

Detecting Brain Tumor by Using Machine Learning and Image Processing Techniques

Ahmet Safa KARAKOÇ
Electrical and Electronical
Engineering Department
Abdullah Gul University
Kayseri/Turkey
ahmetsafa.karakoc@agu.edu.tr

Ahmet Yasir CILVEZ
Electrical and Electronical
Engineering Department
Abdullah Gul University
Kayseri/Turkey
ahmetyasir.cilvez@agu.edu.tr

Abstract— In this article, the classification of two types of brain tumors explained. In the method of the research, image filtering, segmentation, enhancement is used in the image processing part, and SVM (support vector machine) is used in the machine learning part. As a result of this research, test images are classified into two classes with a minimum 62% accuracy rate.

Keywords— Brain tumor, GLCM, SVM, image processing, machine learning

I. INTRODUCTION

Brain tumors are one of the most dangerous illnesses that can be seen in all people of every age. The reason brain tumors are highly dangerous is that they are located in the nervous system. Each year, approximately 12000 people are diagnosed with this disease. According to Davies [1], between the 20-44 age, while the rate of survival is 50%, after 65 years old the survival rate decreases to 5%. Because of the importance of brain tumors, diagnosing earlier will help to intervene rapidly. In addition, by using machine learning and image processing methods, the missed diagnoses cases can be detected truly. Brain tumor detection is possible by transferring MRI images to digital software. Transferred images can be classified according to tumor size and location.

In this project; we will detect the brain images in our selected dataset using image processing and classify the diagnosed tumors with machine learning techniques by using MATLAB. In the literature review part, there are studies about brain tumor detection and classification methods that were useful for this project.

II. LITERATURE REVIEW

K-means clustering is one of the pixel-based segmentation methods that is introduced in [2]. This method is applied on two different data sets which were taken from the gray feature vectors of the original MR brain images and the converted RGB features to detect the tumor objects in magnetic resonance (MR) images. Color translation, K-means clustering, and histogram clustering are the applications that combined in this method. Because of not causing extra overhead of the color space translation, this method is easy to applying and is efficient.

Gray level co-occurrence matrix (GLCM) was used to extract the second-order statistical texture features for motion

estimation of images which are introduced in [3]. In that project, angular second moment, correlation, Inverse difference moment, and entropy features are extracted by using Vivado FPGA. These features have high classification accuracy and require less extraction time that means they have some advantages. To conclude; GLCM is one of the most efficient feature extraction methods.

IBkLG classifier has been used for the classification of normal, benign, and malignant tumors in brain MRI images by developing an automated method [4]. The IBkLG method is one of the successful methods that reach approximately 100% classification accuracy because of handling real-valued features. In this proposed method; the system has the ability to get rid of noisy images by using median filtering. For the preprocessing, the filtered image is converted gray to RGB and to l^*a^*b color space. After these steps; the image has transferred to a color-based segmentation process that makes the images ready to feature extraction. For the feature extraction, GLCM, Histogram of oriented gradients, and canny edge detection have been used. The extracted features' mean, standard deviation, skewness, Root mean square value, and median absolute values are calculated. From these values, a dataset table has been created to use in IBkLG classifier. IBkLG classifier has reached 86.6% classification accuracy.

DWT+PCA+SVM is a developed model to detect and classify the normal and abnormal MR images of the brain. [5]. DWT stands for discrete wavelet transform that extracts the information of the brain images without any loss. This method has advantages that can able to keep both frequency and location information. PCA stands for principal component analysis. PCA is used the reduce the dimensions of the images which will cause the worsened classification accuracy and heavy computation burden. For the SVM section, kernel SVMs types have been tried for classification. Among them, GRB SVM has reached the most accurate method for classification. At the end, the DWT+PCA+SVM method's accuracy level is 99.38%.

Skull stripping is one of the preprocessing methods for MRI brain images [6]. This method is the process of brain tissue segmentation from the surrounding region that is the skull and nonbrain area. Morphology-based methods have been used to separate the brain area from the skull. This

method has 5 steps. These are image thresholding, morphological operators that fill the gaps in the extracted objects, edge detection, creating a binary mask, and combining the binary mask. In conclusion, this algorithm has provided more efficient MRI images that are not affected by the surrounding region of the brain.

