BRAIN TUMOR SEGMENTATION USING RANDOM FOREST ALGORITHM

Priyanka Prakash Ghadge
PG Student
Department of Electronics and Telecommunication,
JSPM's Jayawantrao Sawant
College Of Engineering, Pune-411028, India.

Abstract: A tumor is a mass of tissue formed by accumulation of malignant cells. The World Health Organization (WHO) defines four grades of tumor. Grade I and II are low grade (LG) tumor while grade III and IV are high grade (HG). This paper presents a system based on a discriminative model used in multimodal MRI tumor segmentation. Early detection of diseases is of the most importance to maintaining or somehow regaining one's health, and thus it contributes to improving quality of life. The combination of various image processing techniques creates an efficient diagnostic tool. One part of the imaging techniques is built around automatic image segmentation, which is much faster than time-consuming analysis by experts. In any case, the earlier the tumor is detected, the better the chances of survival life. In addition to sensitive automatic detection, precise segmentation of tumors is also required for efficient treatment and intervention planning. In particular, brain tumor segmentation consists of separating the different tumor tissues from normal brain tissue. Accurate and reproducible segmentation and characterization of abnormalities can be considered indispensable in medical diagnosis. The main goal of the most efficient treatment of brain tumors is early discovery, identification and diagnosis. Detection of tumor in Magnetic Resonance Imaging (MRI) is very imported which detects tumor automatically. It gives information about abnormal tissues which is helpful for planning treatment. In this paper brain tumor detection done with the help of random forest algorithm (RF)

Index Terms - Random Forest, Brain Tumor Segmentation, Feature Selection, Variable importance.

I. INTRODUCTION

The human body is made from many types of cells and tissues. Each and every cell has a specific function which helpful for human working. These many cells in the body growing an orderly manner and form some new cells. The formation of new cell helps to keep the human body healthy and properly working in time. When some cells lose their capability to control their growth, they growing in unorderly manner. These unwanted cells formed form a mass of tissue which is called tumor and these unwanted tissues when present into brain called as brain tumor. There are tumors which can be benign and malignant. Malignant tumors are most cancer tumor while benign tumors are generally not cancerous but after some time benign tumors will dangerous. A tumor is unwanted cells of malignant tissues. According to World Health Organization (WHO) there are four grades of tumor. When higher is the grade, then more cancerous. When the tumor is having grade I and II are the least less malignant tumors, called as low-grade tumors. LG tumors are usually benign or stating stage of tumors. In this paper we describe the best-performing systems based on discriminative model is used in multimodal MRI tumor segmentation. The important factor in the medical diagnosis includes the medical image data obtained from various biomedical devices that uses different imaging techniques like X-ray, CT Scan, MRI. In this paper we take number of images by internet. Here we will analyze the number and type of features which are used and the classification algorithm applied. The performances of the systems presented and of our own system are compared in the experimental results section. In this project we present a discriminative model for tumor detection from MRI images. The main part of the model is built around the random forest (RF) classifier.

II. OBJECTIVE

Objective of this paper is to diagnosis cancerous cells of human body by using random forest algorithm

III. RELATED WORK

Present state there are many brain tumor segmentation methods that have been developed. These have been implemented and published mainly for the Brain Tumor Image Segmentation Benchmark, organized yearly since 2012.Here Two main categories of model is present these are generative and discriminative models. Generative methods is used to attempt determine the probability distribution. Which shows relative function between the input and the target outputs. These are based on the Bayes theorem and are based on prior knowledge using anatomic properties of tissues. These methods assume standardized data acquisition, registration and alignment in order to be converted into a generally usable probabilistic model. Where discriminative models have ability to learning the classification function directly from a manually labeled training dataset. Here the main drawback is the requirement for a substantial amount of data in order to create sufficiently general and high-performing classifiers via supervised learning. Today's leading architectures in the field of medical image processing and brain tumor segmentation are based on two major methods: the random forest decision tree ensemble and deep learning via convolutional neural networks (CNN) Zikic et al combine a discriminative model using 40 decision trees for taking decision in the classification done with 2000 context-aware attributes, where combining all these with a generative model using tissue-specific probabilities which are used for each patient for identification.

