

# Brain Tumor Detection and Classification in Magnetic Resonance Imaging (MRI) using Region Growing, Fuzzy Symmetric Measure, and Artificial Neural Network Backpropagation

Lugina Muhammad \*1, Retno Novi Dayawanti, SSi., MT. \*2, Rita Rismala, ST., MT. \*31

\* School of Computing, Telkom University Bandung, Indonesia

<sup>1</sup>lugina21@gmail.com
<sup>2</sup>retnonovi@telkomuniversity.ac.id
<sup>3</sup>ritarismala@telkomuniversity.ac.id

#### Abstract

Brain tumor is one type of malignant tumors that occurs because there is an abnormal and uncontrolled cell division activity. There are several ways to diagnose brain tumors, for example use MRI images. Through the MRI images, the radiologist can see the brain anatomy without performing surgery. However, this process is still done manually and could lead to misdiagnose. In addition, the different characteristics of brain tumor makes the diagnose more difficult. Therefore, we need a system of Computer-Aided Diagnostic (CAD) that will help radiologist in identifying brain tumors.

In general, the CAD system consists of two major processes, namely image segmentation and feature extraction and classification. One example of segmentation is Region Growing that will classify the pixels based on certain criteria. However, the manual selection of seed point is a drawback of this method. The examples of feature extraction methods are Fuzzy Symmetric Measure (FSM), and First and Second Order Statistics. FSM values can be used to calculate the symmetry of the image brain, while the first and second order to represent feature in the image. As for the classification process, Artificial Neural Network Backpropagation method is widely used for its ability to resolve nonlinear dan complex problems.

This research implements CAD system that uses Region Growing, Symmetric Fuzzy Measure, and Backpropagation Neural Network for detecting and classifying the brain tumors. In addition, the modification of converging square is conducted to select a seed point automatically. After testing, the system generates a 100% accuracy and BER is 0 in the case of distinguishing between normal and tumor brain. Besides, the average accuracy in classifying the types of brain tumors achieved 89.72%, the BER 0.1 for training data, and the average accuracy of 84.44%, BER 0.16 for the testing data.

## **Index Terms**

e-commerce, Collaborative Analysis of Requirement and Design, User Centered Design, Information

# I. INTRODUCTION

# A. Background

The development of technology in processing digital image and artificial intelligence hold an important role in medical field. The use of Magnetic Resonance Imaging (MRI) by radiologist for identifying anatomy structure and diagnosing brain tumor is one of the examples [2]. Brain tumor is one type of malignant tumors that occurs because there is an abnormal and uncontrolled cell division activity.

Through this MRI image, radiologist can diagnose brain tumor without doing surgery. However, the diagnosis is conducted manually based on the radiologist interpretation. Besides, this process is also considered as a tough work for its complex characteristic of brain tumor [3]. That is why; Computer Aided Diagnosis (CAD) is needed to assist the radiologist in diagnosing the MRI image of the brain.

Generally, CAD system consists of two large processes; image segmentation and image feature extraction and classification. Region growing is an example of the image segmentation method which is classifying the pixels based on certain criteria. However, the manual selection of seed point is considered as a drawback of this method.

Furthermore, the sample methods from the feature extraction are Fuzzy Symmetric Measure (FSM) and First and Second Order Statistics. FSM values can be used to calculate the symmetry of the image brain [4]. While the First and Second Order to represent feature image. As for the classification process, Artificial Neural Network Backpropagation (ANN) method is widely used for its ability to resolve nonlinear dan complex problems. However, ANN has a weakness in the problem of determining the values of the optimal parameters. There is no definitive guide to determine the best parameter values.

# B. Research Problems

This research focuses on four problems:

- 1) How is the way to do the automatic selection of the seed point in the Region Growing?
- 2) How does threshold value influence Region Growing?
- 3) What is the exact threshold value of FSM so that the system can distinguish between normal brain and tumor brain?
- 4) What is the optimal value of ANN Backpropagation in order to be able to classify the type of tumor brain?

