Journal of
Engineering and
Applied Sciences

MEDWELL PUBLICATIONS
SINCE 2000

Classification of Brain Tumor MRI Image using Random Forest Algorithm and Multilayers Perceptron

Indah Soesanti, Meidar Hadi Avizenna and Igi Ardiyanto
Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia

Key words: MRI, feature extraction, statistical texture, GLCM, classification, random forest

Corresponding Author:
Indah Soesanti
Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia

Page No.: 3385-3390
Volume: 15, Issue 19, 2020
ISSN: 1816-949x
Journal of Engineering and Applied Sciences

Abstract: Magnetic Resonance Imaging (MRI) is a medical technique commonly used by radiologists to visualize organ structures in humans without surgery. Based on histopathological appearance, the World Health Organization (WHO) classifies premier tumors into Low Grade Glioma (LGG) and High Grade Glioma (HGG). The process of selecting a tumor area is usually done manually by a radiologist, the process takes a lot of time and effort. To help provide a second opinion for radiologists in the classification of LGG and HGG brain tumors, a computerized system is needed to process ROI, feature extraction and MRI image classification. This study aims to compare the classification results with the ROI process and without the ROI process. 1000 images in the form of 500 LGG Flair MRI images and 500 MRI images of Flair HGG were processed by determining the ROI of tumor images compared to without the ROI processing being performed. The feature extraction process uses statistical texture histogram equalization method by calculating variance, skewness, kurtosis and GLCM texture using Energy, Contrast, Entropy, Homogeneity, Correlation, SumAverage, Variance, Dissimilarity, Auto Correlation. Finally, the Random Forest model is used to classify LGG and HGG class images and be evaluated by k-fold validation validation with k = 7. The results obtained from the proposed method of accuracy, sensitivity and specificity reached 83.6% accuracy, 80.88% sensitivity and 86.84% specificity. Shows that the method used to classify with ROI results in an increase with an accuracy of 4%, sensitivity increases by 4.46% and a specificity of 3.33%. So that, the results obtained accuracy of 87.6% accuracy, 85.34% sensitivity and 90.17% specificity.

INTRODUCTION

For more than twenty years, research on brain tumors conducted by the National Cancer Institute Statistics (NCIS) says brain tumor disease increases by 10% each year[1]. And according to data from the International Agency for Research on Cancer (IARC), that more than 126,000 people in the world have brain tumors each year with >97,000 of them dying[2]. This figure shows that patients with brain tumors are very high and increase each year. To fight the disease, many researchers use several approaches through knowledge of medicine, mathematics, to computer science to obtain effective healing methods[3].

Medical imaging plays an important role as the main character in diagnosing brain tumors[4]. Magnetic Resonance Imaging (MRI) is one of the best technologies currently used to diagnose brain tumors. There are various types of brain tumors that are classified according to their original cells. Premiere tumors are tumors that start from glioma cells and are classified based on histopathological appearance using the World Health Organization (WHO) system to Low Grade Glioma (LGG) and High Grade Glioma (HGG).

The manual assessment of the desired area (ROI) to determine the location of the tumor is carried out by the radiologist as the diagnosis decision provider for the patient. Manual segmentation around the tumor margins based on slice per slice takes time and takes 12 min or more per tumor[5], so it is necessary to do a study of the comparison between the classification results with ROI and without ROI to classify the MRI HGG image (High Grade Glioma) and MRI images of LGG (Low Grade Glioma) as supporting doctors to diagnose the next stage.

Literature review: Previous studies related to the preprocessing process with the median filter to suppress speckle noise in brain MRI images[6] followed by segmentation with fuzy k-means and watershed algorithms but no classification process was carried out. This study contributed in the form of an increase in segmentation with fusion k-means by conducting a median filter treatment. The same study related methods for detecting brain tumors by doing preprossessing by increasing the important features needed for the next process. RGB MR brain images are converted to grayish images and median filters are used to remove noise from MR brain images[7].

The research related to the texture histogram feature extraction method on brain MRI images[8] continued with classification using a multinomial logistic classifier regression with ridge estimator. Research related to the Gray Level Co-occurrence Matrices (GLCM) texture feature extraction method in brain MRI images[9] was continued by classification using the two layer feed forward neural network method.

