IJCSC 0973-7391 Volume 7 • Number 1 Sept 2015 - March 2016 pp. 202-204 available online at www.csjournalss.com

Brain Tumor Disease Identification Using Random Forest Classifiers

Meet Oza[1],Rupal Kapdi[2]
[1] Student, B. Tech. (IT)
[2] Assistant Professor,
Institute of Technology, Nirma University, Ahmadabad, Gujarat
{13bit033,rupal.kapdi}@nirmauni.ac.in

Abstract: Detection of bodily disease on the bases of magnetic resonance imaging generated is a field that is in rapid growth and with the multiple image processing tools and vivid and varied algorithms available to us, like image processing tool 3D slicer and Random Forest Algorithm. Using the 3D slicer, image is segmented to parts where in the training set a human may detect the tumor and same is done for the dataset that may be provided. The segmented image is then divided into nodes under Random Forest Algorithm. Then by detecting the true and false nodes, the tree is designed based on the selected true nodes. The result of the random forest is then improved using various methods. As Random Forest is robust to noise and refrains from over fitting, it is the prominent choice[13]. In addition, it also offers the possibilities for explanation and visualization of the output.

Keywords: Brain Tumor Segmentation, Machine Learning, Decision Tree.

1. Introduction

Brain Tumor Imaging is done for the ease to detect the tumor to be either malignant or benign. The process of imaging starts with the MRI or a CTScan. The process of MRI uses the Magnetic resonance to capture the image of the brain in a 3d format. The 3D image is in the format ".mha". These files are captured in a "top view", "side view" and "front view". Hence to determine the exact location of the tumor and to recognize the shape and size of the same one requires the generation of segmentation through software which generates the 3D format from the three views given. Once the segmentation is done on the training set by manually marking the infected tumor, the software is ready to take in the new test dataset. It will automatically classify the new tumor infected area and when justified correctly, itcan be used as next generation of training data. The training data used in next population will continue enhancing the population there by increasing the accuracy of the process.

The ensemble method that is used here is the random forest method. Each ensemble method uses a decision tree classifier. Random forest being a forest uses multiple tree formats to generate and enhance the same output for the population of test cases of segmented images. Random forests are usually built using bagging in tandem. It uses random attribute selection. The iteration procedure can be done by the following steps:

For each iterative loop, a training set, $S_{i,}$ of n tuples is sampled with the replacement from S, that is, each $S_i$being bootstrap sample of S hence the tuples may occur more than other samples in the same dataset, also same may remain unselected throughout. Let X be the number of attributes to be used for determining the split at each node, X being much smaller than all the present attributes for the current scenario. To construct the decision tree classifier, D_i, at each node one needs to randomly select the number of available attributes. From the pool of attributes, these randomly selected attribute determine the split in the node, thereby reducing the data set by a ratio for the root node. The method of CART is used to grow the tree. Once grown to the maximum, grown trees are then never pruned. This being the method for Random input selection, called Forest-RI[2]. When the random linear combination for the input is used for the random forest, it is called Forest-RC[7]. Rather than taking the subset of the existing attributes it uses existing attributes and linearly combines them to form a whole new dataset so as to create the new generation and then produce the output for the same. For this the existing features are added to each other with coefficients in the range of [-1, 1]. This may be one on F finite number of selected attributes to generate L number of linear combinations and then the search is made over the same for the best split[11].

2. Method

The process of the whole image segmentation is describes as follows:

The image dataset may have to be obtained prior to the general procedure so as to train the computer for the test cases. Here the dataset taken are the BRATS2015 training dataset in the file format of ".mha"[3]. The slicer tool is a tool that helps in detecting the tumor by image segmentation of the .mha file in a 3D format.

DOI: 10.090592/IJCSC.2016.030

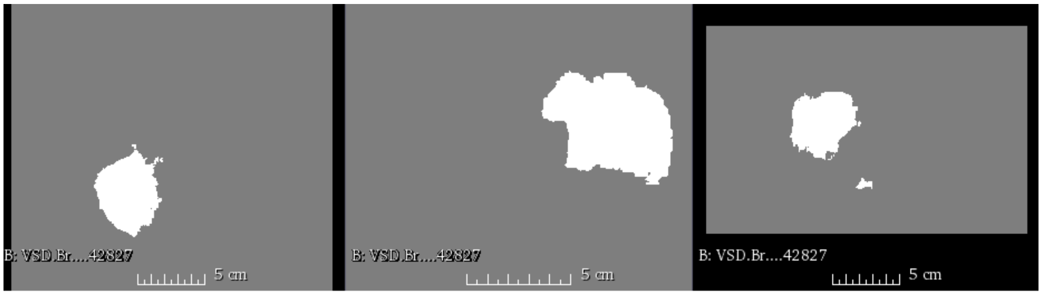

Figure 1. Information of tumor in VSD.Brain_3more.XX.O.OT42827

The above shown picture is for the specific data with image number VSD.Brain_3more.XX.O.OT42827.
The brain image is then segmented into the slicer where the tumor is recognized and the training dataset is prepared for the test data set to be correctly classified into malignant or benign. The 3Dstructure of the given particular image,VSD.Brain_3more.XX.O.OT42827 is shown as follows:

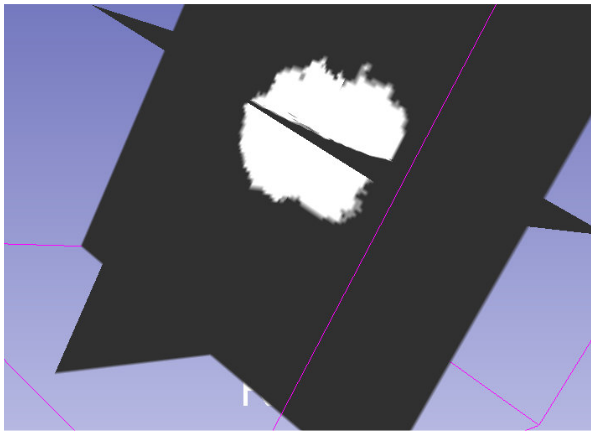

Figure 2. 3D plot of Brain Tumor

Random forest is comparable to ad boost foraccuracy[7]. They are yet more robust to outliers. Chances in random forest for errors are also very low as they being highly decisive in dataflow and tree growth. The random forest tree helps maintain the strength of the individual classes so as not to increase their co-relation. Random forest is not at all sensitive to the individual attributes and features as they are randomly selected for each iteration. Usually up to $\log_2 d+1$ are selected.

3. Enhancement of the Algorithm

Attribute Evaluation[1]**:** Gini Index as the feature evaluation is used for the random forests. It being fast is sure a benefit but has drawbacks when being compared to other heuristics, especially it cannot detect conditional dependencies even when strong one between the attributes. When certain dependencies may arise in any sort of case, the evaluation may result mostly untrue and lead to failed conclusion. Gini Index measures the impurity of the class value distribution before and after the split as per the evaluated attribute. So in gini index the addition of Minimum Description Length(MDL) can be significant to the attribute evaluation property.

Weighted Voting[1]: For all the iterations we need to have the highest and best preferred instances. Similarity of the indexes is also of utmost necessity. More similar the indexes, more high the possibility of the instances producing the continuously updated and refined output with increasing dataset. Hence with weight as per the correctness of the dataset in the case, when selected for new iterations, it shall produce refined output and correct diagnosis of the disease.

4. **Conclusion**

The dataset of images is highly imperfect in case of machine vision hence to evaluate the dataset is even more complicated. The data must be refined and segmented before use and the training dataset should be complete and correct. The test dataset shall be perfectly segmented and the iteration when done withcontinues upper evaluation of the attributes shall make it better and better, improved dataset over time. The weighted voting shall help increasing the input of refined and perfect dataset as iterative population and will enhance the statistics with time. Thereby increasing the correctness of the diagnosis of the tumor being malignant or the benign one.

References

[1] Robnik-Šikonja, M. (2004). Improving random forests. In *Machine Learning: ECML 2004* (pp. 359-370). Springer Berlin Heidelberg.

[2] Shinde, A. B., &Devale, P. R. (2015). Brain Tumor Disease Identification Using GNSWF-Based Feature Extraction and Random Forest Classifiers.*Brain*, *4*(10).

[3] Chang, H., Han, J., Spellman, P. T., &Parvin, B. (2012). Multireference level set for the characterization of nuclear morphology in glioblastomamultiforme. *Biomedical Engineering, IEEE Transactions on*, *59*(12), 3460-3467.

[4]Chang, H., Han, J., Borowsky, A., Loss, L., Gray, J. W., Spellman, P. T., &Parvin, B. (2013). Invariant delineation of nuclear architecture in glioblastomamultiforme for clinical and molecular association. *Medical Imaging, IEEE Transactions on*, *32*(4), 670-682.

[5] Gordillo, N., Montseny, E., &Sobrevilla, P. (2013). State of the art survey on MRI brain tumor segmentation. *Magnetic resonance imaging*, *31*(8), 1426-1438.

[6] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ...&Lanczi, L. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *Medical Imaging, IEEE Transactions on*,*34*(10), 1993-2024.

[7] Havaei, M., Jodoin, P. M., &Larochelle, H. (2014, August). Efficient interactive brain tumor segmentation as within-brain kNN classification. In*2014 22nd International Conference on Pattern Recognition (ICPR)* (pp. 556-561). IEEE.

[8] Boughattas, N., Berar, M., Hamrouni, K., &Ruan, S. (2014, October). Brain tumor segmentation from multiple MRI sequences using multiple kernel learning. In *Image Processing (ICIP), 2014 IEEE International Conference on*(pp. 1887-1891). IEEE.

[9] Tang, H., Lu, H., Liu, W., & Tao, X. (2015, April). Tumor segmentation from single contrast MR images of human brain. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on* (pp. 46-49). IEEE.

[10]Navab, N., Hornegger, J., Wells, W. M., &Frangi, A. F. (Eds.). (2015).*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings* (Vol. 9351). Springer.

[11]Sallemi, L., Njeh, I., &Lehericy, S. (2015). Towards a Computer Aided Prognosis for Brain Glioblastomas Tumor Growth Estimation.*NanoBioscience, IEEE Transactions on*, *14*(7), 727-733.

[12]Gupta, M., Rao, B. P., Rajagopalan, V., Das, A., &Kesavadas, C. (2015, September). Volumetric segmentation of brain tumor based on intensity features of multimodality magnetic resonance imaging. In *Computer, Communication and Control (IC4), 2015 International Conference on* (pp. 1-6). IEEE.