Learning Texture Features from GLCM for Classification of Brain Tumor MRI Images using Random Forest Classifier

ASHWANI KUMAR AGGARWAL
Department of Electrical and Instrumentation Engineering
Sant Longowal Institute of Engineering and Technology
Longowal-148106, Punjab, INDIA

Abstract—In computer vision, image feature extraction methods are used to extract features so that the features are learnt for classification tasks. In biomedical images, the choice of a particular feature extractor from a diverse range of feature extractors is not only subjective but also it is time consuming to choose the optimum parameters for a particular feature extraction algorithm. In this paper, the focus is on the Grey-level co-occurrence matrix (GLCM) feature extractor for classification of brain tumor MRI images using random forest classifier. A dataset of brain MRI images (245 images) consisting of two classes viz. images with tumor (154 images) and images without tumor (91 images) has been used to assess the performance of GLCM features on random forest classifier in terms of accuracy, true positive rate, true negative rate, false positive rate, false negative rate derived from the confusion matrix. The results show that by using optimum parameters, the GLCM feature extracts significant texture component in brain MRI images for promising accuracy and other performance metrics.

Keywords—statistical texture features, image matching, descriptors, dissimilarity, entropy

Received: April 15, 2021. Revised: February 16, 2022. Accepted: March 20, 2022. Published: April 19, 2022.

1. Introduction

Brain tumors occur due to abnormal growth of cells [1]. Like other tumors, the brain tumors may be benign tumors or malignant tumors. A malignant tumor usually spreads more rapidly than a benign tumor [2]. If a brain tumor starts from the brain, it is known as primary brain tumor, whereas a tumor which spreads to brain from other parts of the body is known as secondary brain tumor. The symptoms of brain tumor vary from a person to person, however, some of the common symptoms are nausea, vomiting, headache, seizures, and loss of balance, etc. There are hundreds of types of brain tumors, the most common types are gliomas, meningioma, and medulloblastoma [3]. The gliomas brain tumor may be low-grade astrocytoma which is a slow growing tumor and is usually benign or glioblastoma multiforme which is rapidly growing and is usually malignant [4]. The meningioma usually starts from brain and usually is benign. Medulloblastoma is more common in children than in adults [5]. The brain tumor imaging modalities help the radiologists and oncologists during pre-therapy (for assess the lesion extent, grading), therapy (delineation), and post-therapy (therapy response, monitoring) [6]. The brain tumor imaging modalities include mass spectroscopy, brain perfusion (CT and MRI), diffusion tensor imaging (DTI), and functional magnetic resonance imaging (fMRI), etc. [7]. The magnetic resonance imaging (MRI) uses a very strong magnet and radio waves to image the brain [8]. All the metal objects need to be removed before taking MRI scan as the metallic objects interfere with the magnetic field and causes erroneous signal. The patients with metallic implants such as pacemaker are sometimes undergone alternative imaging modality by the radiologist. The MRI imagers are closed type of open type. In case of open type MRI scanners, the image quality is not as good as that obtained with closed type MRI scanners [9]. The magnet produces a strong magnetic field and rf coils send the radio waves in the brain. After the radio waves are stopped, the MRI imager receives the energy signals from the body to image the brain [10]. The MRI images are sometimes taken with a dye to enhance the contrast of the image. This paper is organized as follows. Section II discuss the related work. Section III presents the dataset and methodology. Section IV presents the results obtained and the discussion. Section V concludes the work.

2. Related Work

The automatic detection of tumors in medical images is challenging task. The feature section and the choice of classifier is very important for the detection results to give a significant accuracy. The features derived from the grey-level co-occurrence matrix have been used to classify the brain tumors using a two-layered artificial neural network [11]. The accuracy obtained using the method is 97.5%. The brain images are classified into four classes viz. astrocytoma, meningioma, metastatic bronchogenic carcinoma, and sarcoma. A total of 16 features are extracted based on correlation, contrast, dissimilarity, energy, homogeneity, variance, and entropy. The conditional entropy is used using evolutionary algorithm [12].