Research related to the Random forest classification method on brain MRI images[10] the study grouped into 5 classes: background, necrosis, edema, enhancing tumors and non-enhancing tumors. Research related to SVM classification method[11]. The proposed algorithm is a combination utilizing FCM grouping and SVM classifier for classification of tumors together with BCFCM for field correction of bias and HAAR wavelet transforms for feature extraction. MLP-related research[7], Texture-based features were extracted using the Gray Level Co-occurence Matrix (GLCM). The proposed texture features include energy, contrast, correlation, homogeneity. For classification purposes, the Multi-Layer Perceptron and Naïve Bayes learning algorithms are used. The difference between this research and previous research is that this study compares the classification of brain FLAIR MRI images with ROI and without ROI into the HGG and LGG classes using the Random Forest classification method and compared with the SVM and MLP algorithms and applies the histogram and GLCM texture feature extraction algorithm. The strengths of the research carried out were the 9×9 dilated morphology process in the ROI process to fatten the ROI area.

MATERIALS AND METHODS

This research was made in three stages. The first stage is the pre-processing stage. The second stage is feature extraction which is a calculation process that produces a number of values that produce a number of values taken as characteristics of MRI images. The third stage is classification to classify LGG and HGG classes. The testing process uses cross validation by grouping data into groups according to the number of folds given. After trying various folds to 10. The classification process uses random forest to classify LGG and HGG classes. Overall the research process can be seen in Fig. 1.

In this study, a CAD method was built to classify the results of MRI image data into LGG and HGG classes. This method consists of several processes, namely the process of pre-processing, processing and output which is the result of previous processes. This study aims to establish a classification scheme to classify LGG and HGG tumors, so that, there are several stages in this study

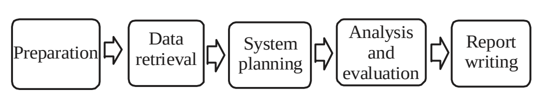

Fig. 1: System design

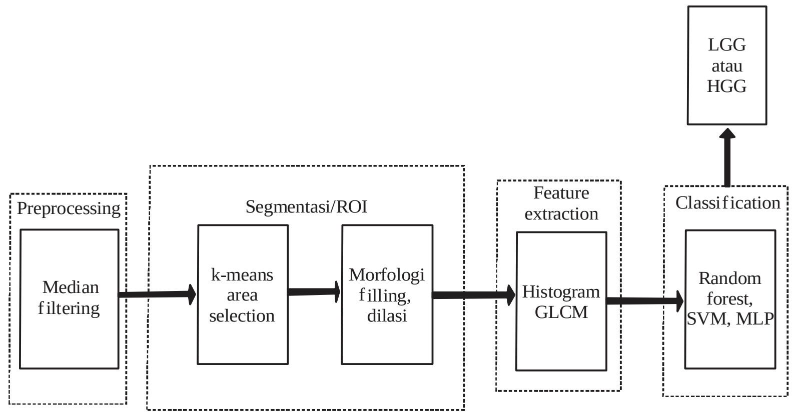

Fig. 2: Proposed method

as shown in Fig. 2. Pre-processing aims to improve image quality, usually used to improve the quality of MRI images with low contrast in order to facilitate the subsequent processing, data from the BRATS data set with an extension of * .mha and slicing and filtering with a median filter measuring 3×3.

k-means algorithm exhibit low-computational cost with higher efficiency when large datasets are clustered[12]. The k-means morphology uses k = 4 as a tumor in area 4 as an ROI area or the results of its segmentation are then eroded to take the edge area from the tumor. The MRI image used in the study will be extracted based on texture. Texture is a mutual relationship between the value of the intensity of neighboring pixels that are repeated in an area wider than the distance of the relationship. The selection of texture-based feature extraction because the image used is an MRI image of the brain.

The final stage of this research is the classification aims to determine whether the image is an LGG or HGG class. Where in a series of previous stages various processes have been carried out, so that, they get a number of features that present each patient. The value of the results will be used to be trained with the random forest algorithm.