# C. Purposes

The purposes of this research are:

- 1) To modify the algorithm of Converging Square with Split and Merge to conduct the automatic selection of seed point in Region Growing.
- 2) To analyze the influence of Threshold Region Growing value towards the segmentation result.
- 3) To gain the exact FSM Threshold value to be able to distinguish between normal brain image and tumor brain image.
- 4) To attain the optimal values of ANN Backpropagation parameter so that the system can classify the type of tumor brain.
- 5) To analyze the performance of Region Growing, Fuzzy Symmetric Measure, and Artificial Neural Network Backpropagation in detecting and classifying the type of tumor brain.

# **II. BASIC THEORIES**

## A. Brain Tumor

Brain tumor is an intracranial tumor that occurs because there is an abnormal and uncontrolled cell division activity. These cells may be derived from the brain itself or spread from cancer that is located in other organs (metastatic tumors) [7]. World Health Organization (WHO) classifies brain tumor into four grades; those are grade I, grade II, grade III, and grade IV. Tumor that goes into grade I and II is a benign tumor, while for grade III and IV is a malignant tumor.

## B. Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is an imaging technique in medical field. It is used to create picture of organs and internal structure of body by using magnetic field and radio wave. It is also used to detect various medical disorders that occur in the body such as tumors, bleeding, or wound.

## C. Digital Image Processing

Digital Image Processing is a technique of manipulating image into another image with a better quality.

1) Digital Image Enhancement: The aim of Image Enhancement is to process the image to become more suitable for certain application compared to the original one. An example is by doing the image filtering using median filter.

2) *Region Growing Method:* Region Growing is a procedure to process the image segmentation through classifying the pixel into a region based on certain criteria [8]. These criteria might possibly be the value of light intensity, gray, texture, or color value. The works of this method are:

- a) Determine the pixel of seed point as the first point in developing a region.
- b) Check the pixels which are close to seed point. If it has the same property with seed point, then insert the pixel into region.
- c) Make the pixels with the same property as a new seed point.
- d) Repeat the b step until there has no other pixel can be added.

3) Split and Merge: Split and Merge method is a way of image segmentation by dividing the image into four same sizes, then each of the section will be split and merge to fulfill certain condition [8]. The general works of Split and Merge algorithm are:

- a) Divide the image into four sections of same  $R_i$ , thus,  $P(R_i) = \text{FALSE}$  for each condition.
- b) Combine the Rijand Rik if the condition of  $P(R_{ij} \bigcup R_{ik}) = \text{TRUE}$ , meanwhile if the condition value is FALSE then split each section into four same subsections.
- c) Stop when there is no more section can be divided.

INTL. JOURNAL ON ICT VOL. 1, ISSUE. 1, DECEMBER 2015

#### D. Fuzzy Symmetric Measure

It is a value used to count the symmetry of Cerebrospinal Fluid (CSF) image in brain image [4]. The value of the symmetry can be counted using:

$$FSM = \frac{1}{1 + (\frac{n_L - b_R}{100})^2} \tag{1}$$

 $n_L$ = The number of white pixel in the left side of image  $n_B$ = The number of white pixel in the right side of image

#### E. Feature Extraction Statistic

Feature statistic shows texture characteristic from an image [6]. There are two methods in taking the value of the texture characteristic itself; the First and Second Order Statistic.

The first order takes the feature based on histogram image. Some parameter features are mean, variance, skewness, kurtosis and entropy.

Meanwhile, in the second order statistic, the probability of relationship between two pixels with certain distance and side orientation is used to get the texture feature. Some values that can be taken are angular second moment, contrast, correlation, variance, inverse different moment, and entropy.

## F. Artificial Neural Network Backpropagation

Artificial Neural Network Backpropagation is one of artificial neural networks with supervised learning. This ANN uses propagation algorithm to reduce the error value. The error value is beneficial for setting the weights of each neuron in backward propagation process; started from the output layer to the input layer. The purpose of this setting is to make the ANN to be able to recognize and adjust the resolving problem.

# III. DESIGN AND IMPLEMENTATION OF SOFTWARE

Brain tumor detection system is divided into two parts generally, that are processing of MRI image and clarification using JST-Backpropagation. The system designed in this research in generally can be seen in Figure 1.



