Performance Analysis of KNN Classifier with and Without GLCM Features In Brain Tumor Detection

Bharanidharan N, Harikumar Rajaguru, Geetha V

Abstract: This paper presents the K-Nearest Neighbour (KNN) classifier combined with Grey Level Co-occurrence Matrix(GLCM) feature extraction technique for brain tumor detection using MATLAB software. Thirty MRI images obtained from the clinical sources are analyzed in this study. Though KNN classifier is less time consuming, it has a disadvantage of less accuracy. To improve the accuracy of KNN classification, GLCM feature extraction is used. Performance of KNN classifier is analyzed with and without GLCM feature extraction.

Keywords: Tumor detection- Classifer- Feature extraction- KNN-GLCM

I. INTRODUCTION

An anomalous growth of cells present in brain is known as brain tumor. This type of abnormal tissue development may affect the functions of brain and causes a pressure that turns into pain [1]. This pressure may also leads to shifting of some brain cells and injury of other healthy cells. Brain tumor may arise at any age and the root cause for this is not exactly known to researchers and doctors in most of the cases. The World Health Organization (WHO) created the tumor standards by which different tumors are classified. WHO classifies the brain tumor into more than 120 categories depending on the location and tumor cell type. The major widespread tumor types includes Meningiomas, Gliomas, Pituitary tumors and Schwannomas [2].

In medical terms, brain tumor is a diversified collection of neoplasms arising from meningles and intracranial cells with the tumor ranging from benign to aggressively malicious [3]. Each and every type of tumor has to be treated in different manner. In some cases, a benign tumor can turn dangerous due to its location in brain. This causes the classification of brain tumor as much difficult task. Generally diagnosis of brain tumor is carried out using Computed Tomography (CT), Magnetic Resonance Imaging (MRI) or neurological exam. Other tests include angiogram, biopsy and spinal tap. MRI is one of the most popular methods to diagnose various diseases. This technique captures the images of internal structures of body in a non-invasive way and provides the platform to visualize the interior tissues of the organs more clearly. MRI is based on the Nuclear Magnetic Resonance (NMR) technique. At the time of MR imaging, the patient will be positioned on a very strong magnetic field. The applied magnetic field makes the protons present in the water molecules of patient body to arrange themselves in either parallel or anti parallel direction with respect to the magnetic field applied. As a second step, Radio Frequency (RF) pulses are sent across the magnetic field which causes the aligned protons to come out of the table state. Then the RF pulse is stopped which makes the protons return to original state which in turn cause an electrical signal. Finally RF coil presents in the MRI scanner will collect the electrical signal, decode it to produce the image [4]. This imaging technique plays a vital role in brain tumor detection since it has ability to distinguish various soft tissues like grey matter and white matter.Doctors choose their treatment methods for a patient after integrating their medical knowledge and MRI brain images. But there will be number of brain images taken at different angles of brain. It will be tedious task to view the number of MRI images for every patient and segmenting brain tumors. Therefore an automated brain tumor detection based on the computer aided techniques will be helpful in doing the above task. Many techniques were proposed in last few decades for brain tumor automated detection but still there is absence of routine automated technique to be used in hospitals due to robustness and accuracy issues [5]. This paper is organized as follows: Second section introduces the GLCM based feature extraction technique, third section deals with the KNN classifier for brain tumor detection and fourth section explains the KNN classifier combined with GLCM feature extraction. Results are discussed in fifth section and the paper is concluded in section VI.

II. GLCM FEATURE EXTRACTION

Feature extraction is the process of reducing the quantity of resources required to narrate the large amount of data. In particular to feature extraction of images, the large amount of intensity values of image is given as input for feature extraction. The output may be some higher level information like mean, variance, shape, color, etc. Depending upon the problem taken, appropriate features should be chosen. Otherwise the results will be very poor[6]. Haralick et al. introduced Grey Level Co-occurrence Matrix (GLCM), one the most powerful technique for image analysis applications. Initially the input intensity matrix of image is converted into GLCM and then GLCM based texture features are

Revised Manuscript Received on December 08, 2018.