RESULTS AND DISCUSSION

Based on the dataset used, MRI image labels are categorized into LGG and HGG. The selection of the image used does not have the patient's eye and there is a tumor to do the next phase. So that, from LGG data, 500 MRI images were used and for HGG 500 MRI images were used. Pre-processing is carried out from the BRATS data set with an extension of * .mha and slicing. In Fig. 3 and 4 the original image and the segmented image are stacked, so that, ROI can be obtained for the feature extraction process. Seen by the line that the

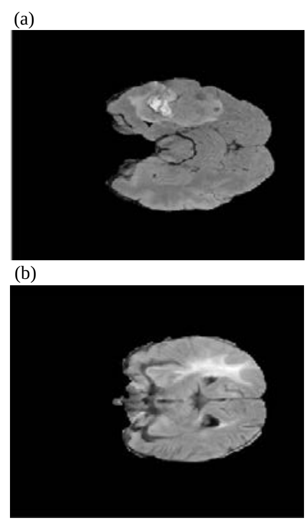

Fig. 3(a, b): LGG and HGG Image

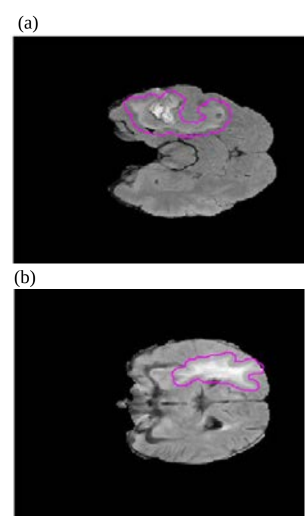

Fig. 4(a, b): LGG and HGG segmented image

Table 1: The evaluation results of the LGG image ROI results with ground truth

	Image LGG	
Images	Over segmented	Under segmented
1	0.157171	38.74263
2	0.103878	37.3615
3	0.416533	35.3092

Table 2: Results of ROI evaluation of HGG image with ground truth

	Image HGG	
Images	Over segmented	Under segmented
1	14.66165414	40.78947368
2	11.74438687	42.31433506
3	4.441260745	40.83094556

Table 3: Result of feature extraction from 4 LGG images with ROI

	Image LGG			
Features	1	2	3	4
Variance	368682380.7	426145785	418668553.1	410203797.1
Skewness	0.002238039	0.076526198	0.055702795	0.041431304
Kurtosis	-1.198474732	-1.265879351	-1.255456673	-1.244091606

Table 4: Result of feature extraction from 4 LGG images without ROI

	Citra LGG (image)			
Features	1	2	3	4
Variance	369342530.7	369651942.7	369833612.2	369455045.3
Skewness	0.002667925	0.000295729	-0.001836186	0.003048928
Kurtosis	-1.20147329	-1.19803926	-1.199904014	-1.200807847

ROI result of the image is suspected to be a tumor. To provide an evaluation of ROI results with ground truth, an evaluation of over segmentation and under segmentation was carried out. Table 1 shows the evaluation results of the LGG Image ROI results with Ground Truth. The results in Table 2-4 above show that images with ROI are validated with Ground Truth that oversegmentation values are lower than undersegmentation.

The feature extraction process is based on the output from the pre-processing which is input to feature extraction. The feature extraction method used in this study is feature extraction based on texture histogram and texture (GLCM). The histogram-based texture extraction used 3 features, namely: variance, skewness and kurtosis. While the GLCM features used are Energy, Contrast, Entropy, Homogensity, Correlation, Sum Average, Variance, Dissimilarity, Auto Correlation. The entire data is used for training and testing with the 7-fold cross validation validation method. The results of 4 images can be seen in Table 5 and 6.

The feature extraction results are carried out by extracting all areas of the brain and extracting from ROI results that were carried out previously, so that, it will be compared to how significant the results obtained with and without the process of selecting the ROI. Classification is the final stage in this study, the classification scenario in this study was done by classifying two classes of brain images, namely the class image of LGG tumors and the image of HGG tumors, then the values of accuracy, sensitivity and specificity were sought. The whole brain image classification is carried out first to determine accuracy, sensitivity and specificity and to see a comparison between the Random Forest classifier and SVM and MLP.

Table 7 is the result of the classification evaluation of the three classifiers namely Random Forest, SVM and MLP on images that were not carried out by the ROI process in the experiment k = 7. These results indicate that the Random Forest classifier gets the highest value from the SVM classifier and the MLP classifier with the advantage of the difference in accuracy >21.1% of the SVM classifier and 6.7%> the MLP classifier. Based on the sensitivity produced, the Random Forest classifier is 19.24% greater than the SVM classifier and 7.49% greater than the MLP classifier, as well as the results of greater specificity 23.34% of the SVM classifier and 5.2% greater than MLP classifier.

