Fast Hybrid Adaboost Binary Classifier For Brain Tumor Classification

1S.Jayaprada                2G.JayaLakshmi            3L.KanyaKumari

1Associate Professor, Department of Computer Science & Engineering, Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India. jayaprada79@yahoo.com
2Assistant Professor, Department of Information Technology, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India. jaya1123@gmail.com
3Assistant Professor, Department of Information Technology, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India. kanyabtech@yahoo.com

ABSTRACT

The brain tumor is a dangerous disease and its characterization is a difficult undertaking for radiologists in light of the heterogeneous idea of the tumor cells. Computer-Aided tasks and its implementation of the current models with their frameworks would suffice the design metrics to recognize and relate the different tumors including brain though the process with scanning of the brain would emphasize on MRI. The models with support vector machine-nearest neighbor, naïve bayes analysis aren’t enough to produce the different scenario at each set of layers and its correlation values of the performance observed. We propose a design model with fast Adaboost binary classifier, ensemble approach with fast boosting algorithm with the pre-trained model's dataset to analyze the different problem which proposes a strategy for multiple application scenarios of feature extraction to provide a classification model for different tumors in the brain to improve different performance parametric values for each trained and test sets with the prediction algorithm with each design formulation.

Keywords:
Tumor, Brain-tumor, Classification, Adaboost, ensemble, CNN, RCNN

1. INTRODUCTION

Medical Advancements in the field of biomedical aspects have defeated different sicknesses over the most recent couple of years yet peoples are as yet, experiencing disease because of its unusual changes in the real-time world. This tumor related to the brain is as yet a critical issue for mankind. The tumor in the brain for this disease has been genuine and most extreme rising afflictions as of till known scenario. In the USA, right around 23,000 patients were distinguished mind tumor malignant growth in 2015 [1].In one more report of disease markers [2], this ailment is consistent in the two grown-ups and youngsters. Roughly 80,000 fresh instances of essential cerebrum tumors were accounted for in 2018 [3]. Meningioma spoke to 36.3% (29,320), Gliomas 26.5% (21,200), Pituitary tumors spoke to almost 16.2% (13,210) and the remaining of the cases had a place with different kinds of cerebrum tumor, for example, Malignant, Medulloblastoma, and Lymphomas. The chief reasons for such ailment are disease related sickness with dismalness. Successful treatment of this illness is significant which depends on its convenient and precise recognition. Presently the treatment of the tumors and its examination has become more costly and hectic states to identify each and differentiate the tumors. So their regenerative and active scenario would affect the different systems utilized to implement such drastic measures. Unpredicted nature of the changes in cell generation and its accumulation as a cluster results in formation of a clad known as tumor. Along these lines, cerebrum tumors are an autonomous and unpredictable proliferation of synapses [4], [5].

2. MOTIVATION

Boosting is an ensemble method that tries to create a strong classifier from a number of weak classifiers. The machine learning technology that discovered to answer problem like that if the data is partial, or the chosen representation is over-fitted with the related data, the representation cannot illustrate the exact performance is
ensemble approaches. Ensemble approaches are mainly modeled into three types, specifically, boosting stacking, and bagging. Bagging is a procedure of permitting redundancy as arbitrarily classifying training statistics to a forecast form in the procedure of knowledge. Boosting is alike to bagging, but it reprocesses the data that the previous representation cannot categorize fine in the learning procedure by given that weight. Stacking is an ensemble method that is as well identified as great learning. It utilizes different forms, such as random forest, k-nearest neighbors, and support vector machine as support learners to make innovative data based on expected results. Next this novel data is used for one more analytical model, which is called meta-learner, to in conclusion obtain the predicted assessment. Low-level processes such as thresholding, edge detection, and morphological techniques, are speedy and be able to be used for brain tumor segmentation. But, the routine of these methods extremely depends scheduled the apparent difference in the intensities among tumor and non-tumor areas. We propose a fast hybrid binary classifier for this. The phrase hybrid is used here as, in other ensemble models, a identical set of weak learners is used excluding here, a various different set of weak learners is used. The hybrid ensemble learning form, assembled by these weak learning models, is functional in the job of brain images tumor, on-tumor classification.

3. LITERATURE REVIEW

AI approaches have been widely utilized in different spaces including clinical diagnostics and precautionary medication. A predetermined number of studies, in any case, contain focused on determination of cerebrum tumor particularly utilizing attractive reverberation imaging (MRI). Generally ML techniques train with test customary ML calculations on MRI information. As of late, a portion of the methodologies have utilized DL for the determination of cerebrum tumor.

