classifier all the base classifiers has different perspectives hence improving its accuracy.

4.3 Novel contributions of this study

This paper also provides evidences in the context of brain tumor classification where the Random Forest and Stacking methods are found effective and superior in terms of precision rate, accuracy rate. Finally, our approach is different from the prior works that focused often on a single model or used a shallower form of ensemble learning to work with the intricate nature of medical imaging data. Also, one of the significant differences of this work is the more extensive assessment with the use of a more significant and accessible database from Kaggle, Brain Tumor MRI Dataset, which increases the work's reliability and the ability to replicate the results. This work also tries to overcome the class imbalance problem by using proper data handling methods, even though method selection is also rarely discussed in similar works.

Furthermore, the study recommends probing into other variations in the application of ensemble methods including the union of homogenous and heterogeneous ensemble to increase the percentage of classification. This makes the way for the future studies in other fields such as the classification of other severe diseases like brain tumor.

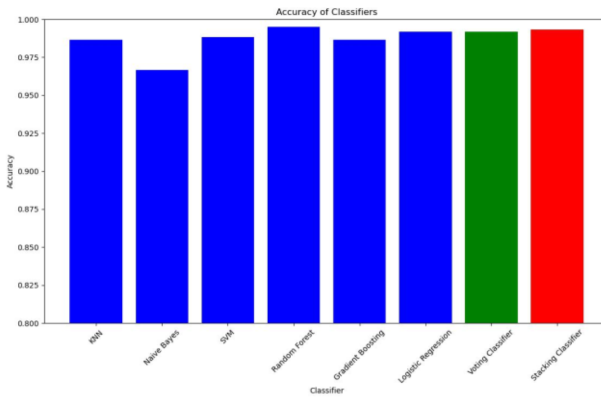

Figure 2: Graph below illustrates a comparative analysis of accuracy across different classifiers

4.4 Insights

a. Accuracy: Random Forest obtained a maximum efficiency of 99.50%, immediately after the Stacking Classifier with an accuracy of 99.33%. These results indicate that both homogenous and Heterogeneous ensemble methods outperform individual base classifiers
b. Precision: All classifiers, including Random Forest, achieved near-perfect precision scores of 1.0. This implies that the classifiers gave very few wrong predictions in the positive class.
c. Recall: The Random Forest and the Stacking Classifier possessed high recall values of 98.74% and 98.32% respectively. This shows their capacity to well categorize positive instances or tumor samples from the dataset.
d. F1 score: Random Forest achieved the highest F1-score of 99.37%, maintaining a balance between recall and precision. The Stacking Classifier also proved to be very effective with an F1 score of 99.15%.

ROC AUC

The Stacking Classifier showed the highest ROC AUC score of 99.71%. Although Voting, Boosting and Random Forest are not much behind this shows that these models have better capability of distinguishing between positive instances and negative instances

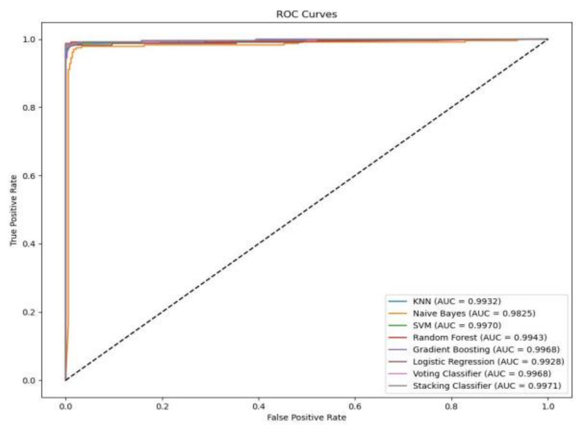

Figure 3: ROC Curve Analysis of Classifier Performance

Consistency of performance

Random Forest performed better across most of the metrics analyzed than the other classifiers. This makes Random Forest accurate and dependable for brain tumor classification problems.

Balanced performance

These results prove that Random Forest has high precision, recall, and F1-score, which means that it successfully detects positive instances (tumor samples),as well as keeping a low false positive rate. This balance is rather important for the proper identification of the problem and further management.