The segmentation and the classification of brain tumors has been done by using texture features derived from GLCM and using wavelets. The brain MRI images are segmented using Otsu thresholding method and then k-means clustering is performed. The texture features are obtained from GLCM along with features extracted using wavelets. The combined features are given to PCA before classifying the images with support vector machines. It is shown that linear kernels outperform than polynomial kernels for the classification of brain MRI images [13]. The image retrieval using features based on ttexture and shape are used [14]. An overview of methods for brain tumor images using features derived from grey-level co-occurrence matrix and classified using artificial neural networks is given [15]. Brain MRI images are classified using various features obtained from GLCM and kernel support vector machines as a classifier. The feature reduction module is used after extraction of features and before giving the features to the kernel support vector machines for classification [16]. The classification of brain tumor MRI images is done using schematic segmentation and support vector machine-based classification of features extracted from grey-level co-occurrence matrix [17]. The images are pre-processed using median filtering. the method is shown to have accuracy of 93.05% for the classification.

3. Dataset and Methodology

3.1 Dataset

The dataset consists of brain tumor MRI images. The total number of images is 245. The number of images with tumor is 154 and the number of images without tumor is 91. The size of the dataset is 7.22 MB. The image size of images in the dataset is not same, for example, some of the images have spatial resolution of 300x168 while others have spatial resolution of 200x200. The data size is also not same as some of the images are 4.46 KB while others are 5.76 KB. The image format of the images in the dataset is jpg format.

3.2 Methodology

The features such as Scale Invariant Feature Transform (SIFT), Speeded-Up Robust features (SURF), and Oriented FAST and Rotated BRIEF (ORB) are used for image classification [18]. These feature extractors extract geometric, color, textural, and shape information to construct a feature vector that represents the small neighborhood around the feature point. SIFT is 128-dimensional feature vector that is robust to many affine transformations, illumination changes, and sensor noise [19]. The SIFT features extracted on the brain tumor MRI images are shown in Fig. 1. As the dimensions of SIFT are 128 and there are hundreds of such features in a single image, it becomes computationally expensive to compute SIFT features for brain tumor MRI image classification and therefore, SIFT has not been used in this work.

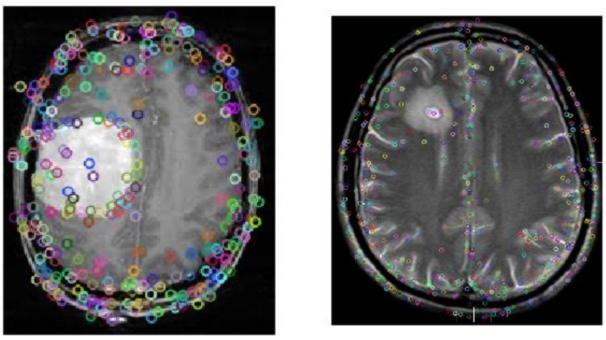

Fig. 1. Extracted SIFT fesatures

The texture of an image is spatial variation of pixel intensities in an image. If the pixel intensities vary a lot, that means the texture is coarse and if the pixel intensities do not vary much, in that case, the texture is smooth. The texture is very useful clue for classifying the image contents. The texture descriptors may be obtained based on the direction, color, and contrast. The texture analysis of an image can be done at various resolutions. The texture features can be classified as statistical, structural, and model based methods. The statistical texture features are extracted based on correlation of neighboring pixels, frequency of occurrence of pairs of pixel intensity values in a neighborhood, and entropy. The statistical features may be first order features (mean, median, variance, entropy) or second order features (relationships between groups of pixels). In this paper, the texture features are extracted using grey-level co-occurrence matrix (GLCM). The GLCM is a second order statistical texture feature extraction method [20]. The GLCM maps the input image into a table that represents the number of occurrences of a pair of pixel values at a certain distance and angle. The angle and distance values are varied over a range of 0-360 degree and 1 pixel to 8 pixels respectively. The GLCM is a two-dimensional array where each dimension of the array is same and is equal to the number of greyscale levels in the image [21]. For 8-bit image, the number of levels is 256, and therefore, the GLCM has 256 rows and 256 columns. There is one such 256x256 matrix for each combination of distance (offset) and angle value. Each diagonal cell represents the texture homogeneity in the image, whereas off-diagonal cells count for texture heterogeneity in the image. The matrix is made symmetric by adding the matrix with its transpose. The matrix is then normalized by dividing each element by the sum of all the elements in the matrix. The GLCM is used to obtain the second-order statistical texture features such as angular second moment, entropy, variance, correlation, inverse difference moment, energy, dissimilarity, homogeneity, and contrast etc. The energy is obtained by taking the square root of angular second moment. The statistical features in an image give several features in the image that are useful for the classification of brain MRI images. The following statistical texture features have been obtained from GLCM for the brain tumor MRI images.