IV. METHODOLOGY

Methodology of the system is illustrated in Figure 4.1. It shows the flow of operation of the system. The first phase shows the data collection process. The data is gathered from human body. This data is then exported to MATLAB environment for further processing and analysis. Random Forest is flexible, easy to use, it is simple algorithm. It avoids over-fitting.

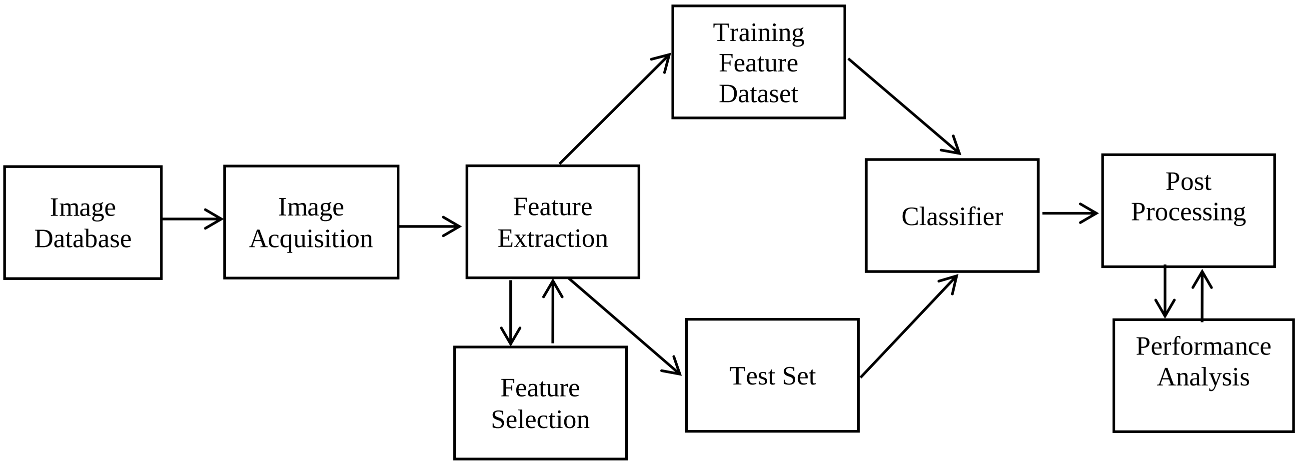

Figure 4.1 Proposed Block Diagram

4.1 Data Acquisition

The database is the highest important step in the discriminative model, which is related to data collection. Taking the acquisition of image samples and the corresponding annotations. The images are taken from internet, it contain four types of images these are T1, T1c (with contrast material Gadolinium), T2 and Flair.

4.2 Preprocessing

Pre-processing and feature extraction are two important steps in brain tumor segmentation. Pre-processing consists of noise filtering and standardization of luminosity and contrast; this means standardization of image pixel intensities. The MRI image consist amount of noise which we should be remove to getting proper tumor image.

In pre-processing process, the MRI images are converted into grey images. Subsequently, the grey images are smoothed using contrast adjustment. Denoising method improves the quality of image which reducing the noise component while improving quality of the image. Decided to use anisotropic diffusion filtering for removing the noise.

Image denoising-Image denoising technique is used to remove noise from images. It is considered as the main technique in the field of image processing in which noise free digital images are developed. In image denoising, the noise in the image is removed by using the image smoothing operation. The noise removal process have quality to eliminate the noise and restore the original image in such a way that getting proper cancerous image. The noise can be present in the form of patterns, dark sports, blur or white sports .The noise is removing with the help of noise on the replacement of pixels. Image denoising is a part of image segmentation which works by removing the hurdles and focusing on the specific objects about which the information is required. There are two methods are used in the image denoising to reduce the noise of an image. First is Spatial filtering-In spatial filters, pixels are set to the accurate size by using the sliding window. The square window used in square, but any type of shapes can be used to removing noise.

Transform domain-The transform domain filter is used to change the signal space so that some processing techniques can be added to the image data. The transform domain filtering consist of wavelet transform and Fourier transform which can used for noise removal process. The sequence for the best segmentation is getting with the help of following process these are bias field correction, after that noise filtering, and finally intensity standardization.

