Analysis of Medical Images using Machine Learning Techniques

S. R. Dogiwal, Mahesh Kumar Joshi
Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur-302017 (INDIA)
Email - mjoshi293@gmail.com, dogiwal@gmail.com
Received 21.01.2019 received in revised form 07.03.2019, accepted 11.03.2019

Abstract: In medical research and development, the computational domain provides support and calculative power to perform unique and innovative tasks. In computer field Image categorization is a huge and an important issue. In the world of medical science everyday doctors are encountered with various types of image related problem. Pattern classification is a solution to categorize the image. In our human body brain play very important role, although due to the sudden growth of the anomalous tissues in the brain caused a very critical disease named as brain cancer. There are two different types of brain tumor such as benign and malignant. When cells and tissues are relatively grow slowly such type of cancer named as benign tumor which is termed as non cancerous tumor while in the malignant type of tumor, cells are growing very fast and causes serious harmful to the brain which causes death. There are so many techniques for extracting the medical images related to brain but MRI (Magnetic Resonance Imaging) is very powerful technique for extracting the digital images of internal slice of the brain. This information is very helpful for medical diagnosis and research purpose. To identification the tumor in human body firstly it goes through image acquisition process and then through the proper image segmentation technique finds out the tumor. The proposed work is motivated from the medical image analysis for medical dataset like brain tumor. In this research for image segmentation Otsu's method is used and to select the appropriate features extraction principal component analysis is used. Gray level co-occurrence matrix is used to calculate some valuable texture parameters and finally using machine learning techniques kernel functions are used to calculate the highest accuracy with the help of support vector machine.

Keywords: MRI, PCA, SVM, IDM, RMS.

1. INTRODUCTION

Now days, information extraction from the web and large databases are common. The users frequently place their queries for obtaining the results according to the queried data. In recent years a great effort of the research in the field of medical imaging was focused on brain tumour segmentation. Despite the undisputed usefulness of automatic tumour segmentation, this is not yet a widespread clinical practice; therefore the automatic brain tumour segmentation is still a widely studied research topic [1]. The principle of our task is to detect the brain tumour from the MRI image of the brain. A brain tumour is an abnormal growth of tissue in the brain. Unlike other tumours, brain tumours spread by local extension and rarely metastasize (spread) outside the brain [2]. All types of brain tumours may produce symptoms that vary depending on the part of the brain involved. These include headaches, problem with vision, vomiting, and mental changes. More specific problems may include difficulty in walking, speaking and with sensation. Each year, more than 17,000 brain tumours are diagnosed in the United States [3]. Here, for feature extraction of digital images wavelet transformation is used and then applying PCA (Principal Component Analysis) to reduce the dimensions of Features. Kernel SVM is the K-fold stratified cross validations are important for improve generalization of Kernel SVM [4]. SVMs are suitable method for binary images classification. Initial stage of brain tumours start from the brain itself, while second stage brain tumours originate from rest part of body. MRI is an imaging technique that provides high resolution images of the brain of human body. Non-invasive imaging methods, e.g. magnetic resonance imaging (MRI), computed tomography (CT) and positron emission tomography (PET), that could diagnose a brain tumor, would avoid unnecessary surgery.

2. LITERATURE SURVEY

The authors Panda et al. suggested that Otsu's segmentation technique is works well for Brain MRI image to segment and PCA performed give the good results for feature extractions. In this study they analysis the four standard kernel functions for SVM with a brain MRI images. In which images consists of two different classes of tumour i.e. Malignant and Benign. In this paper suggested that segmentation technique is works well for Brain MRI images to segment and performed PCA results for feature extractions and reduction [5]. Zhang and Wu proposed technique for identifying MR brain images from abnormal MR brain images, find GRB kernel as the successful one[6].

Another author proposed an interactive segmentation method that enables quickly and efficiently segment tumours in MRI of brain. The extraction of the tumour and the size of the interested tumours successfully obtained in the MATLAB coding for image processing and segment the different part of the brain from the brain CT mages. By calculating the area, there are slight differences between the sizes of the tumour for different slice of brain images [7].