Generalization ability

The generalization ability and high accuracy along with the ROC AUC score of Random Forest indicate that it performs well on unseen data. This characteristic can be useful for using the classifier in clinical practice for the analysis of new data on patients.

Ensemble learning effectiveness

The fact that the Stacking Classifier performed almost as well as the Random Forest demonstrates how ensemble learning can enhance classification accuracy. The nature of using many base classifiers with a meta-classifier as illustrated by the Stacking Classifier is that it can achieve better results than using a single learning algorithm.

Robustness to imbalanced data

While Random Forest and the Stacking Classifier may appear imbalanced in the accuracy of the brain tumor dataset, the performance remained stable in all the metrics. This robustness implies that ensemble methods are applicable in managing imbalanced datasets which are prevalent in medical data analysis.

5 Conclusion

Ultimately, the extensive experimental analysis conducted in this study provides insightful findings into the field of ensemble learning approaches for brain tumor classification. This also supports our claim of how both Random Forest, a homogenous ensemble method, and Stacking Classifier, a Heterogeneous ensemble method outcompete individual base classifiers consistently in terms of F1-score, accuracy, ROC AUC, recall, and precision. Notably, as much as both ensemble methods bonded very well, Random Forest turned out to be slightly more effective than the other, performing slightly better in most of the metrics. This robust performance supports Random Forest for brain tumor classification, which is a challenging task, and it re-establishes it as one of the primary techniques used in medical image analysis.

However, while popularizing Ensemble methods one has to admit that the field of medical diagnosis is constantly changing and new ideas are needed. Thus, our work provides a foundation for further research initiatives intended to expand the scope of ensemble learning approaches in the classification of brain tumors. One area that could potentially be further investigated relates to the concern of incorporating superior feature selection techniques to boost the discriminant ability of ensemble classifiers. The utilization of Convolutional Neural Networks and Deep-Learning architectures provides a promising perspective to utilize the enormous possibility of neural networks for pattern extraction from medical imaging data.

Moreover, it would be interesting to investigate the possibilities of improving classification results even more using a combination of homogenous and heterogeneous methods. Since such hybrid methods combine the strengths of disparate ensemble methods, these novel approaches could open new possibilities to achieve very high levels of efficiency in the identification of tumors in the brain.

Furthermore, the combination of multiple genetic, histological, and clinical data is another promising avenue of research in the future. Taking full advantage of the abundant data contained in these various forms of data sources might provide deeper and more comprehensive insights into the nature of the brain tumors which would lead to more precise and individualized methods of diagnosis. Overall, the results of this study not only corroborate the applicability of ensemble learning algorithms like Random Forest for brain tumor classification but also pave the way for further research initiatives. By accepting the ongoing changes in fields such as machine learning and medical imaging, researchers will be able to push further and unlock possibilities in diagnosing brain tumors and improving the entire field of medical diagnostics.

References

[1] Azam, M., Ahmed, T., Ahmad, R., Ur Rehman, A., Sabah, F., & Asif, R. M. (2021). A Two-Step Approach for Improving Sentiment Classification Accuracy. Intelligent Automation &Soft,Computing,30(3) https://doi.org/10.32604/iasc.2021.019101

[2] Tandel, G.S., Tiwari, A., Kakde, O.G., Gupta, N., Saba, L., & Suri, J.S. (2023). "Role of ensemble deep learning for brain tumor classification in multiple magnetic resonance imaging sequence data." *Diagnostics*, 13(3), 481. https://doi.org/10.3390/diagnostics13030481

[3] Tandel, Gopal S., Ashish Tiwari, Omprakash G. Kakde, Neha Gupta, Luca Saba, and Jasjit S. Suri. "Role of ensemble deep learning for brain tumor classification in multiple magnetic resonance imaging sequence data." *Diagnostics* 13,no.3(2023):481 http:/doi.org/10.3390/diagnostics13030481

[4] Seetha, J., & Raja, S.S. (2018). "Brain tumor classification using convolutional neural networks." *Biomedical & Pharmacology Journal*, 11(3), 1457-1467.https://doi.org/10.13005/bpj/1511