4.3 Feature Extraction

In the tumor segmentation there are many studies helps to find required characteristics with a high correlation to the appearance of the brain tumor in MRI. These research efforts, no proper feature sets have been found yet now. This is the reason for using a large feature set to finding tumor, with having special features and little correlation to the goal of classification. Firstly approach defines a smaller feature set, but for better accuracy this dataset is enlarged for increasing classification performance and quality of image. For each feature we defined many types of low-level characteristics that having ability the intensities in the neighborhood of the voxels studied. In our application following features are used for tumor detection. These are first order and second orders.

First order operators features are (mean, standard deviation, max, min, median, Sobel, gradient)

Standard Deviation

$$S=\sqrt{\frac{\sum(X_i-\bar{x})^2}{n-1}} \qquad (1)$$

Higher order operators are used in modes (Laplacian, Difference of Gaussians, entropy, curvatures, kurtosis and skewness)

Laplacian of Gaussian –To reduce the noise effect the image is smoothed.

$$G(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (2)$$

Kurtosis- Kurtosis is a parameter that describes the shape of a random variable's probability distribution.
It is a measure of the peakdness of probability distribution of real valued random variables .It is standardized fourth center moment of distribution.

$$K=\frac{E(X-\mu)^4}{\sigma^4}$$

$$excess\_k=k-3 \qquad (3)$$

4.4 Classification

The classifier is the main part of system which have ability to find out cancerous and noncancerous image .In the field of data collection there are many well-known classification algorithms are used to identify the tumor, such as Naive Bayes, C4.5 tree, k-NN, k-means, Neural Networks, SVM, AdaBoost, Random Forest (RF). The important classifiers used in this field have been implemented in the WEKA Data Mining Toolkit for detection of tumor. Using this toolkit, compared several classifiers and have chosen to use RF for our application. There are many advantages of RF are, high accuracy, easy handling of large databases, estimating variable importance, computing the proximities between instance, generating the error as forest building progresses. It have ability to handle missing value, it is used to solve both classification and regression problems. The RF classifier was firstly introduced by L. Breiman. RF classifier have a large collection of binary decision trees based on two random processes. These two random process are training data sets and testing data sets. First, the training set is randomly sampled with replacement for obtaining the bootstrap set.
The second process is introduced in the building process of trees. In each and every node only a small part is randomly chosen to identify the features, are used to search for the best split. The training set and bootstrap set have the same size, N and, accordingly, the bootstrap set contains an instance of the training set that is different by approximately 2/3, while the rest is made up of repeated samples. Approximately 1/3 of the training samples are left out from the bootstrap set. These instances form the out-of-bag (OOB) set. Thus, every tree is grown on its own bootstrap set and tested on its OOB set.

4.5 Post processing

Can help to improve classification performance. Performance Evaluation- This is last step of segmentation is evaluation of the results obtained. It can be assessed by the different coefficients like-Dice coefficient, Jaccard, similarity, precision, sensitivity, specificity.

V. RESULTS AND DISCUSSION

Results obtained from random forest transform are as follows

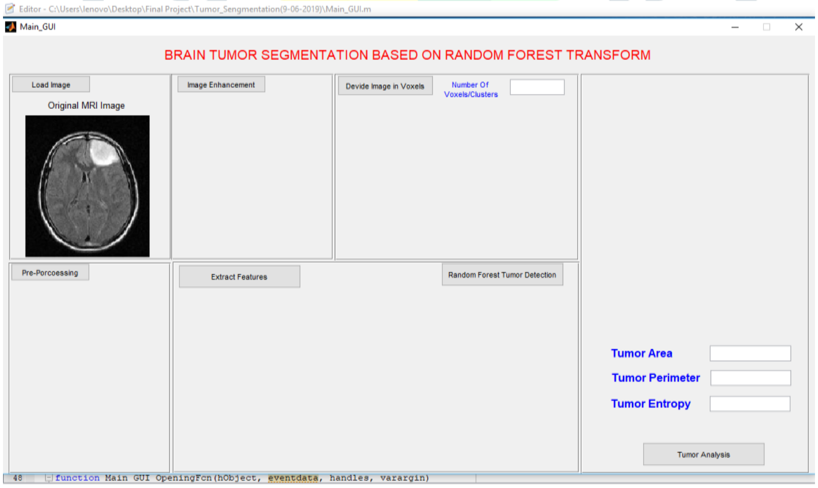

Figure 5.1 Original MRI Images

The above figure 5.1 shows original MRI image for identify brain tumor.