The author of the concerned paper Kharrat et al. proposed an efficient detection of brain tumor algorithm introduced. It's based on mathematical morphology, wavelet transform and K-means technique [8].

3. PROPOSED METHOD

The human body brain tumor recognition system mainly consists of five stages as follow:

- Image Acquisition: Acquisition of the brain image is the first step of any image analysis process. Brain MRI images can be captured of collect from hospitals.
- Image Pre-processing: In this step, acquired images can be covert in to required format, i.e. resize the image, remove the noise etc.
- Feature Extraction: features extraction technique is classify the image in to different sub images according to energy level of pixels.
- Classification: In this process classify the image data into different categories according to nature of image.
- Accuracy assessment: By applying different techniques or methods to find out the accuracy of the different category of image data.

Algorithm for performance evaluation:

1) Load or input image from different media.
2) Convert the input image into binary image. i.e. Otsu Binarization.
3) Apply the Fourier and discrete wavelet transform. The DWT is a efficient and effective method for finding the feature. It allows analysis the various level of super resolution MRI images.
4) Apply the PCA.
5) Find the Mean, Energy, Variance, Entropy, RMS, Homogeneity, Smoothness, Contrast, Correlation, Kurtosis, Skewness, IDM, Standard Deviation, and MSE etc. for the performance evaluation.

The above algorithm procedure can be understood using the below block diagram fig. 1.

Fourier transformation is very useful in representation and analysis of stationary signal in witch the frequency components do not change with time. However, in case of non-stationary signal, where frequency components change with time, sometime it becomes a necessary to know not only which frequency components are present but also where they present. Above transforms are not good. Images are usually non-stationary two-dimensional signal. The term wavelet means small wave. These small waves are generated from bigger one through scaling and translation.

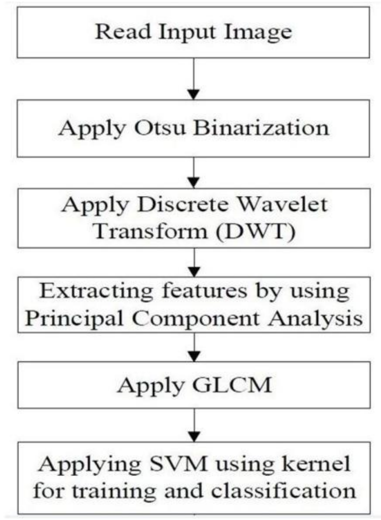

Figure 1: Proposed Approach

There are various types of wavelets. some of them are Haar wavelets, Mexican hat wavelets, Modulated Gaussian wavelets, Meyer wavelets, Doubechies wavelets[9]. First one of above is the simplest to understand and implement, while the last one is the most effective as it satisfies various requirement of wavelets such as compact support, orthogonally condition and regularity conditions. A. Lifting Scheme The lifting scheme is a process for reconstruction of images. In this scheme, wavelet transform decomposes into a set of levels. In the forward lifting scheme, each level divides the value of level into an odd half and even half for further processing[10].

Texture features can be extracted from the input image by using wavelet transform. Wavelet transform is able to analyze an image by decomposing it into a series of subband images.

These subband images can be considered as one form of wavelet coefficients, where texture features are computed. DWT can be performed by iteratively filtering a signal or image through the low-pass and high-pass filters, and subsequently downsampling the filtered data by two[10].

This process will decompose the input image into a series of subband images. Figure 2 illustrates an example of DWT, where h and g represent the low-pass and high-pass filter respectively, while the symbol with a down arrow inside a circle represents the downsampling operation.

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (or sometimes, principal modes of variation).

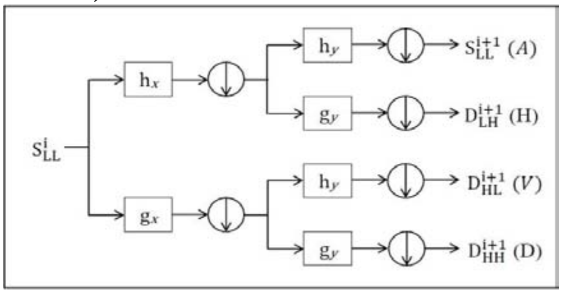

Figure 2: Discrete Wavelet Transform

Gray level co-occurrence matrix has proven to be a powerful basis for use in texture classification. Various textural parameters calculated from the gray level co-occurrence matrix help understand the details about the overall image content. Texture contains important information about the structural arrangement of surfaces. The textural features based on gray-tone spatial dependencies have a general applicability in image classification. SVMs (Support Vector Machines) are a useful technique for data classification. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one “target value" (i.e. the class labels) and several “attributes" (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