Many studies in the literature used single classifier techniques for brain tumor classification. Many studies declared that classifier ensembles suggest improved performance, compared through single classifiers [11]. The ensemble of classifiers has been used for the data mining domain [12].

S. Mary Praveena and Ila Vennila.[3] proposed a structure that utilized an arrangement called tri-design. With scenario of different data set from the Neural net observations different scenario of the tumors especially related to brain are investigates and analyzed. The creators additionally considered information expansion methods to acquire results’ precision. This exploration accomplished an exactness of 98.69% by utilizing engineering for upgrading grouping and identification.

Subsequently, these authors W. Yang Q. Feng M. Yu Z. Lu Y. GAO Y. Xu Et Al.[4] improvise a three dimensional structure modeling of CNN that naturally reviews glioma from customary multi-arrangement scanning of each observed tumor part by characterizing a locale of intrigue consequently. The trained model and its performance accuracy was turnout to be 89.5 %, notwithstanding the exactness of anticipating tumor that was accomplished at 92.98% dependent on areas of interests.

Sugandha Agarwal, O.P. Singh and Deepak Nagaria., [6] additionally received the idea of profound exchange for arrangement of pictures and utilized a similar information source talked about in [5]. The design model and its functionaries were classified and represented with test and train scenario of each set of case utilized. To characterize patients requirements and their identification of the tumor with each section of the images from the patients are observed with the values of 98%.

Convolve nets with as a CN network model, by B. Cheng, M. Liu, H. I. Suk, D. Shen and D. Zhang., [8] characterized cerebrum tumors. The investigation improved the degree of precision of the design and its features with patients were at 86.50% by imparting the different component layers with each set of parametric criteria to enhance the design with 64 levels by getting a variety the guides of Caps Net at some layer implemented with convolution.

Mr. Manish Saraswat and Mr. Ajay Kumar Sharma., [9] the design model with layers connecting with convolution and rectification have characterized with different specific scenario of difference observed from the results which are actuated with the value of 98.5% on preparing and 84.19% for approval. The investigation accomplished a
precision proportion of 91.9% by utilizing the previously mentioned framework arrangement with back propagation NN.

Sushma Laxman Wakchaure and Anil Khandekar., [10] built up an engineering dependent on CNN for highlights extraction. To improvise a specific structure and its importance we have different learning layers and its arrangement with 3X3 model. The investigation professed to accomplish a precision of 81% that was additionally upgraded to highlight arrangement model of CNN dependent on ELM (outrageous model for to learn different machine algorithms represented).

4. DATASET DESCRIPTION:

The dataset used is taken from the Kaggle. The dataset have a total of 253 MRI images. Out of them, 155 are marked “yes”, which shows that there is a tumor and the remaining 98 are marked “no”, which says that there is no tumor.

![Figure 1: Sample image of an MRI](image)

The above figure 1 shows the Tumor in the left side of the brain and in the second image, there is no tumor formation.

Implementation of Existing models: CNN and mask RCNN:

CNN:

The CNN Model is designed as a neural network using Keras library with different convolutional and pooling layers.

Steps:

1) Pre-Training the CNN model: for this we create a “train generator” and “validation generator” to store split images of the training and the test sets into two classes (yes & no).

2) Training the CNN model: we fit the image data to the trained neural network for around 150 “epochs” and at the end of the 100th epoch, it was observed that the trained CNN model has validation accuracy of 70.85% which denotes that model can correctly classify about 71% of the test set images as Tumor/Non-tumor.
Figure 2: Results of Brain Tumor classification using CNN

After the training, the “accuracy” and “loss” of both “train generator” and “validation generator” for all the 100 epochs(iterations) are plotted.

Figure 3: Accuracy and Loss graph

Mask RCNN:

The mask RCNN model is used to construct a definite detector to spot out on the position of the tumor on the MRI scan. For this implementation annotations are added for scans with VGG Image Annotator (VIA). And dataset is divided into 4 sets training, test, validation and annotations. Mask R-CNN to make an actual detector which will point out on the position of the tumor on the scan.

Steps:

1) Configuration for training on the brain tumor dataset.
2) Get the x, y coordinates of points of the polygons that build up the draw of each object illustration. These are stored in the contour attributes.
3) Generate occurrence masks for an image. A Boolean array of shape [height, width, instance count] with one mask per instance.
4) Training and validation the dataset.