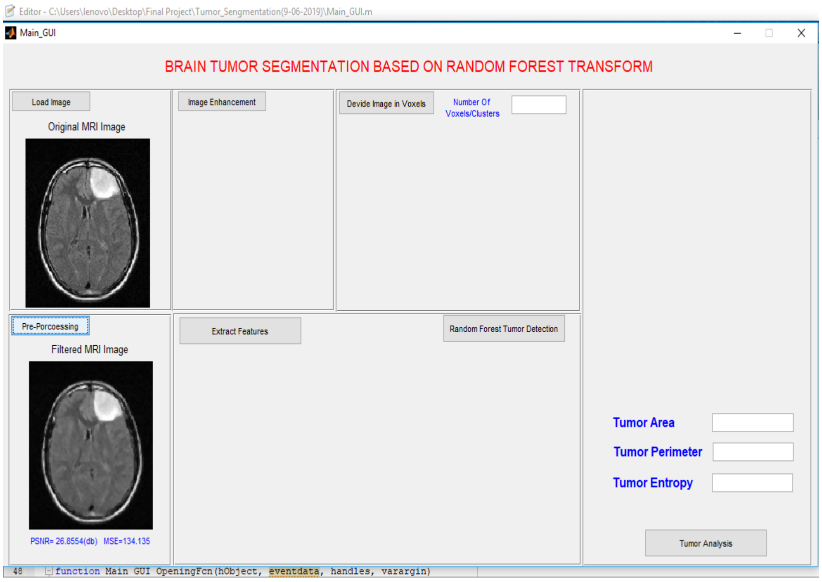

Figure 5.2Filtered MRI image

The above figure 5.2 shows filtered MRI image. Filter MRI image is noise free image. It helps to improve the quality of image. The denoising method improves image quality by reducing the noise component while preserving the quality of the image. Decided to use anisotropic diffusion Filtering. Image denoising-Image denoising technique is used to remove noise from images.

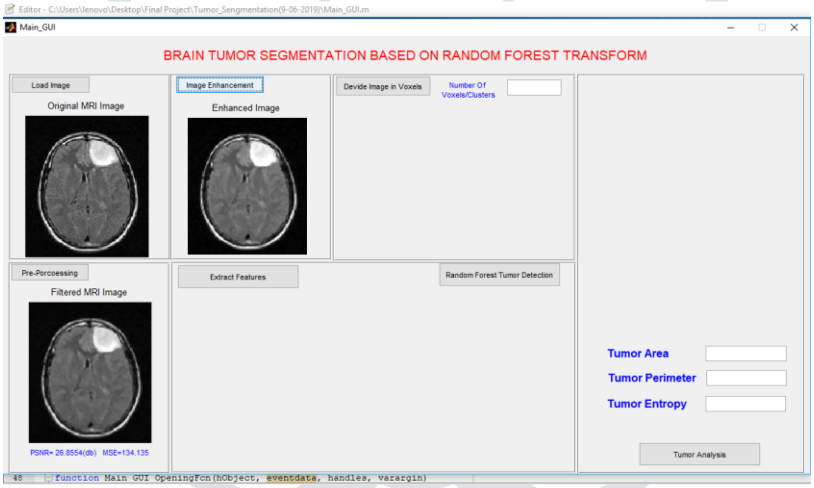

Figure 5.3Enhanced MRI image

Above fig.5.3shows enhanced MRI image. It helps to remove noise, sharpen, brighten an image.