Bharanidharan N, Department of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India

Harikumar Rajaguru, Department of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India

Geetha V, Department of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India

computed. Usually GLCM will be square matrix whose size depends on number of grey levels present in the image. All the intensity grey levels are placed as index for both rows and columns. The GLCM values a(k,l) is compute vy finding the number of repetitions of 'k' pixel horizontally adjacent to 'l' pixel i.e., every element in GLCM represents the number of times pixel 'k' is horizontally adjacent to pixel 'l' as shown in fig. 1

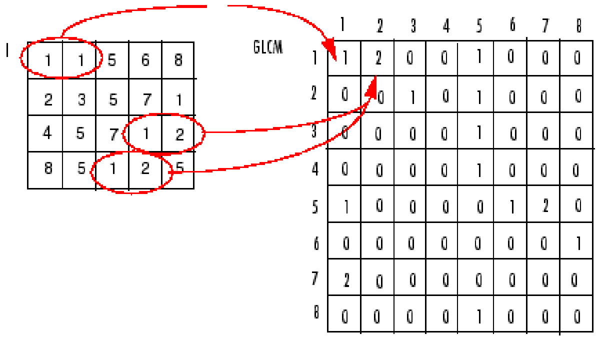

Fig.1. GLCM Generation

Due to the obscure arrangement of various tissues like white matter, grey matter in the brain MRI, extraction of appropriate features is a crucial task. Textural analysis could will help to improve the diagnosis of tumor and identification of tumor stage. The statistics feature formulas which will be useful are listed below.

Energy- It involves the summation of squared elements of GLCM values. The value of energy range between 0 to 1 and constant image intensity values will have energy as 1.

$$Energy = \sum_{i,j} p(i,j)^2 \quad (1)$$

Homogeneity- It is a measure of proximity of a distributed elements in GLCM to the Diagonal values of GLCM. The value of homogeneity varies from 0 to 1 where the highest value is achieved when a GLCM is diagnonal matrix.

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (2)$$

Contrast- It is a parameter that relates the intensity contrast between a pixel and its neighbor. Local changes of GLCM will be considered to compute contrast. Constant image intensity values will have contrast as zero.

$$Contrast = \sum_{i,j} |j|^2 p(i,j) \quad (3)$$

Dissimilarity- This parameter relates the difference in intensity level between a pixel and its neighbor. Constant image intensity values will have dissimilarity as infinite.

$$Dissimilarity = \sum_{i,j=1}^{N} P_{i,j}\ |i-j| \quad (4)$$

III. KNN CLASSIFIER FOR DETECTION OF BRAIN TUMOR

Generally a classifier is used to categorize the image based on their characteristics. K-Nearest Neighbour (KNN), one of the simple yet powerful classifier used to categorize objects depending on the nearest training samples present in the feature space. It is mainly used when there is no prior knowledge about the distribution of data. KNN classifier comes under the group of instance based learning techniques. After training, the test object will be categorized as the majority class represented by its K neighbours. If K=1, then KNN algorithm will deduced to Nearest Neighbour (NN) classifier [6-10]. The algorithm for KNN classifier is as follows:

- Calculate the distance between a test sample and all the training samples (Samples with known class)
- Arrange the distance in the ascending order and choose the K number of training samples with minimum distance on comparison with other training samples.
- In the K number of training samples, find out the label class which belongs to the majority training samples
- Assign that label to the test sample.

The operation of KNN classifier mainly depends on the choice of K and distance metric used. If the value of K is very small, then there are chances for leaving many data points unlabelled while the larger value of K may leads to overlapping of classes.

IV. KNN CLASSIFIER COMBINED WITH GLCM FEATURE EXTRACTION

For brain tumor detection, various complicated techniques like Neural Networks (NN), Fuzzy logic, Swarm intelligence, etc can be used. Even though they provide good accuracy of classification, they are more time consuming. KNN classifier consumes less time for classification but lags in accuracy. Hence KNN classifier combined with GLCM feature extraction is used and this method provides trade-off between accuracy and time consumption. The 50 MRI images of brain are obtained from the clinical sources and 30 brain images belongs to the non-tumor class and 20 brain images belongs to the tumor class. Out the 30 non-tumor subjects, 20 subjects will be used for training and 10 subjects will be used for testing. Similarly, out of the 20 tumor subjects, 10 subjects will be used for training and 10 subjects will be used for testing. Two cases are considered and performances of both the cases are analyzed.