Fig. 1: Flowchart systems for detecting the type of tumor brain

#### A. Image Data Collection

Image data of brain MRI is taken from internet public repository. The image of normal brain is taken in CASILab at The University of North Carolina at Chapel Hill. Meanwhile, the image of tumor brain with Lower Grade Glioma or Glioblastoma Multiforme is available in The Cancer Genome Atlas (TCGA) Portal Data.

The number of the used data for each MRI image of normal brain, malignant tumor, and benign tumor is 27 images. Specifically for malignant and benign tumor, the data is divided into two; training data and testing data which are used in ANN Backpropagation process. The data is taken randomly using cross folding with 18 training data and 9 testing data, so that the data is more representative.

#### B. Image Processing

The first process is conducted to transform the MRI brain image to make it more suitable for the next process.



Fig. 2: Brain image before the resetting of direction orientation (a) Brain image after the resetting of direction orientation (b)

1) The conversion of Image to JPG and Image Orientation Setting: The image is converted from .mha format and .dicom format into .jpg by using YAKAMI DICOM Tools application. After the conversion, the size and direction resetting is conducted in order to have the same size and direction. The process is done manually by using Photoshop CS5 application with  $165 \times 200$  pixels.

2) The Conversion of Image from RGB to Grayscale: Furthermore, the image is modified from RGB color into grayscale. It aims at making the three of images color channels become one color channel so that it will be easily processed using Region Growing algorithm.

3) Image Enhancement Using Median Filter: The Image enhancement is conducted by using the median filter with mask 3x3size. The improvement of image quality is useful for dismissing the noise of image.

#### C. Image Processing

The first process is conducted to transform the MRI brain image to make it more suitable for the next process.

1) The conversion of Image to JPG and Image Orientation Setting: The image is converted from .mha format and .dicom format into .jpg by using YAKAMI DICOM Tools application. After the conversion, the size and direction resetting is conducted in order to have the same size and direction. The process is done manually by using Photoshop CS5 application with  $165 \times 200$  pixels.







(b)

Fig. 3: the brain image before the enhancement (a) Brain image after the enhancement using median filter (b)

2) Sharpening the Image Contrast: Changing the contrast and brightness value is done in this process as to get a clearer visual of tumor area. The contrast value is 0.8 and brightness value is 0.2 to get a sharper image.

#### D. Segmentation with Region Growing Algorithm

The image segmentation using Region Growing technique is conducted. It works by comparing the pixel intensity value of the seed point with all around them. However, the Region Growing algorithm has a drawback in selecting the seed point; it should be done manually by users. That is why; the modification towards one of the selection algorithm on the seed point is taken as the solution for this drawback; that is through converging square [5] based on split and merge method.



Fig. 4: the illustration of Seed Selection algorithm which split the image from k size into subsections with k/2 size

1) Seed Selection Algorithm: The modification of converging square algorithm is done by changing the image split process based on the split and merge method. Firstly, the MRI image with  $200 \times 165$  pixels is enlarged to  $256 \times 256$  so that the image can be split into 4 subsections with  $k/2 \times k/2$  size. (Figure 4) Hereinafter, the image is changed into binary image and split into 4 subsections with the same size. Then, the white pixel number of occurrences will be counted in each of subsection. The subsection with the highest number of white pixels occurrences will be selected and split into 4 smaller subsections. This process is repeated until each of the smaller subsection has 1 pixel and 1 pixel with the biggest number is found. (Figure 5) After pixel seed point is found, intensity value of the seed point is subtracted by intensity value of neighbors pixel to be compared with a threshold value. In this research 0.05 is used as threshold value. If the difference of the intensity of pixel seed point and its neighbors still in threshold, the value is chosen to enter a region and used as the next seed point. The process is repetitive until there are no pixel can be chosen. (Figure 6) The experience about e-commerce website which are visited by



Fig. 5: Flowchart of Seed Point Selection Algorithm

the customers can be seen in Figure 6.

#### E. Feature Extraction

The featured extraction form image segmentation is the FSM value and First and Second Order Statistic feature.

