International Journal of Engineering & Technology, *7 (3.29) (2018) 146-149*

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Segmenting and classifying MRI images for brain tumors using CNN

Niharika Sharma [1] *, Prakhar Mishra [1], K. S. Sadhana [1], P. Nithyakani [1]

[1] Dept. of Information Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203
**Corresponding author E-mail: prakhar.mishra@hotmail.com*

Abstract

Gliomas are one of the most prevalent and aggressive form of brain tumours in the world. Patient's usually go on to live a very short life after the initial diagnosis. Therefore, it is crucial to successfully and quickly outline a method for diagnosing the same in it's very earliest stages.Magnetic Resonance Imaging, or MRI as it is more frequently called is a noninvasive method of imaging parts of human anatomy. MRI's utilise robust fields of magnetism, along with waves that have frequencies corresponding to the radio waves in the spectrum to develop precise pictures to get a sense of the happenings inside the human body. The current, most widely used method of diagnosis for brain gliomas involves an oncologist or radiologist reading the MRI image and using his knowledge and experience regarding the same to reach a diagnosis. However, this manual method of diagnosis is very tedious and has been prone to errors in the past. Therefore, it essential to develop an automatic method for the same.Most of the techniques used currently for segmenting brain tumours were initially developed for other diseases, the most common use among them being the separation of white matter lesions. Most of the current methodologies can be broadly categorised into two families- 1.General Probabilistic Methods- Probabilistic methods are a remarkable method to establish the validity of combinatorial entities with distinct characteristics. Although the basis of their existence lies in probability, they are not bounded by it and can be used to solve and evaluate theorems across different branches of Mathematics.2.Discriminative Approaches- They are also sometimes referred to as Conditional Models. We utilise CNN's for faster and accurate processing of the data.

Keywords: *MRI's; CNN; Contrast; Contours; ANN.*

1. Introduction

We can chronicle gliomas as the most prevalent form of brain tumours with the one of the lowest post-diagnosis life-spans, usually not more than 12-15 months. They can be characterised into two distinct categories- High Grade Gliomas (HGG) - the more malignant and the Low Grade Gliomas (LGG) - which are considerably benign and easier to treat than HGG. Surgery and chemotherapy are the most common treatment tracks for patients after diagnosis. The current, most widely used method of diagnosis for brain gliomas is a "Rough Measure", which involves a specialist physician reading the MRI image and using his knowledge and experience regarding the same to reach a diagnosis. But, this manual method of diagnosis is very tedious and has been prone to errors in the past. This emphasises the need for a semi-automatic or automatic method for diagnosis. However, it can be an arduous procedure, as MRI scans for different patients are highly variable when it comes shape, borders and location of these tumours and MRI image processing also bears the problem of the having the same image sequence with different ranges of intensity. Brain tumour segmentation uses probabilistic models, which mostly include a function for likelihood that is applied across the entire data. To lay it out in non-technical language, tumours are considered outliers of the given data, as they are inaccuracies, or more clearly abnormalities of what is wrong with the actual underlying tissue. Some of the other techniques that they suggest involve the usage of probabalistic approaches, which in the case of segmentation of brain can be described as "the separation or segmentation of images using atlases, so as to normalise the variable size and location of the tumours. In this project, we investigate the possibility of deep Convolutional Neural Network with small-sized kernels with the objective of segmenting out gliomas from a patient's MRI scan. The advantage and reasoning behind using small kernels is that, using smaller sized kernels we can increase the number of layers of convolution, and at the same time have the effective field of reception of bigger kernels. As an example, we can consider a 3X3 kernel, which will have the same receptive field as a 7X7 kernel, but will have much lesser weights, hence increasing g the speed of processing. This in turn has an added benefit of making the model less likely to overfit to the testing data and applying more linearities. We understand that there is a large amount of variability across different MRI's of brain tumours and have hence applied thorough Bias Field Correction and Intensity Normalisation techniques during pre-processing, so as to enhance the correctness of the data used for training and hence, as a result, of the neural network itself.