III. DATASET DEFINITION

The dataset was taken from Kaggle [7]. In this dataset; there are 3 different tumor types that are glioma, meningioma, and pituitary and normal brain images. The description of the tumor types will be introduced in order.

Gliomas are a general name used to describe a group of tumors that occur in the glial cells, the supporting tissue of the brain. Symptoms include headache, seizures, nausea, and vomiting. (Figure1). Meningiomas are slow-growing and mostly benign tumors that arise from the membranes surrounding the brain. Although they are not technically a brain tumor since they do not originate in the brain, they cause a number of symptoms by creating pressure on the brain with their growth. (Figure2). Pituitary tumors are abnormal growths that develop in your pituitary gland. Some pituitary tumors result in too much of the hormones that regulate important functions of your body. Some pituitary tumors can cause your pituitary gland to produce lower levels of hormones. (Figure3).



Figure 1

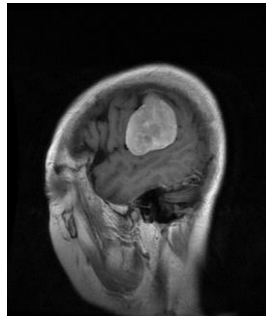


Figure 2

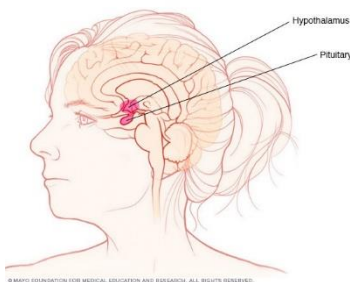


Figure 3

Dataset has been separated into 2 files that are training images and test images. In the training image file, there are 841 different images with different types. In the testing image file, there are 156 different images. For our project, we did not use all images. Because the shooting angles of the images are different and to use these images together will decrease the accuracy. So, the MRI images that are taken by the same shooting angle were trying to selected.

IV. METHODOLOGY

A. Image Processing (Preprocessing)

Before starting classification, some methods have to be applied in Image Processing to make the image workable. Firstly, the skull stripping method (Figure4) has been implemented to remove the surrounding area of the brain. This method helps to focus on the brain tumor part. Secondly, median filtering (Figure5) has been applied to remove the noises from the brain MR images via a built-in function in MATLAB. After the filtering is completed, normalization has been applied. Normalization has been used to change the range of pixel intensity values. The next methods after these steps are image enhancement and image segmentation. Image enhancement is used to adjust the digital images to make them more suitable for display. For this implementation “imadjust” function has been used in MATLAB. Image segmentation’s goal is to change the representation of the image to make it easier to analyze and more meaningful. As a final step morphology operator has been applied to brain images. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. (Figure 6) [11]. After all, steps completed, MRI images are ready for GLCM that will extract the features of the images.