Figure 4: Results of Brain Tumor classification using RCNN

Region-based convolutional neural network shows accuracy results 83% for the same dataset which have been used for CNN

5. PROPOSED MODEL

Fast Adaboost binary classifier:

The novel extensions of the Fast Adaboost binary classifier is to improve tumor classification rate. The key idea is to present the boosting algorithm with a synopsis of the training dataset such that the learning algorithm can rapidly and competently prepare the classifier.
According to literature review, it was originate that brain tumor classification is extremely important as elevated accuracy is needed when human life is concerned. As Adaboost learning algorithm is believed to be computationally capable but training is time intense in image processing and classification, we propose to build classifier model based on boosting ensemble methods which selects a reduced set of training points, thus reducing large computational difficulty for training and rising the speed of the training progression.

If any classifier is given unlabeled image, it cannot classify that image with the knowledge obtained through the training process. So we have to train the classifier with a set of images with labels.

The same dataset considered for CNN and RNN is considered for proposed algorithm implementation.

**a) Data Augmentation:**

Clearly dataset indicate that we have data imbalance i.e., number of observations per class is not uniformly distributed. This data imbalance is solved using data augmentation because this is very important part in medicine as there would be very less number of unhealthy patients than the number of healthy patients in many instances. We performed image enhancements like flip & rotate. Data augmentation transformed the dataset into more number of images i.e., increases the size of the training set.

**Flip and Rotation:**

Random flipping creates a mirror likeness of an original image beside one (or more) chosen axis. This operation helped AHM, especially benefitting background tumor information i.e., with respect to their location within the brain which would be or else complex for example having brain tumors positioned only in the left/ right hemisphere. This process is followed through right exclamation to suit the original image size.

After data augmentation, the dataset contains almost equal number of **1085** images of “yes” class and **979** images of “no” class.

This data set now divided keen on training and test data sets. The training dataset contains 80% (1651 images) and test set contains 20% (413 images).

Removed any colors in the image to convert it to the gray scale by CvtColor function (python)
b) Data Pre-processing:

Pre-processing of image prior to segmentation is vital for correct discovery of tumor. In this phase, we execute noise and artifacts decrease and sharpening of edges. There are slight chances of noise being there in the MRI images. Thus the key job of the preprocessing is to file edges in the image.

Filters be used to remove the noise from the image as preprocessing step. Here we used median filter for reducing the intensity difference between one pixel and its neighbor’s pixels. All images are preprocessed using median filter, edge detection, binarization using thresholding.

c) Segmentation:

Segmentation is the procedure of dividing the image into smaller objects. So that objects in the similar cluster are alike as possible and objects in dissimilar clusters are as different. The segmentation procedure involves binarization using a thresholding technique, enhancing the image using morphological transformation, segmentation and edge detection.

The preprocess image is segmented by means of thresholding and tumor is identified. The white part in the image shows the tumor region in an MRI.
Feature Extraction:

Figure 8: Tumor identification in segmentation process

Figure 9: Process of features extraction.

Feature Extraction:

GLCM (texture-based features) is used to extract features from images as we need to perform a binary classification of them using a fast Adaboost binary classifier which needs these features to get trained on. After obtaining the segmented images from the above process, we extract GLCM features and store them. GLCM indicates for Gray-Level Co-occurrence Matrix. Texture Analysis Using the Gray-Level Co-occurrence Matrix (GLCM) is a numerical system of exploratory surface that considers the spatial link of pixels. Here feature describe as properties in class region, it have every shape feature.

Feature selection and extract has been used for 17 features from Tumor image array.

1) Low Variance Filter to remove features that have low variance data sample. These Features are: Solidity, Euler number, Extent.

2) Normalize the rest of features to create feature values Convergent as likely using Standard deviation & Mean.
3) Lastly, PCA was used to perform linear dimensionality reduction. The output obtained proves that only three features: homogeneity, correlation, dissimilarity will have an effect on or have the best effect in classification problem.

![Figure 11: PCA output for 14 input features](image)

As we know set of several models functioning together on a only one set is called an ensemble. The technique is called Ensemble Learning. It is a great deal and more useful to use all different models rather than any one. To execute fast adaboost binary classifier we mixed different classifiers where the class which has been predicted generally by the weak learners resolve be the ultimate class forecast of the ensemble form.