The FSM value is used to distinguish between normal brain and tumor brain image. The value of normal brain usually bigger than 0.1 [4]. In this point, the threshold value is 0.15 which is gained after several trials. If the FSM value of MRI brain is above 0.15, so the system will inform that the brain is having a normal condition. In such condition, the system will not move to the next feature extraction as well as into the Artificial Neural Networks. Meanwhile, if the value is under 0.15, means the brain is having tumor. Thus, the system will extract the first and second order statistic.

The first and second order statistic feature is used to distinguish between the malignant tumor and benign tumor. Those are the parameters that will be taken; mean, variance, skewness, kurtosis, entropy, angular second moment, contrast, correlation, variance<sub>2</sub>, inverse different moment, entropy<sub>2</sub>, and maximum probability. Those values will then be the input vector on Artificial Neural Networks Backpropagation.



Fig. 6: Flowchart of Region Growing Algorithm

# F. Artificial Neural Networks Backpropagation

After attaining the information from MRI tumor brain, the next process is the classification process which uses ANN-BP. The input towards the ANN is the information of feature extraction; the first and second order statistic. Then, the ANN will generate the output in terms of classification result on MRI brain image which has been categorized as malignant or benign tumors.

There are two processes in the classification step; training process (Table I) and testing process (Table II).

# TABLE I: Training Algorithm

Step	Algorithm
1	The initialization of weight and refraction on each neuron
2	Each of input neuron receives Pi, input data and send the data to the neuron in hidden layer
3	Each of the hidden layer neuron counts all the input weights
4	If the number of hidden layer (nHidden) is more than one:
	The counting result on hidden layer neuron is accepted by the next hidden layer neuron and counted for the
	whole input weights
5	The counting result of hidden layer neuron is received by output layer neuron, and then the counting of the
	whole input weights is conducted
6	Output layer neuron receives the target pattern data, and then the error counting is conducted
7	The result of the sixth step is used to set the weights and refraction of hidden layer
8	The resetting process of weights and refraction on output layer neuron is conducted
9	The resetting of weights and refraction on hidden layer neuron is conducted
10	Checking the stop condition

# TABLE II: Testing Algorithm

Step	Algorithm
1	The initialization of weights and refraction based on the collected data of training.
2	Input the testing data $U$ , then count the $A1$ ,
	$A1 = f(B_j + \sum_{i=1}^{n} W1_j U_i)$ If <i>nHidden</i> is more than one
	$A1_{H} = f(B_{Hj} + \sum_{i=1}^{N} W 1_{Hj} A 1_{H-1})$
3	i=1
5	After attaining the value of $A1_H$ then could $A2$ ,
	$A2 = f(B_k + \sum W2_k A1_j)$
	j=1
4	Check the value of A2, if $A2 \ge$ one threshold value, then $A2 = 1$ , else $A2 = 0$

## G. Testing Scenario

There are three scenarios going to be tested in this research, those are:

- Testing the parameter that is used in the process of digital image setting, especially for testing the algorithm of seed point selection and threshold value of Region Growing in the segmentation process of tumor area. The values that are going to be tested are:
  - The mask window size on median filter,  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$
  - The contrast and brightness value on image sharpening with (0.6, 0.4), (0.7, 0.3), and (0.8, 0.2) values.
  - The threshold value on Region Growing with tested value of 0.01, 0.025, 0.05, 0.075, 0.1.
- 2) Testing the FSM threshold value in the case of distinguishing between MRI image of normal brain and tumor brain. The values that are going to be tested are 0.05, 0.1, and 0.15.
- 3) In this scenario, the parameter of ANN will be tested to gain the best result of brain tumor type classification. Those parameters are the number of neuron on hidden layer, learning rate and epoch.

#### H. Performance System Analysis

The system performance is measured by using Sensitivity, Specifity, Accuracy, and Balanced Error Rate (BER).

In the classification of normal and tumor brain, Sensitivity shows how good the system in identifying the brain tumor while specificity shows how good the system in identifying the condition of normal brain.