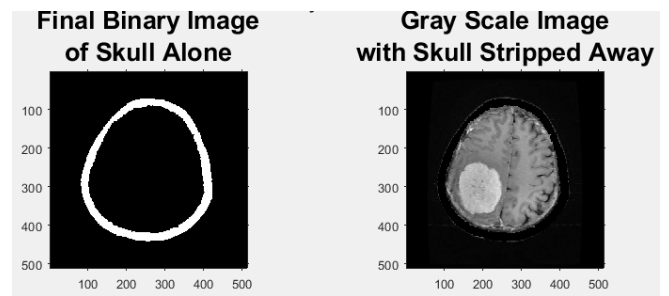


Figure 4 Skull Stripping

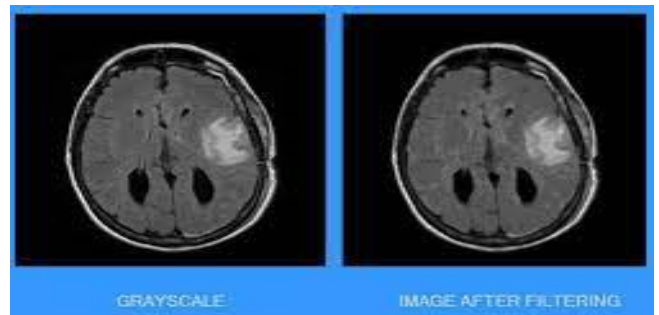


Figure 5 Median Filtering

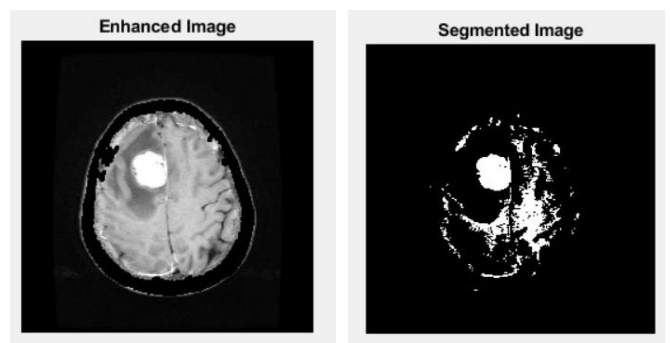


Figure 6

Figure 7

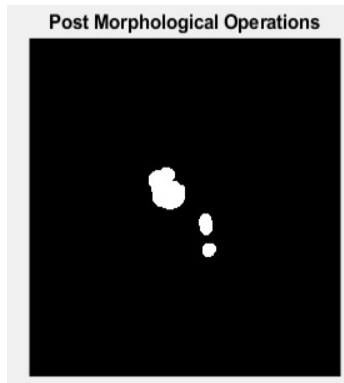


Figure 8

B. Feature Extraction (GLCM Texture)

When the image processing part is completed, to extract the features from the images GLCM Texture has been used. The gray level co-occurrence matrix is a statistical method that examines the texture by considering the spatial relationship of pixels.[6] The GLCMs are stored in a three-dimensional matrix that $(i \times j \times n)$. n represents the number of GLCM's calculated where n is the number of GLCMs calculated.

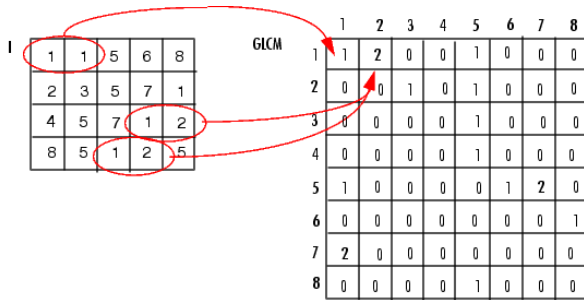


Figure 9 GLCM Algorithm

By implementing GLCM, some characteristic features are extracted. Contrast, Correlation, Energy, Homogeneity are the most important features among them. Contrast calculates the local variations of the GLCM matrix. Correlation calculates the joint probability of occurrence of the pairs. Energy shows the sum of squared elements in GLCM. Homogeneity calculates the closeness of the distribution of elements in GLCM [8],[9].

In this project, 22 different features were extracted of each brain image in the dataset. An Excel file was created to keep these values. These values will be used for classification. These features are listed below;

Autocorrelation, Contrast, Correlation, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, Homogeneity, Homogeneity, Maximum probability, Sum of squares, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation1, Information measure of correlation2, Inverse difference (INV), Inverse difference normalized (INN), Inverse difference moment normalized.