[5] Ayadi, W., Charfi, I., Elhamzi, W., & Atri, M. (2022). "Brain tumor classification based on a hybrid approach." *The Visual Computer*, 38(1), 107-117. https://doi.org/10.1007/s00371-020-02005-1

[6] Biratu, E.S., Schwenker, F., Ayano, Y.M., & Debelee, T.G. (2021). "A survey of brain tumor segmentation and classification algorithms." *Journal of Imaging*, 7(9), 179. https://doi.org/10.3390/jimaging7090179

[7] Bonte, S., Goethals, I., & Van Holen, R. (2018). "Machine learning-based brain tumour segmentation on limited data using local texture and abnormality." *Computers in Biology and*

Medicine, 98, 39-47. https://doi.org/10.1016/j.compbiomed.2018.05.005

[8] Solanki, S., Singh, U.P., Singh, S., Chouhan, S., & Jain, S. (2023). "Brain tumor detection and classification using intelligence techniques: An overview." *IEEE Access*. http://doi.org/10.1109/access.2023.3242666

[9] Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., Imran, M., & Shoaib, M. (2020). "A deep learning model based on a concatenation approach for the diagnosis of brain tumor." *IEEE Access*, 8, 55135-55144. https://doi.org/10.1109/access.2020.2978629

[10] Sharif, M.I., Khan, M.A., Alhussein, M., Aurangzeb, K., & Raza, M. (2021). "A decision support system for multimodal brain tumor classification using deep learning." *Complex & Intelligent Systems*, 1-14. https://doi.org/10.1007/s40747-021-00321-0

[11] Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). "Multi-modal brain tumor detection using deep neural network and multiclass SVM." *Medicina*, 58(8), 1090. https://doi.org/10.3390/medicina58081090

[12] Shah, H.A., Saeed, F., Yun, S., Park, J.H., Paul, A., & Kang, J.M. (2022). "A robust approach for brain tumor detection in magnetic resonance images using fine-tuned EfficientNet." *IEEE Access*, 10, 65426-65438. https://doi.org/10.1109/access.2022.3184113

[13] ZainEldin, H., Gamel, S.A., El-Kenawy, E.S.M., Alharbi, A.H., Khafaga, D.S., Ibrahim, A., & Talaat, F.M. (2022). "Brain tumor detection and classification using deep learning and sine-cosine fitness grey wolf optimization." *Bioengineering*, 10(1), 18. https://doi.org/10.3390/bioengineering10010018

[14] Raza, A., Ayub, H., Khan, J.A., Ahmad, I., Salama, A.S., Daradkeh, Y.I., Javeed, D.U., & Rehman, A. (2022). "A hybrid deep learning-based approach for brain tumor classification." *Electronics*, 11(7), 1146. https://doi.org/10.3390/electronics11071146

[15] Kurdi, S.Z., Ali, M.H., Jaber, M.M., Saba, T., Rehman, A., & Damaševičius, R. (2023). "Brain tumor classification using meta-heuristic optimized convolutional neural networks." *Journal of Personalized Medicine*, 13(2), 181. https://doi.org/10.3390/jpm13020181

[16] Ali, Arshad. (2018). Coarse Classification of Terrain Image Information by Using Sobel Edge Detection Technique. International Journal of Computer Network and Information Security. 18, 149-154.

[17] Ali, Arshad. (2022). Remote Monitoring of Lab Experiments to Enhance Collaboration Between Universities. Informatica. 46. 169-177. 10.31449/inf.v46ix.xxxx.

[18] Ali, Arshad. (2022). A Framework for Air Pollution Monitoring in Smart Cities by Using IoT and Smart Sensors. Informatica. 46. 129-138. 10.31449/inf.v46i5.4003.

[19] Singh, R., & Agarwal, B.B. (2023). "An automated brain tumor classification in MR images using an enhanced convolutional neural network." *International Journal of Information Technology*, 15(2), 665-674. https://doi.org/10.1007/s41870-022-01095-5

[20] Woźniak, M., Siłka, J., & Wieczorek, M. (2023). "Deep neural network correlation learning mechanism for CT brain tumor detection." *Neural Computing and Applications*, 35(20), 14611-14626. https://doi.org/10.1007/s00521-021-05841-x