Enhanced Brain Tumor Classification using Feature Optimization

Vellanki Priyanka
Dept. of Electronics and Communication Engg.
Amrita School of Engineering
Amrita Vishwa Vidyapeetham
Bengaluru, India
vellankipriyanka1@gmail.com

Vendra Lakshmi Sowjanya
Dept. of Electronics and Communication Engg.
Amrita School of Engineering
Amrita Vishwa Vidyapeetham
Bengaluru, India
vlsowjanya03@gmail.com

Yoshithaa B
Dept. of Electronics and Communication Engg.
Amrita School of Engineering
Amrita Vishwa Vidyapeetham
Bengaluru, India
yoshithaabalaji21@gmail.com

Neelima N
Dept.of Electronics and Communication Engg.
Amrita School of Engineering
Amrita Vishwa Vidyapeetham
Bengaluru, India
n_neelima@blr.amrita.edu

***Abstract*— Excessive cell growth causes cancers that can metastasize to other organs. 80% of CNS cancers are brain tumors. So, improving survival rates is essential which in turn requires early detection, especially in this present lifestyle. Brain tumors cause significant challenges in the medicine industry and are ranked as the fifth most prominent cancer type over 100 types of cancers. Despite the advancement in AI, accurate cancer detection and classification are not present which makes it inconclusive. Many models state their efficiency in detecting brain tumors but the diverse nature of tumors presents challenges in ongoing research. This paper aims to compare different models of deep learning that can be used for the classification of specific tumors like Glioma, Meningioma, Pituitary tumors as well as detecting No tumor. The paper has shown a good accuracy by using RFE with Logistic Regression among all the methods mentioned.**

Keywords—Brain Tumor, Feature Extraction, RFE, Random Forest, Mutual Information, Chi-squared

I. Introduction

In today's fast paced society where money comes first before anything, health concerns always take a backseat. Different factors such as dietary habits, exposure to toxins and radiation are the causes of cancers. Apart from all of these factors changes in DNA are also causing cancers which in turn is related to our lifestyle. Brain tumors are formed due to the excessive cell growth. The tumors can be of two types - Malignant and Benign. The diverse nature of tumors in terms of its size, location, and type possesses challenge for the doctors to help patients. Given the considerable burden of brain tumors, with over 300,000 new cases every year worldwide and significant deaths in countries like USA and India shows the urgency for effective treatment. The tumor count includes an approximate amount of 54% in men and 46% in women. Nearly 4500 children across the globe are said to be diagnosed with brain tumor.

The study and funds spent on Oncology now is more than ever, which silently -indicates the growth of a violent disease that is going to create a havoc in future generation. The claimed treatment to remove and control tumor are: Tomotherapy and Radiation therapy. Treatments strategies vary depending on what kind of cancer, what stage is it and existing treatments. Recent developments in Russia on cancer vaccines signify potential breakthroughs in cancer treatments. Thus, the early and efficient diagnosis is essential for saving many lives and providing the patients with a better treatment.

In medical diagnosis of brain tumor, the detection and classification algorithms play a vital role in precise and efficient analysis of the MRI images. A technique proves to be efficient for a dataset while not a better choice for another, which indicates the level of normalization that is needed for every dataset, and the need to include the most intricate feature detail so that it can detect the type of tumor more precisely. Machine learning and CV techniques has emerged as a favorable stream for automating this process by potentially enhancing the speed and diagnosis.

Trained models like ResNet, DenseNet and CNN are utilized to make this process more efficient. Computational efficiency of the models in processing and classifying of an MRI image into various classes like Pituitary, Glioma, Meningioma or as no tumor plays a major role than calculation of accuracy. This paper aims to explore various algorithms for Brain tumor detection and classification, which can aid the situation presented above.

The key challenge that was encountered during this process was in the extraction of relevant features from the brain images, which is a task traditionally labor-intensive and prone to errors. Thus, the method that performs feature extraction, must be developed critically taking into consideration the chances of that being able to aid in predicting a type of tumor in a most accurate manner. By automating feature extraction and analysis, the methodologies presented here aims to provide a more efficient and reliable approach to brain tumor detection and classification.