4. MEASUREMENT OF PERFORMANCE

Mean: Mean is most basic of all statistical measure. Means are often used in geometry and analysis.

$$\bar{X}= \frac{\sum_{i=0}^{n} X_i}{n} \tag{1}$$

Standard Deviation: A standard deviation filter calculates the standard deviation and assigns this value to the center pixel in the output map.

$$\boldsymbol{s} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{(n-1)}} \tag{2}$$

Entropy: The entropy measures the randomness of the distribution of the coefficients values over the intensity levels.

$$Ent = -\sum_{X_{i=1}} p(i,j)\log_2[p(i,j)] \tag{3}$$

RMS: RMS is the measure of root mean square value of an image.

$$\bar{X}_{rms} = \sqrt{\frac{\sum_{i=0}^{n} X_i^2}{n}} \tag{4}$$

Variance: It is a measure of how far a set of numbers is spread out. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean

$$\boldsymbol{s^2} = \left(\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{(n-1)}}\right)^2 \tag{5}$$

Smoothness: It is a measure of relative smoothness of intensity in a region.

$$\boldsymbol{SM = 1 - \frac{1}{1+s^2}P(i,j)} \tag{6}$$

Kurtosis: It is a measure of the shape of the probability distribution of a real-valued random variable. It is closely related to the fourth moment of a distribution. A high kurtosis distribution has longer, fatter tails, and often (but not always) a sharper peak.

$$\boldsymbol{Kurt = \frac{1}{s^4}\sum_{X_{i=1}}^{n}(X_i - \bar{X})^4 P(i,j)} \tag{7}$$

Skewness: It is a measure of the asymmetry of the probability distribution of a real-valued random variable. The skewness value can be positive or negative, or even undefined. Qualitatively.

$$Skew = \frac{1}{s^3}\sum_{X_{i=1}}^{n}(X_i - \bar{X})^3 P(i,j) \tag{8}$$

IDM: IDM **(**Inverse Difference Moment) is a measure of image texture as called homogeneity that measures the local homogeneity of an image.

$$IDM = \sum_{X_{i=1}}^{n} \frac{1}{1+(i-j)^2} P(i,j) \tag{9}$$

Contrast: It is a measure of intensity of a pixel and its neighbour over the image. In the visual perception of the real world, contrast is determined by the difference in the colour and brightness of the object and other objects within the same field of view.

$$Con = \sum_{i,j}|i-j|^2 p(i,j) \tag{10}$$

Correlation: This measures the linear dependency of grey levels of neighbouring pixels.

$$\boldsymbol{Cor} = \sum_{i,j}\frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i\sigma_j} \tag{11}$$

Energy: Energy feature measures the uniformity of the intensity level distribution. If the value is high, then the distribution is to a small number of intensity levels.

$$Energy = \sum_{i,j} p(i,j)^2 \tag{12}$$

Homogeneity: It is the closeness of the distribution of elements in the GLCM.

$$Hom = \sum_{i,j}\frac{p(i,j)}{1+|i-j|} \tag{13}$$

5. RESULTS

In simulation, the proposed method was tested on several images. Results on five such MRI images are shown in fig. 3, fig. 4, table 1, table 2, table and table 4. In fig. 3 and fig. 4 Benign Tumor and Malignant tumor MRI Images with various features of the data set have been considered like Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM, Contrast, Correlation, Energy, Homogeneity etc. for the classification methods.