The Accuracy itself is the measurement of system effectiveness in conducting the whole classification while BER is the possibility of system in doing error on the classification. Sensitivity, Specifity, Accuracy and BER is counted as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specifity = \frac{TN}{TN + FP}$$
(3)

$$Akurasi = \frac{TP + TN}{TP + FN + TN + FP} \tag{4}$$

$$BER = \frac{1}{2}\left(\frac{FN}{TP + FN} + \frac{FP}{TN + FP}\right) \tag{5}$$

# IV. THE TEST AND ANALYSIS OF SYSTEM

#### A. The Measurement Analysis of Mask Window on Media Filter

The use of median filter is to omit noise to improve the quality of image. The size of optimal mask in this research is 3.

#### B. The Analysis of Contrast and Brightness Values on Sharpening the Image Contrast

Sharpening the image contrast is aimed at making a clearer image on the tumor area. The contrast and brightness value is (0.8, 0.2). It is selected because it creates an image with the highest level of acuity without omitting some parts of image. The use of Seed Selection algorithm for seed point process will be analyzed along with the threshold region growing value to





Fig. 7: The image before being segmented (a) After being segmented using T = 0.05 (b)

27

determine the algorithm ability on resulting the right and exact segmentation. Based on the analysis of all brain MRI images, 0.05 threshold value is attained which is becoming the best segmentation in this project.

Besides, the under-segmentation happens for 0.01 and 0.025 threshold value. While over-segmentation happens for 0.075 and 0.1 threshold value. In addition, by using the algorithm of seed selection, the tumor areas can be segmented correctly without users participation.

# C. The Analysis of FSM Threshold Value

The determination of FSM Threshold value refers to paper [4]. III shows the result of FSM Threshold value testing.

FSM Threshold	Accuracy (%)	BER
0.05	74.07	0.064
0.1	96.29	0.009
0.15	100.00	0

TABLE III: The testing result with FSM value

Based on the above testing result, FSM Threshold value with the highest accuracy and low BER is 0.15. It shows that there is clear difference between MRI image of normal brain and MRI image of a brain that has brain tumor so that system can distinguish the two classes well.

## D. The Analysis of Neuron Number on Hidden Layer

TABLE IV: The Average of testing result on the number of hidden layer neuron

Number of Hidden Layer	Accuracy (%)	Training Time (s)	Attained MSE
10	82.78	47.50	0.1180
20	85.00	80.99	0.1061
50	89.45	184.09	0.0737
100	91.11	355.50	0.0668

It can be inferred based on the above table, that the number of hidden layer neuron influences the result of accuracy. When the number of neuron is 10, there happens an improvement of accuracy from 8.33% into 82.78% and 91.11% when the neuron is 100.

Besides, the number of neuron also influences the training time of ANN. From the table above, the increasing number of neuron is proportional with the increasing number of training time as needed. This research selects 100 neurons for its ability on creating the highest average of accuracy compared to others.

## E. The Analysis in the changing of Learning Rate Value

TABLE V	V: the average of	f Learning Rate Tes	ting result
arning Rate	Accuracy (%)	Training Time (s)	Attained MSE
0.0001	02.11	194 50	0.1129

Learning Rate	Accuracy (%)	Training Time (s)	Attained MSE
0.0001	83.11	184.50	0.1138
0.001	89.72	184.27	0.0765
0.01	91.39	185.27	0.0622
0.1	91.39	181.94	0.0604
0.25	91.94	184.77	0.0594

From the testing result, it can be seen that 0.0001 Learning Rate can provide a big MSE compared to others. It also can be seen that the value of MSE Learning Rate 0.0001 is 0.1138 which is bigger than Learning Rate 83.11%. Thus, the smaller MSE, the bigger ANN will be.

The difference of time, as resulted, is not significant. As a matter of fact, in 0.25 learning rate, 2.83 seconds longer is needed opposed to 0.1 learning rate. Other factors also influence this; such as the hardware ability.

From the testing result, 0.25 learning rate is selected for its ability on resulting the highest average of accuracy; 91.94% with 184.77 seconds on training time.