1) Energy feature: The energy feature is the square root of angular second moment feature. This feature represents the uniformity in the image texture. The energy feature obtained from GLCM is given by (1). The maximum value of the energy is unity as GLCM is a normalized matrix.

$$E = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \{P(i,j)\}^2} \qquad (1)$$

2) Dissimilarity Feature: The dissimilarity represents the heterogeneity in the image texture. The dissimilarity feature is obtained by multiplying the GLCM cell values with linear weights as given by (2).

$$D = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |i-j| P(i,j) \qquad (2)$$

3) Homogeneity Feature: The large homogeneity in the image texture is represented by large values of the diagonal elements of the GLCM. If the image pixel values are same, the homogeneity is maximum. A large contrast reduces the homogeneity in the image. The homogeneity feature is given by (3).

.

$$H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i,j)}{1+(i-j)^2} \quad (3)$$

4) Contrast Feature: The contrast feature is a measure of difference between the smallest pixel value and the largest pixel value in a group of pixels. The contrast feature is given by (4).

$$H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)(i-j)^2 \quad (4)$$

A feature vector is obtained from these four features by extracting these features for various values of offset and angles. A step size of 1 and 3 pixels is taken and an angle of 0, 45, and 900 degree is taken.

The learnt features are given to a classifier to classify the brain tumor MRI images. The decision tree, support vector machines, k-nearest neighbor, and random forest classifiers are some of the common classifiers used for image classification tasks. In this paper, we have used random forest classifier for the classification of the brain tumor MRI images. The other statistical tests like Chi-square test and t-tests are conducted to validate the effectiveness of the method.

4. Resuls and Discussion

The dataset is divided into training and testing datasets. The training and testing sets are prepared from the given dataset by dividing it into a ratio of 0.8:0.2. The method is repeated several times to order to make the model robust towards various types of noise. The features are extracted from the training dataset. The random forest classifier is trained using the features extracted from the training dataset. Thereafter, the features are extracted from testing datasets.

4.1 Actual and predicted Class labels

The classifier is tested for the performance on the features extracted from the testing dataset. The results of the actual class labels and predicted class labels for 18 test images is shown in Fig. 2.

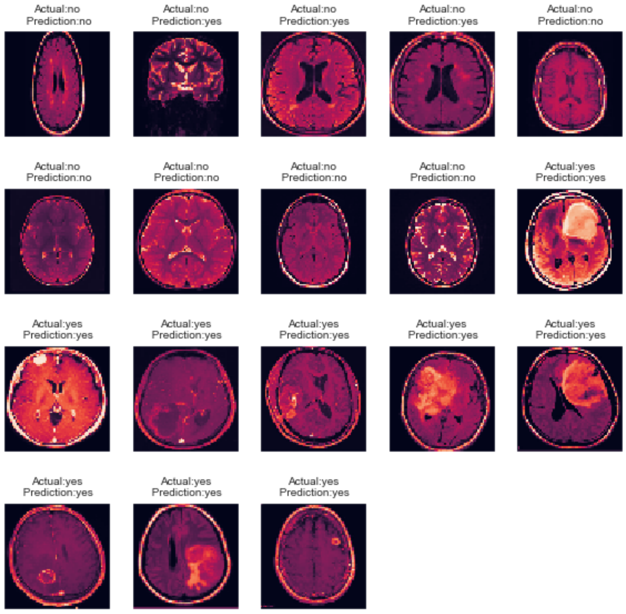

Fig. 2. Actual and Predicted class labels

It is observed that all the images with actual class 'yes', are predicted correctly in all the 9 test images. However, the out of 9 images with actual class 'no', only 6 are predicted correctly as 'no' whereas, rest of the 3 are predicted falsely as 'yes'.