Table 8 is the result of the evaluation of the classification of the three classifiers, Random Forest, SVM and MLP at the trial period k = 7. These results indicate that the Random Forest classifier gets the highest value from the SVM classifier and the MLP classifier with the difference in accuracy with a greater 29% of the SVM classifier and 10.7% greater than the MLP classifier. Based on the sensitivity generated, the Random Forest classifier is greater than 21.47% of the SVM classifier and 6.91% greater than the MLP classifier, as well as greater specificity results of 33.94% of the SVM classifier and 14.65% greater than MLP classifier.

From Table 8 shows that the classification results with Random Forest, SVM and MLP, it can be seen that Random Forest has a better performance than SVM and MLP in the two class classification, namely LGG and HGG. So that, this research will focus on the Random Forest classifier. Then if the results of the difference between the Random Forest ROI and not the ROI are observed, the difference in accuracy is 4%, the sensitivity is 4.46% and the specificity is 3.33%. The results can be seen in Table 8. The difference is the result of an increase if the method to do the ROI is used with the Random Forest classifier.

In this study the classification of MG HGG and LGG images was carried out using the random forest method. The image used is 1000 images. The feature extraction method used in this study is feature extraction based on texture histogram and texture (GLCM). The histogram-based texture extraction used 3 features, namely: variance, skewness and kurtosis. The feature extraction used is GLCM and Histogram. While the GLCM features used are Energy, Contrast, Entropy,

Table 5: Result of feature extraction from 4 HGG images with ROI

	Image HGG			
Features GLCM	1	2	3	4
Energy	0,001097695	0,002664027	0,002663829	0,002683572
Contrast	59,84638382	26,54180496	29,8774038	34,74424223
Entropy	10,27102087	9,577775932	9,672128932	9,686710674
Homogeneity	0,303056505	0,418854195	0,400690858	0,401048533
Correlation	0,913442005	0,961824417	0,959036645	0,952302782
Sum average	0,00830574	0,007535845	0,007709353	0,007716072
Variance	0,078188176	0,077746723	0,081136799	0,080896797
Dissimilarity	5,376930049	3,165585491	3,380985803	3,598333245
Auto correlation	1447,717212	1257,940412	1314,536849	1312,859262

Table 6: Result of feature extraction from 4 HGG images without ROI

	Image HGG			
Features GLCM	1	2	3	4
Energy	0,000886776	0,001035051	0,000986139	0,001014829
Contrast	64,5602459	52,16119949	52,7893875	52,84675784
Entropy	10,57249272	10,39671524	10,41617734	10,37476568
Homogeneity	0,271292363	0,309086345	0,297308357	0,298661771
Correlation	0,90376643	0,923202727	0,922720166	0,921686169
Sum average	0,008303101	0,008276989	0,008310387	0,008325226
Variance	0,077363485	0,077570128	0,077856855	0,077633923
Dissimilarity	5,888584033	5,089937499	5,233665905	5,230743466
Auto correlation	1441,247022	1441,029428	1451,184066	1454,383794

Table 7: Random forest classification, SVM and MLP image evaluation results without ROI at K = 7

	Evaluation parameters						
Schema	TP	TN	FP	FN	Accuracy (%)	Sensitivity (%)	Specificity (%)
Random forest	440	396	60	104	83.6	80.88	86.84
SVM	331	294	169	206	62.5	61.64	63.50
MLP	422	347	78	153	76.9	73.39	81.64

Table 8: Random forest classification, SVM and MLP image evaluation results with ROI at K = 7

	Evaluation parameters						
Schema	TP	TN	FP	FN	Accuracy (%)	Sensitivity (%)	Specificity (%)
Random forest	454	422	46	78	87.6	85.34	90.17
SVM	198	388	302	112	58.6	63.87	56.23
MLP	371	398	129	102	76.9	78.43	75.52