II. Literature survey

Machine learning models and Deep CNN are widely used for the research on identification of different kinds of brain tumors. Each model's performance varies with respect to the dataset that is being tested on, the methodology that is being applied, the features being calculated, the reduction techniques performed, and so on. Thus, each model carries its own pros and cons. So, choosing of a model must be done, by keeping in mind a few parameters like desired computational time, performance, accuracy, precision, recall, input size, f1-score, etc.

The authors in [1] have suggested a CNN model with maximum pooling layers as 5, batch normalization, ten convolutional layers, and fully connected layers with Softmax applied in the output. A new feature selection approach 'EKbHFV' was proposed in [2], where modified version of GA was included for the best feature selection. Bothof them have shown incredible performance but the only disadvantage is that they need significant computational resources and time, which implies the need of models that can perform feature extraction and classification in a less amount of time, with enhanced performance.

The implementation of MM-LinkNet as an encoder-decoder network with ResNet152 as the backbone, along with three robust CNN models for different datasets was implemented in [5][6]. They proved to be efficient models but the complexity of the structures needs powerful hardware for training and optimization which may pose as a challenge for optimizing computational resources and infrastructure. However, in addition to that, the method used in [5] poses serious challenges related to manual dataset annotations. Thus, these techniques indicate the need to develop a model, where such data structures are being used which reduces the need of much computational efforts required to perform the extraction procedure.

Different techniques such as Skull stripping, bias correction, filtering, etc., are used in [3] along with Machine learning and Deep learning. However, the paper does not provide a very detailed technical analysis of each method. In [4], the methodology involved preprocessing with pre-trained models like EfficientNet50 and ResNet50, but the complexity involved is the biggest disadvantage when compared to simpler approaches that have been already introduced. These methodologies suggest the need to introduce such models, to reduce the complexity involved in the process of extraction and will provide better results with less computational time and better accuracy.

In [13] they proposed a method named Deep Dense Inception Residual Network (DDIRNet), which strengthens both Inception and Residual networks. It incorporates three key elements which are Inception residual network, deep dense layers and Regularization techniques. Although it has high classification accuracy compared to existing models and showed robustness to noise in MR images but it has a problem with requirement of data which need to be huge for training purpose. It might become computationally expensive compared to all the existing simpler models.

[14]A CNN model was developed to accept an input image and it was passed into many subsequent convolutional layers and a feature map of reduced size was obtained after the max pooling stage. The process tried to decrease the spatial data used, by half. Adam Optimizer was used as it's known for handling sparse gradients and noisy data. Categorical Cross Entropy loss function was used to calculate the loss value. Though the model boasts an accuracy of 93.3%, it had a few drawbacks like small dataset, etc. Also, the usage of Holdout Validation method could lead to a partial estimate of performance as the dataset involved may not cover all the necessary features involved in detection of tumor. This tells us to utilize models that would evaluate the dataset in an impartial manner, which would lead to higher accuracy estimation.

For brain tumor recognition [15] a Multi-level Attention Network (MANet), which includes both spatial and cross-channel attention was utilized. BraTS benchmark datasets and Figshare was used to evaluate the performance of the model. Though the model achieved higher accuracy, it didn't include the functionality to test the No Tumor case. Thus, the need to perform multi-task classification was highlighted through it.

III. Proposed Methodology

The methods presented in this paper mainly focused on extracting the features from the MRI scans and then classifying them into Meningioma, glioma, pituitary tumor, and no tumor. To perform this objective, four different techniques have been performed, which involves the use of varied techniques for feature extraction and classification. Each technique has its own merits and demerits, which will be dealt in this section of the paper.

The characterization scheme used can be broadly classified into 3 types: Statistical Measures, Texture Analysis, and Geometric features.

The features described in table I are used for further processing and are generated using [8]. A dataset consisting of 4000 images, had been used to obtain the results published in the below sections.

TABLE I. FEATURES USED FOR CLASSIFICATION IN THE METHODOLOGY

Characterization Scheme	Features
Statistical Measures	Mean Standard Deviation Kurtosis Skew
Texture Analysis	ASM (Angular Second Moment) Contrast Energy Homogeneity Correlation Dissimilarity
Geometric Features	Area Convex Area Eccentricity Euler Number Extent Orientation Solidity

For all the below methods, the % reduction in each parameter is calculated using eq.1

$$Percentage\ reduction = \left(\frac{No.of\ features\ before - No.of\ features\ after}{No.of\ features\ before}\right) * 100 \quad (1)$$