2. Methodology

The proposed method consists of a tumour segmentation part, feature extraction part and finally the classification part. All the modules are explained below in each block.

a) Bias Field Estimation

The first step we have to do in the process is to reduce the noise in the image. Noise can be removed by Bias Field Corrected Fuzzy C-Means method. The method is a proposed in [18] which will be used to reduce the noise content in the image The pixel intensity of MRI in identical cumulative distribution values can be reasoned to variations in the radio-frequency turmoil or to difficulties related to the acquiring sequences. Results obtained consists of a gradually shifting shading antique over the picture that might result in erroneous data with the ordinary distinction based on the variations of intensity.

Target capacity of the standard FCM calculation is selectively altered to enable the labelling of a pixel for identical cumulative distribution values allowing it to get affected by the tags in its adjacent surrounding. The surrounding effect regularises as well puts forth the bias implementation to a piecewise-homogeneity tagging. Regularisation as these are useful for the implementation of tumour segmentation methods. The mathematical equation for the bias estimation is given below

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{p} \| x_{k-v_i} \|$$

The method is used to estimate the noise in the image and then we would subtract the noise from the noisy images. By subtracting we would be able to reduce the noise from the original noise.

b) Gliomas Tumour Segmentaion

The first part of segmentation is the thresholding. After thresholding we use the method which is used in [16], [17] T F Chan method. Here we use this active contour method to detect objects in a given image, based on techniques of curve evolution. This method would first use the threshold image to plot the edges and then uses the iteration method to identify the border of the segmented images.

After a specific iteration the border would be extracted and then the coordinates of the area would be noted down. With those coordinates we would create a matrix with the binary value and then the binary matrix will be multiplied with the MRI image. The image contains only the tumour part and the area is cropped using the bounding boxes.

c) Extracting Features

Machine learning is a field of study wherein extracting features begins from a preliminary set of data which later develops value-set pairs for adding meaning to the dataset. This leads to an environment rich in developing effective and non-repetitive results on data. The concept of extracting features is to highlight ID of all classes.

At a point when the information to be calculated is greater to process and it is suspected to be repetitive, it can be changed into an arrangement of features which can likewise be renamed as highlight vector. Deciding an arrangement of the underlying feature is called selection of features. The features are relied upon to contain the diverse data from the information, with the goal that the coveted assignment can be performed by utilising the decreased portrayal rather than the entire information.

Gray Level Co-occurrence Matrix-Extracting Features: GLCM grid is a method used for analysing images and calculating the extent to which their features can be harnessed with the surface component extraction. For example, let's take a sample of image that is composed of different proportions of pixels with each proportion having a significant degree of mixes of dim levels. The tabulated columns are prepared in order to track the similar proportions of mixes of a particular section in an image. The scale of composition of an image can be calculated using the contextual data saved in the grid which gives a calculative idea of the distribution of the mixes of dim levels.

Algorithm

The feature is created in the following

- Consider an image v to be extracted for features.
- Consider the image to be surrounded by a list of patches X that have the image as the center of median with a specific dimension.
- Let us consider that the only existing data is X, we now initialise p, q parameters of the grid which are the level of occurrence of the intensity p and intensity q. They together have an inbuilt relationship amongst them. The addition of the attributes contributing to p and q of the grid will add up to a cumulative figure that signifies the association between p and q in X.
- Create a transposable grid that could be used with the co-occurrence in grey level in the form of a matrix. Join this grid to its previously formed grid in order to form a symmetrical grid wherein we have the association between p and q closely resembling to the association between q and p. After implementing the steps mentioned till this point, the parameters will add up to double the cumulative figure observed in X. Hereafter, for any value of p, the parameters of p and q along with the mentioned p will add up to a cumulative figure having intensities of p being displayed in the patch association with another sample of data.Shape Feature Extraction: For feature value we calculate descriptors.
- The descriptors that highlights the physical features of an image can be significantly divided by their uniformity in accordance to the transitions that are specified in any physical feature. Of them, quite a good numbers of descriptors are uniform in accordance to their consonance which implies that consonant features will possess a similar set of values of their descriptors.
- Such set of descriptors are known to display no irregular behaviour when tested with various isometric engrained values related to their physical features. The upper hand of having these descriptors revolves around the fact that they can be accommodated in various distorted entities due to a simple reason that these distortions don't possess the ability to stretch as one of their character traits. These descriptors are primarily based on the calculation of length running through the top layer of an entity. It could also base on any other entity possessing uniform isometric behaviour.
- Apart from the descriptor that has been explained in the paper, there exists other various types of descriptors relying on the physical state of an object. These descriptors are based on plots that contains a median axis in the center. As an example, there are plots that record the data of those properties of geometric configurations that are unaltered by any distortions. But these plots are comparatively less effective than descriptors which displays the similar physical feature that can be built through numbers and figures. Hence, by reviewing all the points mentioned above, we come to a conclusion that every descriptor has a specific role to play and a specific task to be implemented than the other. Which brings us to the point that in every case of feature analysis, descriptors of favourable properties must be chosen.