4.2 Confusion Matrix

A confusion matrix is obtained for evaluate the performance of the method. The confusion matrix is put in Fig. 3.

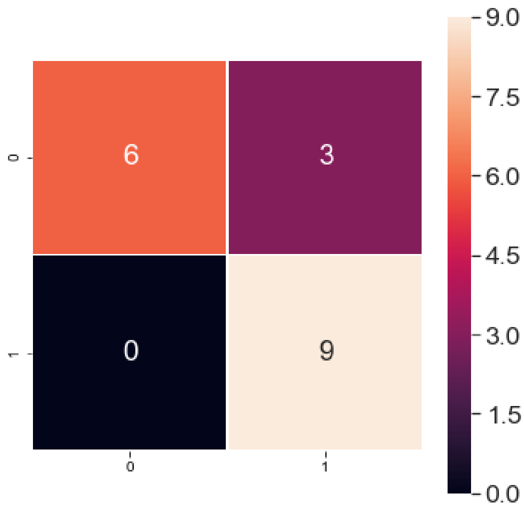

Fig. 3. Confusion matrix

It is observed from the confusion matrix that there is no predicted class 'no' which is actually 'yes'. The number of images tested is 18. Out of 18 images, 9 belong to class 'yes' and 9 belong to class 'no'.

4.3 Performance with various Offset and Angle of GLCM

The GLCM is obtained for various combinations of steps (offsets) and the angles to count the number of occurances of a pair of pixel intensity values. A comparison of calssification accuracy for various values of steps and angles is given in Table I.

TABLE I. COMPARISON OF CLASSIFICATION ACCURACY FOR VARIOUS OFFSET AND ANGLE VALUES OF GLCM

Classification Accuracy			
Offset (pixels)	***Angle (degree)***	***32 x 32***	***64 x 64***
1, 3	0, π/4	61.1%	83.3%
1, 3, 5	0, π/4	72.2%	66.7%
1, 3	0, π/4, π/2	72.2%	66.7%
1, 3, 5	0, π/4, π/2	66.7%	72.2%
1, 3, 5, 7	0, π/4	44.4%	66.7%
1, 3, 5, 7	0, π/4, π/2	83.3%	66.7%

The performance of the method drops significantly if the spatial resolution of the image is reduced. For the same combination of offset (1 pixel and 3 pixels), and angles (0, π/4), the accuracy reduces from 83.3% to 61.1% on reducing the spatial resolution from 64 x 64 to 32 x 32.

5. Conclusion

The brain tumor MRI images are classified into two classes viz. images with tumor and images without tumor. The grey-level co-occurrence matrix obtained from the brain tumor MRI images gives statistical texture features which are computationally fast as compared to SIFT features. the random forest classifier classifies the images with an accuracy of 83.3 % on a limited dataset. the performance of the method could be improved by applying some pre-processing operations on the images. Another feature extraction methods may be combined with GLCM features to enhance the accuracy. The method can be extended by using additional layers in the model.

Acknowledgment

The authors are thankful to their colleagues for proofreading this paper before submission.

References

[1] Surawicz, Tanya S., et al. "Brain tumor survival: results from the National Cancer Data Base." Journal of neuro-oncology 40.2 (1998): 151-160.

[2] Albini, A., et al. "A rapid in vitro assay for quantitating the invasive potential of tumor cells." Cancer research 47.12 (1987): 3239-3245.