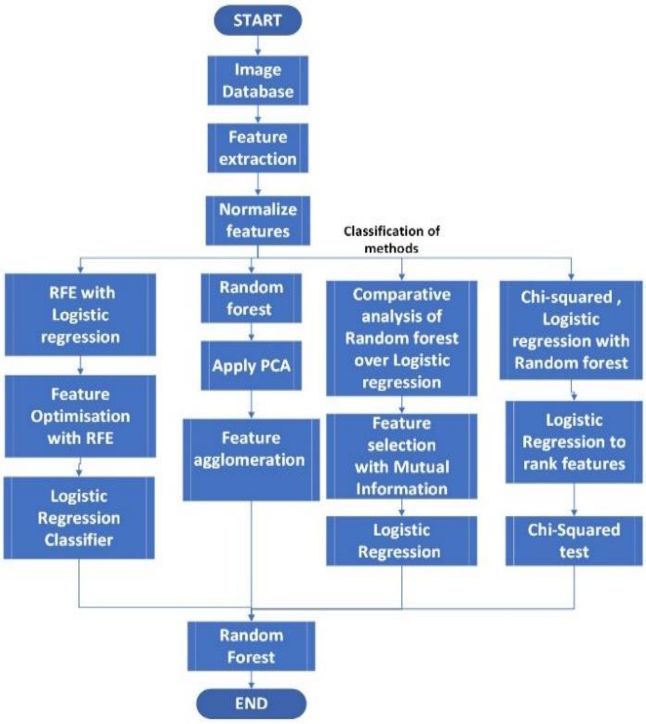


Fig 1. A Flowchart illustrating methodologies Brain Tumor classification

A. RFE with Logistic Regression

After feature extraction, Recursive Feature Elimination (RFE) is used to filter the feature set by selecting the best number of features repeatedly selecting the most informative. Then comes the turn for Logistic Regression which provides coefficients for each feature depending on their correlation with the target variable. It reduces overfitting and enhances models' reliability as well as it is necessary to filter out some more features which are not linked for the classification of tumor. In addition to that Random Forest Classifier [10] is also used. This gives the methodology robustness in terms of accurate classification of MRI images into their predefined classes.

B. Random Forest Classifier

Statistical, Texture and Morphological features are extracted from images in the first step and then they are normalized. PCA is applied to reduce the dimensionality of the feature space and then random projection is employed to further reduce dimensionality. Feature agglomeration is used to group similar features. Random forest Classifier [10] is employed for training on the preprocessed data from the above steps and evaluated performance on the basis of accuracy, precision, recall and, F1 score. The classifier is very efficient in classification of the different entities, here for which we are using it for brain tumors. Comparative analysis of Random Forest Classifier over Logistic Regression after performing Mutual Information

An extensive feature extraction process is performed on the brain tumor dataset, where the parameters mentioned in the above table are computed. For scaling of features and to maintain uniformity in their ranges MinMaxScalar operation is executed. Mutual Information technique is then performed on them for selection of top features relevant for classification. High mutual information scores highlight the informative features, while low scored suggest less relevance. Thus, this technique helps in improving computational efficiency, reducing dimensionality and mitigates overfitting risks. Two classifiers, Random Forest [10] and Logistic Regression, are trained on the selected features. In brain tumor classification, Random Forest provide insight into feature importance, aiding in understanding which features contribute most to the classification task and Logistic Regression is used for its simplicity and interpretability, as it classifies an image into two classes, in a simple yet effective manner.

It was seen that on performing the comparative analysis as provided in table II which involves the usage of Random Forest Classifier and Logistic Regression for feature extraction proves that Random Forest Classifier provides better Accuracy and a significant increase in other parameters which are involved in the process than the latter.