d) Classification

Convolutional neural network:

CNN's as they are commonly referred to, are in the field of image recognition defined as those types of ANN's that are modelled after the human brain's anatomy and try to mimic it's operational functionality. Millions of neurons are connected to each other and transfer and receive responses, in a manner that closely resembles a chain reaction. The most well defined neural networks are those that have a well defined hidden layer, i.e. the number of layers that lie between the input and the output layers and are the processing centres of the CNN.

CNN's have successfully proved their merit across various branches of computing. The working of of the program is dependent on the characteristic picturisation as maps. They are referred to as feature maps and are connected to the properties of the previous kernel in the network. They use the method of backpropogation to try to train the values of the network. In this method, the weights are calculated at every stage of the neuron. The final weight is calculated at the output layer and the error is calculated at the last node. This error is then 'back' propagated across the network and the weights are adjusted accordingly so as to minimise the final error. The advantages of using back propagation is that it allows for the program to not over and under fit to the data. Convolutional Neural Network (CNN), containing 4x4 kernels is intended to investigate a more profound design with a deeper architecture as well as having a constructive outcome against overfitting. The primary responsibility of the system is to identify two different grades of tumours, namely HGG (High-grade gliomas) and LGG (Low-grade gliomas) through separate CNN architectures. Both the categories of gliomas tumours require distinct operability in order to treat the patient suffering from the ailment. During the process, the statistics of training sets are falsely increased by pivoting the training sets. To fuel the process, the samples of high grade gliomas are used to augment the frequency of infrequent low grade gliomas classes. An added advantage of the back propagation method is that the program becomes relatively simple to train to the data.

We use multiple layers for the convolution phase and then place them on top of each other, in what resembles a stack. The initial layers of the network greatly augment their characteristics such as dimensions, and are formatted in the succeeding layers. Some of the techniques that are important for the project are as follows:

- Assignment of data: A crucial step to achieve normalisation of data across the dataset. We use multiple steps for pre-processing including fragmenting and segmenting out essential characteristics from the block of data made available to us as training data. We assign the processed data to another variable which is then called when the network is trained.
- Data Transfiguration: For transformation of data of a form that is not linear we use RELU. The function is stated below:

$$f = \max(0, x)$$

We discovered that more accurate outcomes were obtained when particular function where used, the most successful of them being the sigmoid function which showed outcomes with minimal delta error while pre-processing. We check the integral value of each node at this stage and all negative values, are assigned the value of zero. This can however have disastrous effects on the outcome of the program as it would destroy the normal supply of gradients and cause complications during the reassignment of weights. This hindrance is avoided by using one of the other forms of the RELU described previously, called the Leaky RELU. The added advantage of using this unit, as the name suggests, is that it makes the entire function 'leaky', that is it adds a small value lesser than zero to the function, thus introducing a tiny slope which ensures that the gradient remains continuous.

We describe it as following:

$$f = \max(0, x) - \alpha\max(0, x)$$

- Aggregating: As the name suggests, it involves the aggregation of adjacent characteristics in the map. This ensures that the resources are made available to the program for future stages. It ensures that all the data is less cluttered and that the program isn't very sensitive to small changes in values, thus making it an essential process for increasing the efficiency of identification.
- Normalisation: The training data used in the project was very large. Like all datasets, there happen to be some unique nodes which have values at either end of the spectrum and do not correctly mimic or reflect the feel of the other points in the data sets. These values are called as 'Outliers' and they can have a catastrophic effect on the outcome of the program as they can heavily skew the balance of weights in the neural net. It is better that such points of interest are identified during the pre-processing step and are normalised, this ensuring a uniformity over the data, which would be greatly helpful during predicting the outcome. Our normalisation method aggregates the remaining data closely and hence allows us to have a clear point for concentration of values.
- Supplementation: We use this step to expand the amount of the data that is available to us for training. With a smaller number of nodes, we have a higher chance of the program overfitting to the available data, this step greatly reduces the risk for such an outcome. We did this by rotating the original image through angles of 30, 45, 90 and 120 and saving them as separate images in the training data. This allowed us to increase the size of our dataset by five time while at the same time provide the classifier more images to learn from and thus greatly reduce over fitting.
- Fallacy Function: As the name suggests, it is the function that is due to calculate the mean error during training. As previously explained, during back propagation, the error between the actual and estimated values is calculated at the node during each step of the training. The neural network them aims to minimise this error through each iteration and uses the fallacy function to calculate the same. This function is defined as follows:

$$H = \sum_{j=voxels} \sum_{k\in classes} c_{i,j} \log(c_{i,j})$$

- Model Architecture: Present model for brain tumour segmentation has adopted the methodology that goes by the name BRATS i.e "Multimodal Brain Tumour Image Segmentation Benchmark".

BRATS has dependably been concentrating on the assessment of cutting edge techniques for the division of mind tumours in multimodal magnetic reverberation imaging(MRI) checks. BRATS uses multi-institutional pre-agent MRI filters and spotlights on the division of characteristically heterogeneous such as history, shape and features of cerebrum tumours, to be specific gliomas. Furthermore, the clinical relevance of this segmentation process centers around the forecast of patient general survival.

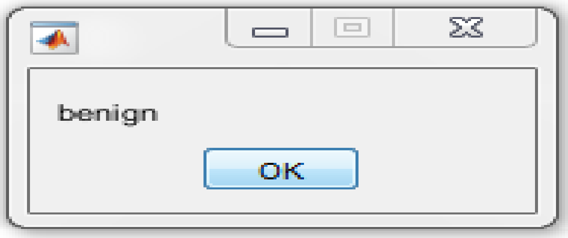
In this method, the algorithms for tumour division depends on systems initially produced for different structures or diseases, most eminently for robotised white issue sore division that has achieved impressive exactness. Meanwhile numerous advances have been tried for their pertinence to mind tumour location and division. We can classify most current tumour division techniques into one of two expansive families of , one of general probabalistic methods and the others that fall under discriminative approaches.

These methods unequivocally build up a parameterized or non-parameterized model for the hidden information. The models discovered more often do not exclude a probability of work relating to the perceptions and an earlier model. From the variations observed by the norm, tumours can be sectioned as exceptions of ordinary tissue, attributed to shape and availability compels. Different methodologies depend on probabilistic chart books. Assessments done quantitatively uncovered extensive difference in fragmenting different tumour sub-districts, showing the trouble of this errand. Later it was found that different algorithms worked best for different sub-regions, but that no single algorithm ranked in the top for all sub-regions simultaneously.

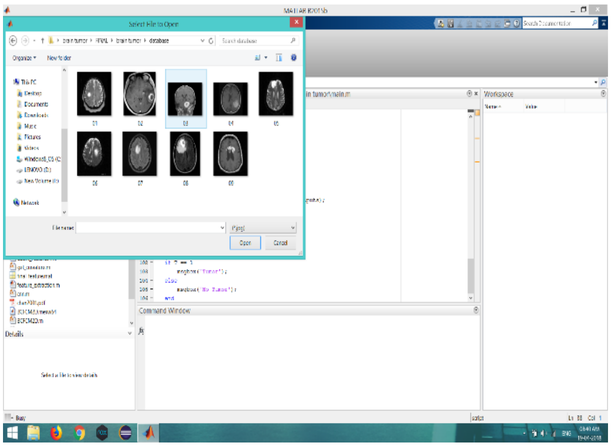
- Output

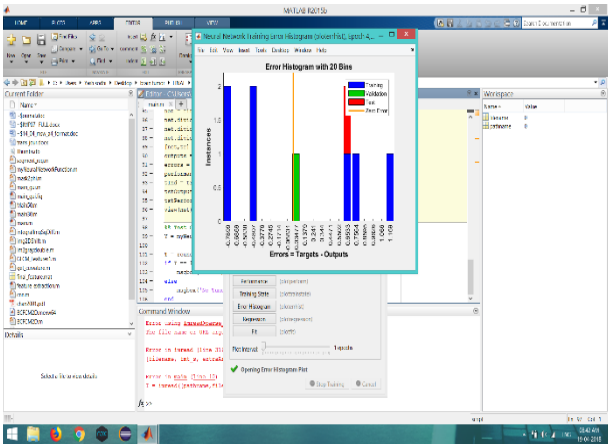
The neural network would process the feature values and predicts the output. The result for the tumour classification is given in the Section III.