Table 1 GLCM Features Values

		GLCM Features											
Sample	tumour	autoc	contr	corm	corr	cprom	cshad	dissi	energ	entro	homom	homop	
1	0	1.0057485	0.0002216	0.9431374	0.9431374	0.0294651	0.0148711	0.0002216	0.9958797	0.0156812	0.9998891	0.9998891	
2	0	1.2535469	0.0044260	0.9716224	0.9716224	0.9331436	0.5064914	0.0044260	0.8396231	0.3172226	0.9977869	0.9977869	
3	0	1.0722465	0.0017964	0.9622397	0.9622397	0.3418847	0.1758909	0.0017964	0.9506325	0.1256565	0.9991017	0.9991017	
4	0	1.1741644	0.0045636	0.9587791	0.9587791	0.7127295	0.3786761	0.0045636	0.8847448	0.2493678	0.9977181	0.9977181	
5	0	1.0040400	0.0001758	0.9360230	0.9360230	0.0206697	0.0104364	0.0001758	0.9970760	0.0116764	0.9999120	0.9999120	
6	0	1.0162824	0.0005045	0.9539765	0.9539765	0.0827915	0.0418691	0.0005045	0.9885333	0.0375267	0.9997477	0.9997477	
7	0	1.4370910	0.0058020	0.9768203	0.9768203	1.2274268	0.6952403	0.0058020	0.7439242	0.4505174	0.9970989	0.9970989	
1	1	0.9979360	0.5673828	2.0039062	3.9612254	0.0155275	0.0002216	0.0020869	-0.8906464	0.1570864	0.9999261	0.9999556	
2	1	0.9125336	0.7756333	2.1705066	3.7535571	0.3141547	0.0044260	0.0284066	-0.9114437	0.6419593	0.9985246	0.9991147	
3	1	0.9747202	0.6234539	2.0487631	3.7964817	0.1244113	0.0017964	0.0131516	-0.9038385	0.4326190	0.9994011	0.9996407	
4	1	0.9389027	0.7095385	2.1176308	3.7190973	0.2462045	0.0045636	0.0291496	-0.8852356	0.5718608	0.9984787	0.9990872	
5	1	0.9985361	0.5659399	2.0027519	3.9701876	0.0115545	0.0001758	0.0016959	-0.8816525	0.1350624	0.9999413	0.9999648	
6	1	0.9942361	0.5762788	2.0110231	3.9174887	0.0371769	0.0005045	0.0043347	-0.9014897	0.2444052	0.9998318	0.9998990	
7	1	0.8504349	0.9291600	2.2933280	3.9056022	0.4464957	0.0058020	0.0356630	-0.9193076	0.7316789	0.9980659	0.9988395	

C. Classification (SVM)

There are many options for classification in Machine learning. To select the best option for this project, the Classification Learner app was used in MATLAB. We trained our dataset via this application. The highest accuracy has reached with Kernel SVM. To train and cross-validate 'fitcsvm' function has been used. The common syntax is:

```
SVMModel=fitcsvm(X,Y,'KernelFunction','rbf',
...'Standardize',true,'ClassNames',{'negClass',
'posClass'});
```

In the SVM model Gaussian RBF (Radial Basis Function) is used. RBF stands for Radial Based Function. RBF kernel is a function whose value depends on the distance from the origin or from some point.[10]

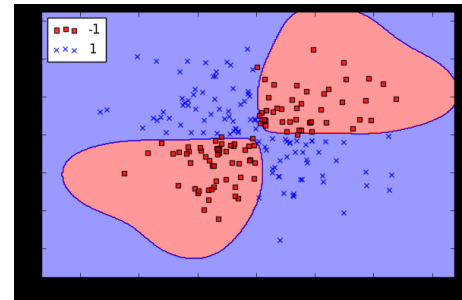


Figure 10 RBF Kernel

V. RESULTS AND DISCUSSION

The methods in the Methodology section were applied to the two different data which contain Glioma and Meningioma tumor MRI images. According to the results of the algorithm, the accuracy rate shows that the algorithm classifies the tumor type. However, the rate is less than the expected value. One of the possible reasons for this issue is the number of low-quality images in the dataset. Moreover, because of the MRI nature, all of the images have different perspectives even though some are similar to each other. Another possible reason for the low accuracy is the image processing part. Skull stripping algorithm removes some important part of the minority of the total images and if a K-means algorithm for segmentation due to the grayscale of the image is used successfully, the results can be more trustable. Even though we used first K-means algorithm to detect tumor and working on just tumor instead of rest of the brain, there are some disadvantages have occurred and it didn't work every MRI images. That's why it was eliminated from the algorithm.