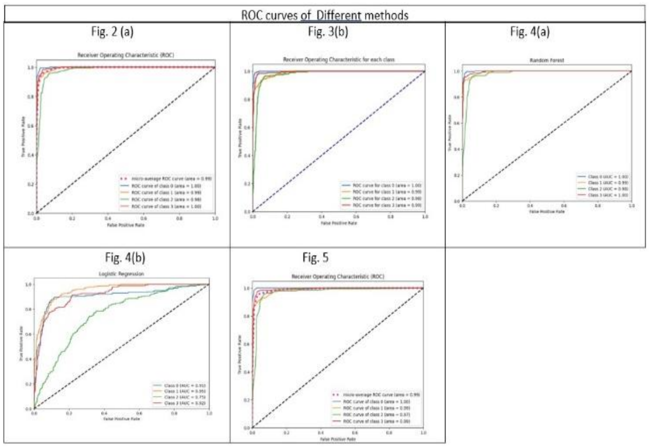
C. *Using Chi-Squared, Logistic Regression, and Random Forest classifier*

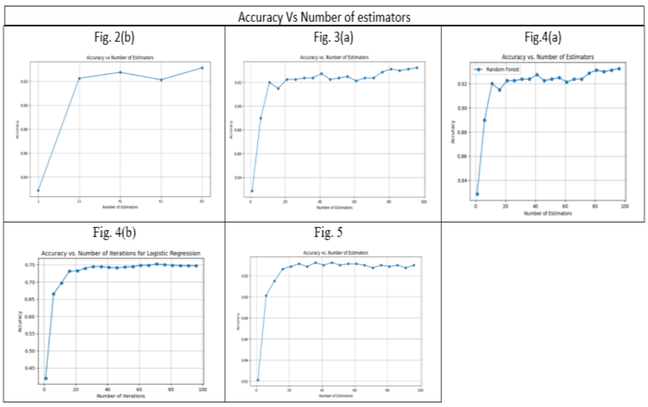
Extraction features uses various techniques such as morphological analysis etc., to obtain distinct characteristics of given tumor images. It calculates statistical measures which are mentioned above in table. After feature extraction using MinMaxScaler, then comes the feature selection techniques to identify most relevant features along with class labels. These class variables are the target variables. Logistic Regression is used to rank the features then chi-squared is being employed to select the top features with strong correlation with the target variables. With the selected features in hand, Random Forest [10] is applied.

Finally, the trained model is strictly evaluated on basis of accuracy, precision, recall and F1 scores.

IV. Results and Discussion

The performance of the classifiers is evaluated using metrics like accuracy and classification reports.





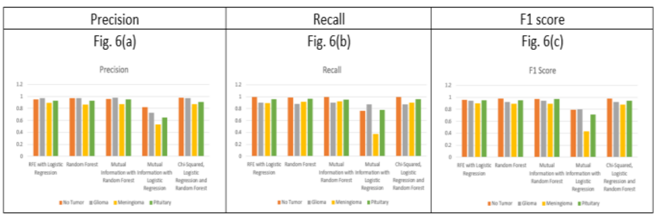


TABLE II PERFORMANCE COMPARISION OF THE METHODS USED

Method types	No. of Parameters before optimization	No. of Parameters after optimization	% Reduction	% Accuracy
RFE + Logistic Regression	**17**	**9**	**47.058**	**94**
Random Forest	**17**	**17**	**0.000**	**93**
Mutual Information + Random Forest	**17**	**11**	**35.294**	**94**
Mutual Information + Logistic Regression	**17**	**11**	**35.294**	**70**
Chi-Squared + Logistic Regression + Random Forest	**17**	**11**	**35.294**	**93**

Fig 2(a),(b) are the plots of ROC and accuracy vs number of estimators for RFE with Logistic Regression. Fig 3(a),(b) are the plots of Accuracy vs. number of estimators and ROC for Random Forest. Fig 4(a) is plotted for Accuracy vs Number of estimators and ROC of Mutual Information with Random Forest. Fig 4(b) is plotted for Accuracy vs Number of estimators and ROC of Mutual Information with Logistic Regression. Fig 5 depicts both ROC and Accuracy Vs Number of estimators of the Chi-Squared, Logistic Regression, and Random Forest method. Fig 6 are the bar graphs representing Recall, Precision, and F1 Score respectively for methods explored in the previous sections for no tumor, Glioma, Meningioma and Pituitary tumors. Fig 6 are the bar graphs representing Recall, Precision, and F1 Score respectively for methods explored in the previous sections for no tumor, Glioma, Meningioma and Pituitary tumors. From the ROC figures, we can observe that for most of the models we discussed the graphs are tending to 1 which signs them as good models.

From Table II it can be seen that both RFE with Logistic Regression and Mutual Information with Random Forest gave the same accuracy with 94% accuracy. In addition to that, both methods have shown similar F1 scores, Precision and Accuracy. But to provide the same accuracy, the former method needed 9 features compared to the latter method which took 11 features.