3. Results



benign





In this paper, we try to use a CNN model for identifying and segmenting brain tumours from MRI scans.The method proposed by us gave highly accurate results. The corresponding graphs for performance validation and error rates depict the same. The accuracy of a properly designed CNN increases as it is fed more training data, and hence the performance of the model proposed in the paper can be increased in the same way, by allowing it to have more training data to learn from.\\ This paper proposes an answer for enhancing the ability of a processor to accurately carry out diagnosis of brain gliomas, which has been a problem that has plagued the medical-tech committee for a while now. In the close element, we expect that we can incorporate our solution, specially within the rural population, a large majority of whom still don't have access to qualified medical personnel.

References

[1] S. Bauer et al., “A survey of MRI-based medical image analysis for brain tumor studies,” Phys. Med. Biol., vol. 58, no. 13, pp. 97–129, 2013.

[2] D. N. Louis et al., “The 2007 who classification of tumors of the central nervous system,” Acta Neuropathologica, vol. 114, no. 2, pp. 97–109, 2007.

[3] E. G. Van Meir et al., “Exciting new advances in neuro-oncology: The avenue to a cure for malignant glioma,” CA, Cancer J. Clinicians, vol. 60, no. 3, pp. 166–193, 2010.

[4] G. Tabatabai et al., “Molecular diagnostics of gliomas: The clinical perspective,” Acta Neuropathologica, vol. 120, no. 5, pp. 585–592, 2010.

[5] B. Menze et al., “The multimodal brain tumor image segmentation benchmark (BRATS),” IEEE Trans. Med. Imag., vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[6] N. J. Tustison et al., “N4ITK: Improved n3 bias correction,” IEEE Trans. Med. Imag., vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[7] L. G. Nyúl, J. K. Udupa, and X. Zhang, “New variants of a method of MRI scale standardization,” IEEE Trans. Med. Imag., vol. 19, no. 2, pp. 143–150, Feb. 2000.

[8] M. Prastawa et al., “A brain tumor segmentation framework based on outlier detection,” Med. Image Anal., vol. 8, no. 3, pp. 275–283, 2004.

[9] B. H. Menze et al., “A generative model for brain tumor segmentation in multi-modal images,” in Medical Image Computing and Comput. - Assisted Intervention-MICCAI 2010. New York: Springer, 2010, pp. 151–159.

[10] A. Gooya et al., “GLISTR: Glioma image segmentation and registration,” IEEE Trans. Med. Imag., vol. 31, no. 10, pp. 1941–1954, Oct. 2012.

[11] D. Kwon et al., “Combining generative models for multifocal glioma segmentation and registration,” in Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2014. New York: Springer, 2014, pp. 763–770.

[12] S. Bauer, L.-P. Nolte, and M. Reyes, “Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization,” in Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2011. New York: Springer, 2011, pp. 354–361.

[13] C.-H. Lee et al., “Segmenting brain tumors using pseudo-conditional random fields,” in Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2008. New York: Springer, 2008, pp. 359–366.

[14] R. Meier et al., “A hybrid model for multimodal brain tumor segmentation,” in Proc. NCI-MICCAI BRATS, 2013, pp. 31–37.

[15] R. Meier et al., “Appearance-and context-sensitive features for brain tumor segmentation,” in MICCAI Brain Tumor Segmentation Challenge (BraTS), 2014, pp. 20–26.

[16] T. Chan and L. Vese, “Active contours without edges,” IEEE Trans. Image Process. vol. 10, no. 2, pp. 266–277, Feb. 2001.

[17] Minimizaion of region-scalable fitting energy for image segmentation. IEEE Trans. Image Processing, vol. 17 (10), pp.1940-1949, 2008.by Li chunming etc.

[18] M.N. Ahmed ET. al. "A Modified Fuzzy C-Means Algorithm for Bias Field estimation and Segmentation of MRI Data" 2002, IEEE Transactions on medical imaging.