#### F. The Analysis in the Changing of Epoch Number

The increasing number of epoch seems influence the MSE value, accuracy level, and training time. It can be seen that the MSE becomes smaller when the epoch increases. The smaller number of MSE also influences the average of accuracy. When epoch is 1000, the accuracy 86.94% with the best attained MSE is only 0.0947. While the epoch is added to be 7000, the best attained MSE is 0.0674 and the accuracy increases become 91.39%. That is why, 7000 epoch is selected in this final result for its ability in attaining the smallest MSE value.

Epoch	Accuracy (%)	Training Time (s)	Attained MSE
1000	86.94	61.86	0.0947
3000	89.44	185.22	0.0766
5000	90.56	309.04	0.0704
7000	91.39	432.25	0.0674

TABLE VI: The average of epoch testing result

# G. The Performace of Testing on the Best Training Result

Based on the above testing and analysis result, the parameter written in VII. is used in this project which later on will be used for testing.

FABLE	VII:	TThe	best	training	result
			~~~~		

The Number of Hidden Layer Neuron	Learning Rate	Epoch	Training Data	Testing Data
100	0.25	7000	36	18

TABLE VIII: the average of testing result on all brain images

Input	Accuracy (%)	BER
Training Data	89.72	0.10
Testing Data	84.44	0.16

#### V. CONCLUSION AND SUGGESTION

#### A. Conclusion

From the result of testing on all tumor MRI images, the algorithm of seed point selection, which is modified from converging square algorithm, can select the seed point automatically. Besides, the number of threshold of Region Growing also influences the segmentation result. The optimal Region Growing threshold in this project is 0.05.

In the classification process of normal and tumor brain images, 0.15 can be attained as FSM threshold value so that it the distinguish between the two classes has accuracy value up to 100%. This is possible because there are clear differences between two classes images.

For the tumor type classification, ANN can achieved 89.72% for the average of accuracy, and BER 0.1 for training data. While the average of accuracy for testing data is 84.44%, BER 0.16. The parameter values use in this project are; 100 for the number of hidden layer neuron, 0.25 for leaning rate, 7000 for the number of epoch.

## B. Suggestion

- 1) The image input for further research should use other medical images.
- 2) Adding the function for preprocessing image automatically
- 3) Using other methods for the process of selecting the seed point in Region Growing Algorithm
- 4) Increasing other constraints as the threshold of Region Growing algorithm
- 5) Using other feature extraction methods
- 6) Using other ANN methods for brain tumor classification.

#### REFERENCES

- [1] Ahirwar, A., "Study of Techniques used for Medical Image Segmentation and Computation of Statistical Test for Region Classification of Brain MRI" in I.J. Information Technology and Computer Science, pp. 44-53, 2013.
- [2] Dzung L. Pahm, C. X., "A Survey Current Methods In Medical Image Segmentation" in *Annual Review of Biomedical Engineering*, pp. 1-27, 1998.
  [3] Jayashri Joshi, M., "Feature Extraction and Texture Classification in MRI" in *Special Issue of IJCCT*, pp. 130-136, 2010.
- [4] K. Somasundaram, T. K., "Fully Automatic method to IdentifyAbnormal MRI Head Scans using Fuzzy Segmentation and Fuzzy Symmetric Measure" in ICGST-GVIP Journal vol. 2, pp. 1-9, 2010.
- [5] Lawrence O'Gorman, A. C., The Converging Squares Algorithm: An Efficient Multidimensional Peak Picking Method, Pittsburgh: Departement of Electrical Engineering and The Robotics Institute.
- Munir, R., Modul 6-Analisis Tekstur dan Ekstraksi Ciri, Bandung.
- News-Medical.Net. (2013, Apr.). Available: http://www.news-medical.net/health/What-is-a-Brain-Tumor-(Indonesian).aspx [7]
- [8] Rafael C. Gonzalez, R. E., Digital Image Processing Second Edition, New Jersey: Prentice Hall.
- [9] Suyanto, Evolutionary Computation Komputasi Berbasis Evolusi dan Genetika, Bandung, Indonesia, 2006.