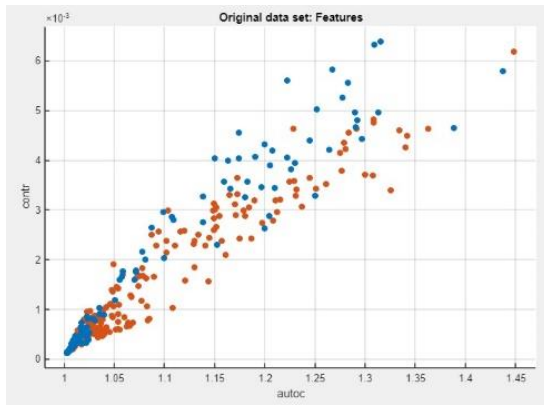


Figure 11 Spreading of Autocorrelation vs Contrast

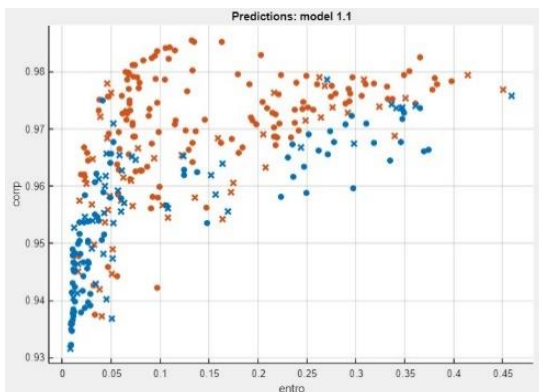


Figure 12 Spreading of Entropy vs Correlation

In order to increase the accuracy rate, MATLAB's own algorithm generated from 'Classification Learner' can be used. Even though the accuracy rate of the Support Vector Machine algorithm is higher and higher, the results of our SVM algorithm gave less accuracy rate. Further, the number of GLCM texture features can be rearranged to obtain more accurate results.

VI. CONCLUSION

In conclusion; different types of brain tumor images have been detected and classified successfully by using image processing and machine learning techniques. Firstly, image processing techniques have been applied to make the image suitable for classification. Secondly, the GLCM Texture Features method has been used to extract the characteristic features of the brain images. Finally, RBF Kernel SVM provides us to classify the brain images. To testing the machine learning algorithm, 200 training examples of each two groups were used and the algorithm was trained. The accuracy rate obtained was a minimum of 62%.

VII. REFERENCES

- [1] https://www.medicinenet.com/brain_tumor_symptoms/vi_ews.htm
- [2] Ming -ni wu1 chia- chin lin2 chin-cheng-chan3, "Brain Tumor Detection Using Color-Based K-Means Clustering Segmentation" December 2007 DOI:10.1109/IIHMSP.2007.4457697, PP 1-3
- [3] P. Mohanaiah1, P. Sathyanarayana2, L. GuruKumar3, "Image Texture Feature Extraction using GLCM Approach" International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013 1 ISSN 2250-3153, PP 1-2
- [4] Chauhan, S., More, A., Uikay, R., Malviya, P., & Moghe, A.A. (2017). Brain tumor detection and classification in MRI images using image and data mining. 2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE), 223-231.
- [5] Yu-Dong Zhang and Lenan Wu, "An Mr Brain Images Classifier via Principal Component Analysis and Kernel Support Vector Machine," Progress in Electromagnetics Research, Vol. 130, 369-388, 2012.
- [6] Swiebocka-Wiek J. (2016) Skull Stripping for MRI Images Using Morphological Operators. In: Saeed K., Homenda W.(eds) Computer Information Systems and Industrial Management. CISIM 2016. Lecture Notes in Computer Science, vol 9842. Springer, Cham. https://doi.org/10.1007/978-3-319-45378-1_16
- [7] <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri?select=Training>
- [8] Avinash Uppuluri (2021). GLCM texture features (<https://www.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features>), MATLAB Central File Exchange. Retrieved June 9, 2021.
- [9] <https://www.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html>
- [10] <https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589>
- [11] <https://www.mathworks.com/help/images/morphological-dilation-and-erosion.html>