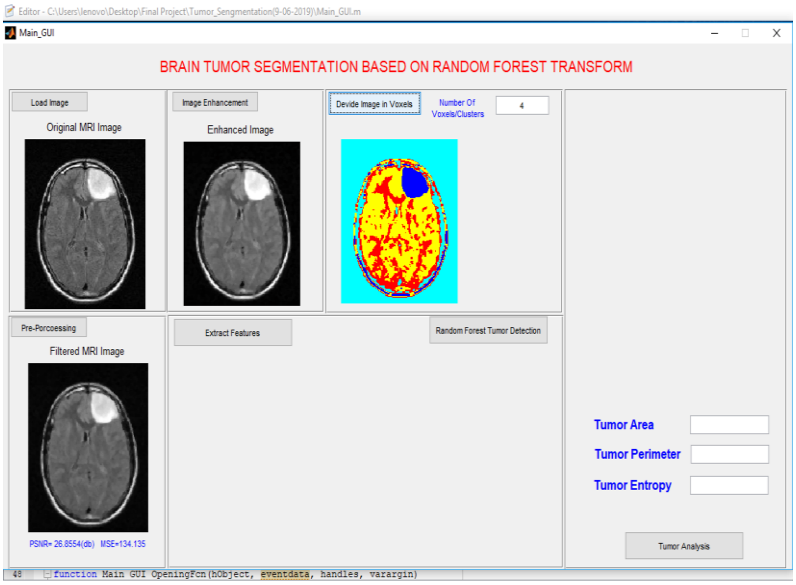

Figure 5.4 Divide MRI image in Voxels

Above fig.5.4shows MRI image into voxels. Voxels are used for divide MRI images into number of clusters in such a way that it identify easily. Color of voxels depends on number of voxels.

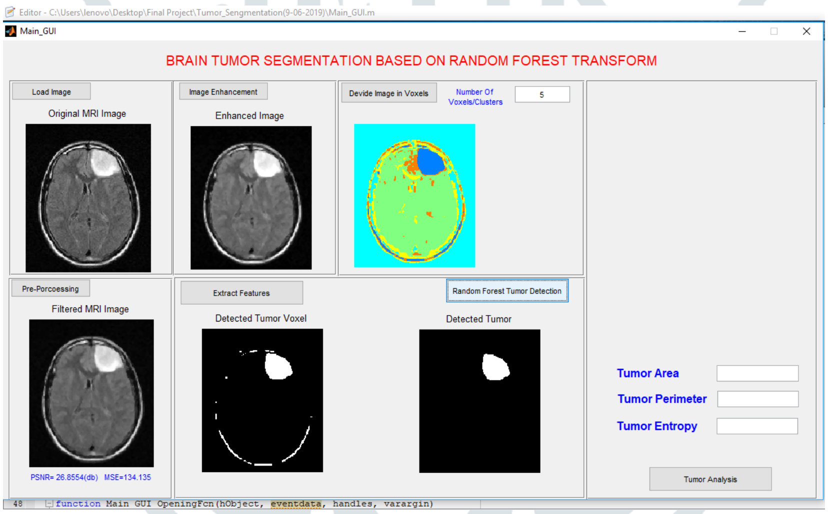

Figure 5.5 Random forest tumor detection and feature extraction

Above fig.5.5showstumor detection by random forest algorithm and feature extraction like area, perimeter, entropy.