[3] Waage, Ingunn Syversen, Ingeborg Vreim, and Sverre Helge Torp. "C-erb B2/HER2 in Human Gliomas, Medulloblastomas, and Meningiomas: a Minireview." International journal of surgical pathology 21.6 (2013): 573-582.

[4] A. K. Aggarwal, "GPS-Based Localization of Autonomous Vehicles," Autonomous Driving and Advanced Driver-Assistance Systems (ADAS): Applications, Development, Legal Issues, and Testing, p. 437, 2021.

[5] Mathew, A. Reema, P. Babu Anto, and N. K. Thara. "Brain tumor segmentation and classification using DWT, Gabour wavelet and GLCM." 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). IEEE, 2017.

[6] A. Kaur, A. P. S. Chauhan, and A. K. Aggarwal, "An automated slice sorting technique for multi-slice computed tomography liver cancer images using convolutional network," Expert Systems with Applications, vol. 186, p. 115686, 2021.

[7] Jellinger, K. "Glioblastoma multiforme: morphology and biology." Acta neurochirurgica 42.1 (1978): 5-32.

[8] Rochkind, Semion, et al. "Extracranial metastases of medulloblastoma in adults: literature review." Journal of Neurology, Neurosurgery & Psychiatry 54.1 (1991): 80-86.

[9] A. K. Aggarwal, "Biological Tomato Leaf Disease Classification using Deep Learning Framework," International Journal of Biology and Biomedical Engineering, vol. 16, no. DOI: 10.46300/91011.2022.16.30, pp. 241–244, 2022.

[10] Huang, Raymond Y., et al. "Pitfalls in the neuroimaging of glioblastoma in the era of antiangiogenic and immuno/targeted therapy–detecting illusive disease, defining response." Frontiers in neurology 6 (2015): 33.

[11] Dimou, S., et al. "A systematic review of functional magnetic resonance imaging and diffusion tensor imaging modalities used in presurgical planning of brain tumour resection." Neurosurgical review 36.2 (2013): 205-214.

[12] Katti, Girish, Syeda Arshiya Ara, and Ayesha Shireen. "Magnetic resonance imaging (MRI)–A review." International journal of dental clinics 3.1 (2011): 65-70.

[13] M. Garg, J. S. Ubhi, and A. K. Aggarwal, "Neural Style Transfer for Image within Images and Conditional GANs for Destylization," Journal of Visual Communication and Image Representation, 2022.

[14] Foster- Gareau, Paula, et al. "Imaging single mammalian cells with a 1.5 T clinical MRI scanner." Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 49.5 (2003): 968-971.

[15] Armstrong, Peter, and Stephen F. Keevil. "Magnetic resonance imaging--1: Basic principles of image production." BMJ: British Medical Journal 303.6793 (1991): 35.

[16] Zulpe, Nitish, and Vrushsen Pawar. "GLCM textural features for brain tumor classification." International Journal of Computer Science Issues (IJCSI) 9.3 (2012): 354.

[17] Chauhan, Sumika, Manmohan Singh, and Ashwani Kumar Aggarwal. "An effective health indicator for bearing using corrected conditional entropy through diversity-driven multi-parent evolutionary algorithm." Structural Health Monitoring (2020): 1475921720962419.

[18] Arora, Kratika and Ashwani Kumar Aggarwal. "Approaches for Image Database Retrieval Based on Color, Texture, and Shape Features." Handbook of Research on Advanced Concepts in Real-Time Image and Video Processing, edited by Md. Imtiyaz Anwar, et al., IGI Global, 2018, pp. 28-50. http://doi:10.4018/978-1-5225-2848-7.ch002

[19] Jain, Shweta. "Brain cancer classification using GLCM based feature extraction in artificial neural network." International Journal of Computer Science & Engineering Technology 4.7 (2013): 966-970.

[20] Kadam, Megha, and Avinash Dhole. "Brain tumor detection using GLCM with the help of KSVM." International Journal of Engineering and Technical Research 7.2 (2017).

[21] Hussain, Ashfaq, and Ajay Khunteta. "Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features." 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2020.