PCA and agglomeration along with Random Forest [10] classifier can be seen to give 93% accuracy but the percentage reduction in features is 0% which is a major setback for that method. The Logistic Regression after performing Mutual Information gave 70% accuracy which is lower than all the methods discussed in this paper. It is not that reliable compared to remaining but has got the same percentage reduction in features which is 35.294% when seen with the model it is compared with.

V. Conclusion

According to recent statistics it is shown that there is very less survival rate of 33%, if detected late. This shows the urgency of early detection of the Brain tumors. This paper showcases five different methodologies to perform the classification on the data obtained from features extracted from MRI images.

The results shows that both Recursive Feature Elimination (RFE) with Logistic Regression and Mutual Information with Random Forest achieved accuracy of 94%. However, RFE with logistic regression is better than Mutual Information with Random Forest, as it only needs 9 features while the latter ones need 11 features to obtain the same accuracy. Lower the number of features required, faster is the computation. It can be made more better by using advanced feature selection techniques. However, selecting a good model depends on type of the tumor, how urgent is the detection of the tumor, and what type of clinical procedures are considered.

It is important to get the accuracy higher so that model can be reliable. According to our work, we found RFE with Logistic Regression is way better than the models we worked on.
We also suggest the feature extraction first and then go for simple but efficient classification models like Random Forest classifier rather than selecting CNN for the same.

References

[1] Deep CNN Architecture for Brain Tumor Classification from MRI Images. Neural Processing Letters (2021) 53:671–700.Published online: 6 January 2021

[2] A Decision Support System for Multimodal Brain Tumor Classification Using Deep Learning Complex & Intelligent Systems (2022) 8:3007–3020 Published online: 9 March 2021

[3] Brain tumor detection and classification using Machine learning: a comprehensive survey) (ZainEldin H, Gamel SA, El-Kenawy EM, Alharbi AH, Khafaga DS, Ibrahim A, Talaat FM. Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization. Bioengineering (Basel). 2022 Dec 22;10(1):18.

[4] A deep learn ing approach for brain tumor classification using MRI images. Published online: 26 May 2022

[5] Multi-Model Semantic Segmentation Model using Encoder-Based Link-Net Architecture for BraTS 2020 Challenge.Gayathri Ramasamy, Tripty Singh, Xiaohui Yuan,ISSN 1877-0509

[6] Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework, Iranian Journal of Science and Technology, Transactions of Electrical Engineering (2021)

[7] Deployment of Breast Cancer Hybrid Net using Deep Learning by Nipun B Nair, Amrita Thakur,Tripty Singh and Prakash Duraisamy 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT) |

[8] Mutlag, Wamidh & Ali, Shaker & Mosad, Zahoor & Ghrabat, Bahaa Hussein. (2020). Feature Extraction Methods: A Review. Journal of Physics: Conference Series.

[9] Chen B, Zhang L, Chen H, Liang K, Chen X. A novel extended Kalman filter with support vector machine based method for the automatic diagnosis and segmentation of brain tumors. Comput Methods Programs Biomed. 2021 Mar;200:105797

[10] Iranzad, R., Liu, X. A review of random forest-based feature selection methods for data science education and applications. *Int J Data Sci Anal* (2024).

[11] Shivalila.H, Neelima.N, Deepa.k, Tolga Ozer, “An evolutionary model for sleep quality analytics using fuzzy system”, Proceedings of the institution of mechanical engineers Part H: Journal of Engineering in Medicine, (SCI indexed) IF-1.8, 2023

[12] A. Das, N. Neelima, K. Deepa and T. Özer, "Gene Selection Based Cancer Classification with Adaptive Optimization Using Deep Learning Architecture," in IEEE Access

[13] Kokkalla, S., Kakarla, J., Venkateswarlu, I.B. *et al.* Three-class brain tumor classification using deep dense inception residual network. *Soft Comput* **25**, 8721–8729 (2021).

[14] Mahmud, Md Ishtyaq, Muntasir Mamun, and Ahmed Abdelgawad. 2023. "A Deep Analysis of Brain Tumor Detection from MR Images Using Deep Learning Networks" *Algorithms* 16, no. 4: 176.

[15] Shaik, N.S., Cherukuri, T.K. Multi-level attention network: application to brain tumor classification. *SIViP* **16**, 817–824 (2022).