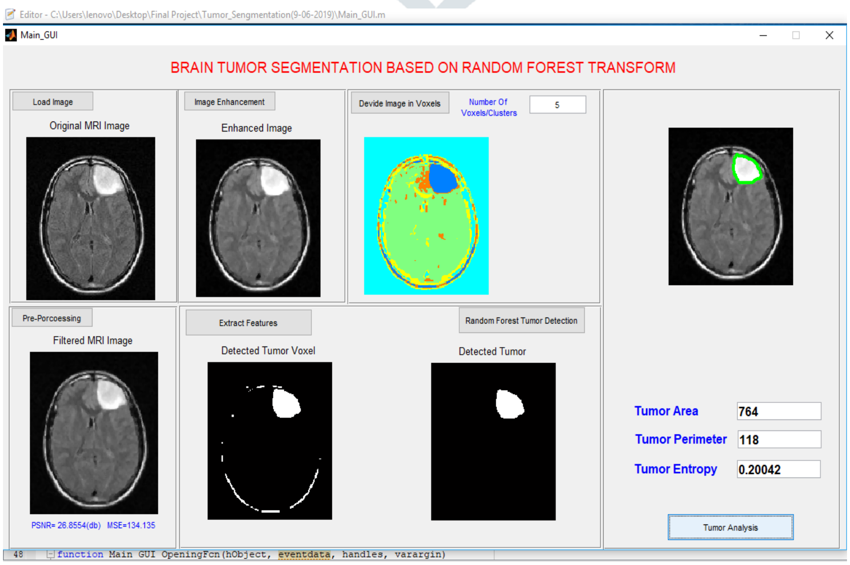

Figure 5.6 Tumor Analysis

Table 5.1 Tumor Analysis

Image	Tumor Area	Tumor Perimeter	Tumor Entropy
1	4915	660	0.91719
2	730	112	0.19349
3	586	112	0.15666
4	1677	304	0.083967
5	1288	173	0.13949
6	1288	173	0.13949
7	57923	6769	0.41293
8	1395	823	0.143992
9	57923	6769	0.41293
10	4300	427	0.23993
11	10941	738	0.2502
12	2092	245	0.17301
13	2208	250	0.18036
14	2286	515	0.42277
15	1956	171	0.15371

VI. CONCLUSION

Different literatures and different techniques are studied for brain tumor detection by random forest algorithm. We have studied of this technique and made survey of literatures. In the various researches, it uses data pre-processing technique and proper classification technique to identify brain tumor in human. In this work we first collect the database for analysis of tumor. Then perform pre-processing and image enhancement, divide image in voxels, feature extraction and random forest tumor detection. We have proposed a feature selection algorithm that can evaluate the importance of new feature sets by comparing them with the existing ones. By this method, we can find the best feature set for the proposed task. In our opinion, the database used considerably limits segmentation performance. Furthermore, the system can be optimized with regard to processing time and efficient memory usage. All these ideas could constitute a meaningful foundation for future research. The important part is the selection of adequate features defines for each voxel. Within this framework, we have proposed a feature selection algorithm that can evaluate the importance of new feature sets by comparing them with the existing ones. By this method, we can find the best feature set for the proposed task. In our opinion, the database used considerably limits segmentation performance. Furthermore, the system can be optimized with regard to processing time and efficient memory usage. All these ideas could constitute a meaningful foundation for future research.

VII. ACKNOWLEDGMENT

There have been many contributors for this to take shape and authors are thankful to each of them. I am highly indebted to Prof. P. A. Kalyankar for their guidance and constant supervision as well as for providing necessary information regarding the research.

REFERENCES

[1] Lefkovits L., Lefkovits Sz., Emerich S., Vaida M., Random Forest Feature Selection Approach for Image Segmentation, under review Eusipco 2016.

[2] Lefkovits L., Lefkovits Sz., Vaida M., An Atlas Based Performance Evaluation of Inhomogeneity Correcting Effects, The 5th Int. Conf. on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, 2015.

[3] Goetz M., Weber C., Bloecher J., Stieltjes B., Meinzer H.-P., Maier-Hein M. Extremely randomized trees based brain tumor segmentation, MICCAI-BRATS Challenge on Multimodal Brain Tumor Segmentation, 2014

[4] Menze B.H., Jakab A., Bauer S., Kalpaty-Cramer J., Farahanik K., Kirby J., et al. The multimodal brain tumor image segmentation benchmark (BRATS). Medical Imaging,
IEEE Transactions on, 2015, 34 (10), 1993–2024

[5] Genuer R., Poggi J. M. Tuleau-Malot C. Variable selection using random forests. Pattern Recognition Letters, 2010, 31 (14), 2225–2236.

[6] GEREMIA E., MENZE B.H., AYACHE N. Spatial decision forests for glioma segmentation in multi-channel mr images, MICCAI-BRATS Challenge on Multimodal Brain Tumor Segmentation, 2012.

[7] http://www.who.int/en/ (Accessed March 2016).

[8] http://www.itk.org/ (Accessed March 2016).

[9] http://www.cs.waikato.ac.nz/ml/weka/ (Accessed March 2016).

[10] J.J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille"Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification."

[11] C. L. Biji, D. Selvathi, and A. Panicker for "Tumor detection in brain magnetic resonance images using modified thresholding techniques"

[12] Njeh, L. Sallemi, I. Ben Ayed, K. Chtourou, S. Lehericy, D. Galanaud and A. Ben Hamida"3D multimodal MRI brain glioma tumor and edema segmentation: a graph cut distribution matching approach,"

[13]Goetz M., Weber C., Bloecher J., StieltjesB.,MeinzerH.-P., Maier-Hein M for "Extremely randomized trees based brain tumor segmentation, MICCAI-BRATS Challenge on Multimodal Brain Tumor Segmentation"