Test accuracy 0.904048  
Training duration 0.06604sec

### Table 3: Results of CNN, RCNN and start-of-art techniques

| ALGORITHMS                  | ACCURACY for 150 iterations (%) |
|-----------------------------|----------------------------------|
| PROPOSED APPROACH FAHM (Fast Adaboost Hybrid Model) | 90.4%                           |
| CNN                         | 68%                              |
| RCNN                        | 73%                              |
| Abiwinanda et al. CNN [13]  | 84.19%                           |

### DISCUSSION

1. Figures from 2-7 representing the CNN and RCNN model of comprising an accuracy of 64% and 73% with 150 epochs.
2. Similarly our proposed design with 150 epochs have been modeled with 160 cycles and learning rate of 0.043. Hence the total learning rate is 100-160 *0.043 = 90.4 %.
3. Classification of the different parameters related to brain tumor are implemented and observed in figure 6 with confusion matrix.
4. Each calculated values and its representation are tabulated as mentioned in table 3.

### CONCLUSION

The detection, segmentation and uncovering of infecting region in brain tumor MRI images are a dull and time consuming job. The dissimilar anatomy construction of human body can be seen by an image processing notion. We have taken heterogeneous collection of weak learners (machine learning models: Logistic Regression Model, K-
Nearest Neighbor Model, Support Vector Machine, and Naive Bayes Model) to construct our hybrid ensemble learning model, applied in the task of classification of brain MRI images. Our proposed system has accomplished the accuracy of 90.4% utilizing the given dataset.

In future, we will implement algorithms with more accuracy even for single image and also for large datsets, and identification of tumor types using artificial intelligence.

REFERENCES

[1] Neelum Noreen, Sellappan Palaniappan, Abdul Qayyum, Ifikhar Ahmad, Muhammad Imran and Muhammad Shoaib, 2019 “A deep learning model based on concatenation approach for brain tumor classification and identification”, IEEE Access, Volume 8.
[2] Q. T. Ostrom H. Gittleman J. Xu C. Kromer Y. Wolinsky C. Kruchko et al. 2016 "Cbtruus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2009–2013" Neuro-oncology vol. 18.
[3] S. Mary Praveena and Ila Vennila,2010 "Optimization Fusion Approach for Image Segmentation using K-Means Algorithm", International Journal of Computer Applications, 2 7.
[4] W. Yang Q. Feng M. Yu Z. Lu Y. Gao Y. Xu et al. 2012 "Content-based retrieval of brain tumor in contrast-enhanced mri images using tumor margin information and learned distance metric" Medical physics, 39 11.
[5] M Monica Subashini and Sarat Kumar Sahoo, 2013 ”Brain MR Image Segmentation for Tumour Detection using Artificial Neural Networks”, International Journal of Engineering and Technology 5 2.
[6] Sugandha Agarwal, O.P. Singh and Deepak Nagaria, 2017 ”Analysis and Comparison of Wavelet Transforms For Denoising MRI Image”, Biomedical & Pharmacology Journal, 10 2.
[7] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng and S. K. Zhou, 2016 ”Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images”, International Conference on Medical Image Computing and Computer-Assisted Intervention.
[8] B. Cheng, M. Liu, H. I. Suk, D. Shen and D. Zhang, 2015 ”Multimodal manifold-regularized transfer learning for MCI conversion prediction”, Brain imaging and Behavior, 9 4.
[9] Mr. Manish Saraswat and Mr. Ajay Kumar Sharma, 2013 ”Genetic Algorithm for optimization using MATLAB”, International Journal of Advanced Research in Computer Science, 4 3.
[10] Sushma Laxman Wakchaure and Anil Khandekar, 2015 ”visualization of 3d view of detected brain tumor and calculation of its volume”, International Journal of Technical Research and Applications, 3 6.
[11] Zolfaghar R.2012, “Diagnosis of Diabetes in Female Population of Pima Indian Heritage with Ensemble of BP Neural Network and SVM” Int. J. Comput. Eng. Manag. 15 4.
[12] Bashir.S. Qamar,U. Khan,F. 2016,”IntelliHealth: Amedical decision support application using a novel weighted multi-layer classifier ensemble framework”, Journal of Biomedical Informatics 59.
[13] Abiwinanda N., Hanif M., Hesaputra S.T., Handayani A., Mengko T.R.2018, ”Brain Tumor Classification Using Convolutional Neural Network” Proceedings of the World Congress on Medical Physics and Biomedical Engineering; Prague, Czech Republic. 3–8 pp. 183–189.
[14] S. Vasavi S. Jayapradha V. Srinivasa Rao, 2011”Extracting semantically similar frequent patterns using ontologies” International Conference on Swarm, Evolutionary, and Memetic Computing, Vol. 2, pp. 157-165.
[15] G.Siva Nagaswarao,M.Kiran Babu, T. Srinivasarao, 2019 “ Efficient scheduling Measures for improving real time systems using prediction process” Journal of Computational and Theoretical NanoScience, 16 5.