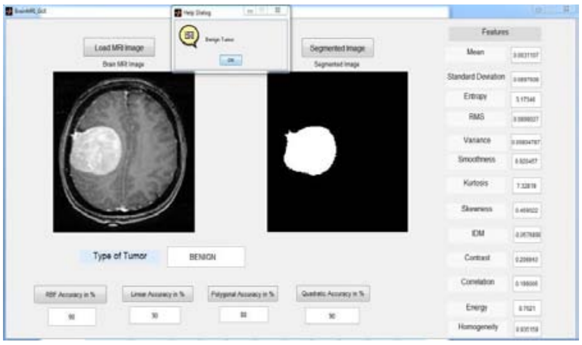

Figure 3: Benign Tumor Detection with Various Features

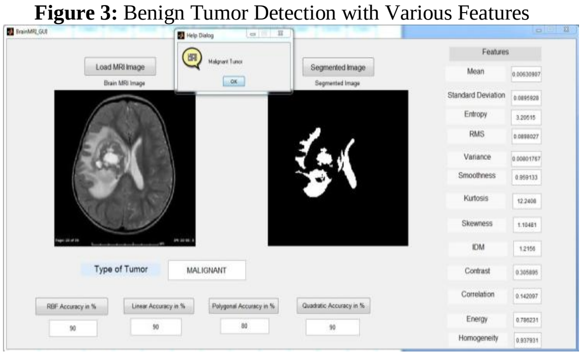

Figure 4: Malignant Tumor Detection with Various Features

These are shown in table 1 in which the benign data set is used and the parameters are calculated.
In table 2 the malignant data set is used and the calculations are performed on it for the different features.

Table 3 shows the accuracy results for different kernel functions for benign data set and this table results proves that linear accuracy method gives the good results.

Table 4 shows the accuracy results for different kernel functions for malignant data set and this table results proves that linear accuracy method gives the good results.

Table 1: Accuracy Observed using different Features for Benign Data

Benign Tumour Data	Image 1	Image 2	Image 3	Image 4
Mean	0.0031089	0.00235179	0.0033116	0.00251283
Variance	0.00803998	0.00804987	0.00802012	0.0080570
RMS	0.0899152	0.0898152	0.0898232	0.0898734
Smoothness	0.919982	0.901233	0.932552	0.904924
Entropy	3.17432	3.270175	3.58127	3.32012
Skewness	0.471291	0.796293	0.644031	0.309932
IDM	0.060312	0.501279	0.53519	0.570182
Kurtosis	7.41412	8.01242	6.28321	6.24193
Correlation	0.200932	0.0929923	0.101026	0.137261
Energy	0.8032	0.78903	0.767239	0.769347
Contrast	0.210239	0.269351	0.23993	0.190236
Homogeneity	0.940236	0.941234	0.930217	0.94206
Std. Deviation	0.0896908	0.0897839	0.0897562	0.0902157

Table 2: Accuracy Observed using different Features for Malignant Data

Malignant Tumour Data	Image 1	Image 2	Image 3	Image 4
Mean	0.00593421	0.00392375	0.00461531	0.00395298
Variance	0.00799794	0.00793456	0.00793467	0.00795667
RMS	0.0903452	0.0901754	0.0902478	0.0905278
Smoothness	0.960346	0.930565	0.935543	0.939789
Entropy	3.19564	3.5982	3.55632	3.54973
Skewness	0.98459	0.538916	0.6193456	0.509456
IDM	0.98488	0.37023	0.495677	0.307863
Kurtosis	11.4532	6.29674	7.1207	5.98346
Correlation	0.146743	0.13572	0.190125	0.167347
Energy	0.804373	0.754785	0.72956435	0.750346
Contrast	0.298305	0.230456	0.245985	0.243689
Homogeneity	0.929805	0.930275	0.925836	0.93056
Std. Deviation	0.0903184	0.0903472	0.0903473	0.0906278

Table 3: Accuracy Observed using different Kernel Functions for Benign Data Set

Benign Tumour Data	Quadratic Accuracy in %	Polynomial Accuracy in %	Linear Accuracy in %	RBF Accuracy in %
Image 1	89	81	87	88
Image 2	78	79	87	77
Image 3	72	71	87	69
Image 4	82	78	85	67
Image 5	72	77	78	82

Table 4: Accuracy Observed using different Kernel Functions for Malignant Data Set