Homogensity, Correlation, Sum Average, Variance, Dissimilarity, Auto Correlation. The last stage is the classification stage using the random forest method. By comparing the image with ROI and without ROI, proceed by comparing with other classifiers, SVM and MLP. The best ROI classification results are random forest compared to SVM and MLP. Comparison of k = 7 fold on the results of evaluating random forest classifier on ROI-generated images and no ROI. The accuracy value increased by 4% while the sensitivity value increased by 4.46% as well as the value of specificity increased by 3.33%. The performance of each of the three models built using RF, SVM and MLP was tested before and after ROI. Based on the tests carried out, it is known from 1000 images, showing the accuracy of the highest Random Forest algorithm compared to Support Vector Machine (SVM) and MLP. The results obtained are so because Random Forest is a versatile classification algorithm that is suitable for analysis of large data sets. The Random Forest is popular because the RF classification model has high predictive accuracy and provides information about the importance of variables for classification.

CONCLUSION

From the results of research conducted to classify brain MRI images with texture-based features and GLCM conclusions can be drawn as follows: The random forest method gets the best results from SVM and MLP on the classification of LGG and HGG brain MRI images with an accuracy of 83.6%, sensitivity 80.88%, specificity of 86.86%. Random forest method with ROI obtained the best results from SVM and MLP on MRI image classification of LGG and HGG brains with increased accuracy to 87.6% accuracy, 85.34% sensitivity, 90.17% specificity.

ACKNOWLEDGEMENTS

The researchers gratefully acknowledges the contribution from the Ministry of Research and Technology, Republic of Indonesia, for funding this research.

REFERENCES

01. Natarajan, P., N. Krishnan, N.S. Kenkre, S. Nancy and B.P. Singh, 2012. Tumor detection using threshold operation in MRI brain images. Proceedings of the 2012 IEEE International Conference on Computational Intelligence and Computing Research, December 18-20, 2012, IEEE, Coimbatore, India, pp: 1-4.
02. Ferlay, J., H.R. Shin, F. Bray, D. Forman, C. Mathers and D.M. Parkin, 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer, 127: 2893-2917.
03. Al-Tamimi, M.S.H. and G. Sulong, 2014. Tumor brain detection through MR images: A review of literature. J. Theor. Applied Inf. Technol., 62: 387-403.
04. Chaplot, S., L.M. Patnaik and N.R. Jagannathan, 2006. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. J. Biomed. Signal Process. Control, 2: 86-92.
05. Aslian, H., M. Sadeghi, S.R. Mahdavi, F.B. Mofrad, M. Astarakee, N. Khaledi and P. Fadavi, 2013. Magnetic resonance imaging-based target volume delineation in radiation therapy treatment planning for brain tumors using localized region-based active contour. Int. J. Radiation Oncol. Biol. Phys., 87: 195-201.
06. Benson, C.C., V. Deepa, V.L. Lajish and K. Rajamani, 2016. Brain tumor segmentation from MR brain images using improved fuzzy c-means clustering and watershed algorithm. Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), September 21-24, 2016, IEEE, Jaipur, India, pp: 187-192.
07. Sharma, K., A. Kaur and S. Gujral, 2014. Brain tumor detection based on machine learning algorithms. Int. J. Comput. Appl., 103: 7-11.
0.8 Avizenna, M.H., I. Soesanti and I. Ardiyanto, 2018. Classification of brain magnetic resonance images based on statistical texture. Proceedings of the 2018 1st International Conference on Bioinformatics, Biotechnology and Biomedical Engineering-Bioinformatics and Biomedical Engineering Vol. 1, October 19-20, 2018, IEEE, Yogyakarta, Indonesia, pp: 1-5.
09. Zulpe, N. and V. Pawar, 2012. GLCM textural features for brain tumor classification. Int. J. Comput. Sci., 9: 354-359.
10. Usman, K. and K. Rajpoot, 2017. Brain tumor classification from multi-modality MRI using wavelets and machine learning. Pattern Anal. Appl., 20: 871-881.
11. Batra, A. and D.G. Kaushik, 2017. SECTUBIM: Automatic segmentation and classification of tumeric brain MRI images using FHS (FCM HWT and SVM). Int. J. Eng. Sci. Comput., 7: 13190-13194.
12. Singh, N., D.P. Singh and B. Pant, 2019. ACOCA: Ant colony optimization based clustering algorithm for big data preprocessing. Int. J. Math. Eng. Manage. Sci., 4: 1239-1250.