Malignant Tumour Data	Quadratic Accuracy in %	Polynomial Accuracy in %	Linear Accuracy in %	RBF Accuracy in %
Image 1	91	78	88	85
Image 2	85	79	89	82
Image 3	75	73	90	74
Image 4	79	78	67	74
Image 5	72	79	84	79

5. **CONCLUSION**

In this presented study, the image extraction techniques and the features extraction techniques are learned. The study addresses the issues of image search in medical domain, and their features by which the images are accurately distinguished from the database. The brain tumor identification could be a nice facilitate for the physicians and a privilege for the medical imaging and industries engaged on the assembly of CT scan and MRI imaging. The MATLAB simulation is carried on totally different brain pictures and tumor is detected using OTSU'S methodology for image segmentation and optimum global thresholding. The nonlinear SVM works in a very great way on elevated dimensional characteristic sets. To extend the periphery of the classification, kernel functions are used. There are various kernels utilized in SVM like linear, polynomial radial basis function (RBF) etc. Kernel SVM provides the clear understanding of the classification and really simple to use in sensible image processing. During this approach some of its methods like RBF, linear, and polygonal are used to find the segmented image. Number of features from the data set have been considered like RMS, Smoothness, Entropy, Variance, Mean, Contrast, Correlation, Standard Deviation, Energy, Homogeneity etc. for the classification methods.

Otsu's segmentation works very well for Brain MRI dataset to segment and PCA performed higher results for feature reduction. Specifically, we have analysed and compared four of the kernel functions of SVM with a paradigm brain MRI data. In future we are able to use others feature reduction technique and can compare and might compare and realize the accuracy. The scientist also can use numerous hybridized learning algorithms with sot of dataset.Natural language processing contain a field in which we can automatically classify sentiments related to text information, this field is named as sentiment analysis.

REFERENCES

[1] Y. Zhang, L. Lenan Wu, "An MR brain images classifier via principal component analysis and kernel support vector machine" Progress In Electromagnetic Research, (**2012),** 130, 369-388

[2] Ghotekar, Bhavana B. and K. J. Mahajan, "MRI Brain image Segmentation and Classification: A Review", International Research Journal of Engineering and Technology (IRJET), (**2016**), 3, 1170-1176.

[3] Vinayadth V. Kohir and Sahebgoud H. Karaddi, "Detection Of Brain Tumor Using Back-Propagation And Probabilistic Neural Network", Proceedings of 19th IRF International Conference, 25th January 2015, Chennai, India, (**2015**) ISBN: 978-93-84209-84-1, 74-80

[4] J. selvakumar, A. Lakshmi, T. Arivoli, "Brain Tumor Segmentation and Its Area Calculation in Brain MR Images using K-Mean Clustering and Fuzzy C-Mean Algorithm*",* IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), Nagapattinam, Tamil Nadu, India, (**2012**), 87-95.

[5] Dilpreet Kaur, Yadwinder Kaur, "Various Image Segmentation Techniques: A Review", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 3, Issue. 5,(**2014**), ISSN 2320–088X, pg.809 – 814

[6] Panda et al., "Role of classification algorithms on brain MRI data"; International Journal of Pharma and Bio Sciences,(**2016**) ISSN 0975-6299; 7(4): (B), 777- 782

[7] Aslam, Daxiang Cui, "Brain Tumor Detection from Medical Images: A Survey", Nano Biomed. Eng., (**2017**), 9, 72-81.

[8] Sheejakumari, Gomathi, "Brain tumor detection from MRI images using histon based segmentation and modified neural network", Biomed Res- India (**2016**) Special Issue ISSN 0970-938X, 134-141.

[9] Deepali Agarwal, S R Dogiwal, " Novel Approach for Digital Image Watermarking Using Lsb and its applications Medical Image Encryption and Decryption Based on DCT Domain", Published in "SKIT Research Journal, ISSN2278-2508, (2018), 8(2), 12-18.

[10] Sanwta Ram Dogiwal, Y. S. Shishodia, Abhay Upadhyaya, "Super Resolution Image Reconstruction Using Wavelet Lifting Schemes and Gabor Filters*",* Published in Confluence, Fifth International conference 2014, Noida, India, (**2014**) 978-1-4799-4236-7/ IEEE.