Case-I: Brain tumor detection using KNN classifier without GLCM. Each image is splitted into 16 regions and four features are calculated for each region. All the 64 features value of the image will be given to KNN classifier.

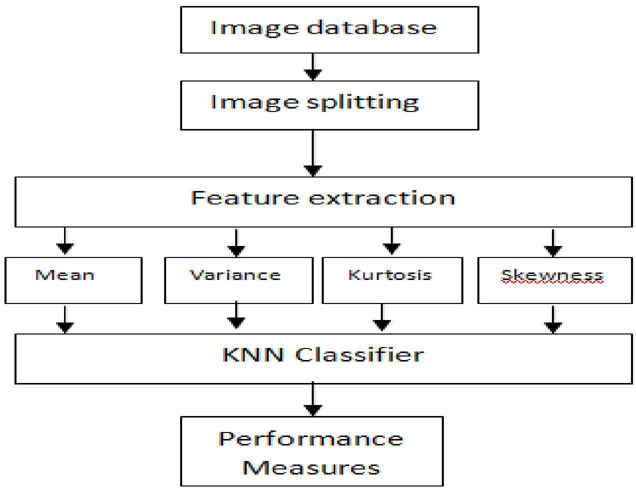

Fig.2. Brain tumor detection using KNN classifier without GLCM

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

Case-II: Brain tumor detection using KNN classifier with GLCM. Each image is splitted into 16 regions and 16 GLCM matrixes are computed. Then from each GLCM matrix, contrast, dissimilarity, energy, homogeneity are obtained using equations (1) to (4). Finally all the 64 features value of the image will be given to KNN classifier.

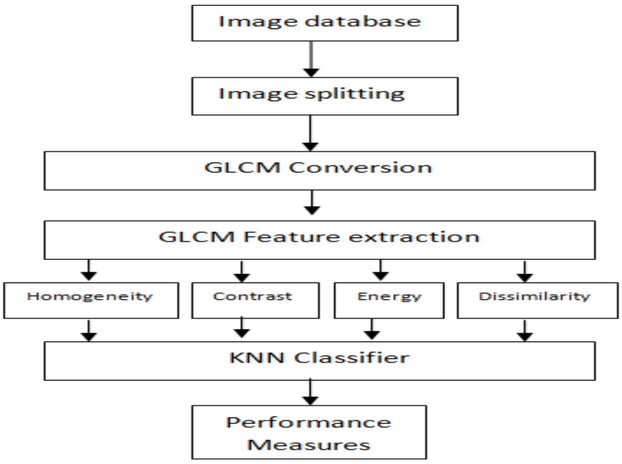

Fig.3. Brain tumor detection using KNN classifier with GLCM

V. RESULTS AND DISCUSSION

The confusion matrix is used to represent the performance accuracy of a classifier on a test data for which original true class is known. Different performance measures such as accuracy, sensitivity, specificity, error rate and precision can be derived from the above said matrix. For binary classification, the confusion matrix will have the following metrics:

True positive (TP): number of true positives - perfect positive prediction

False positive (FP): number of false positives - imperfect positive prediction

True negative (TN): number of true negatives - perfect negative prediction

False negative (FN): number of true negatives - imperfect negative prediction

Table 1: General Confusion Matrix for Brain Tumor Detection

	Predicted as tumor	Predicted as non-tumor
Actual tumor	TP	FN
Actual non-tumor	FP	TN

Error rate

Error rate (ERR) is computed as the fraction of total number of imperfect predictions to the total number of test data. The finest possible error rate is 0.0, whereas the very worst is 1.0. Minimization of this error rate will be the prime objective for any classifier.

$$ERR = \frac{FP+FN}{TP+TN+FN+FP} \qquad (5)$$

Accuracy

Accuracy (ACC) is found as the fraction of total number of perfect predictions to the total number of test data. It can also be represented as 1 – ERR. The finest possible accuracy is 1.0, whereas the very worst is 0.0.

$$ACC = \frac{TP+TN}{TP+TN+FN+FP} \qquad (6)$$

Sensitivity (Recall or True positive rate)

Sensitivity (SN) is found as the fraction of number of perfect positive predictions to the total number of positive predictions. The finest possible sensitivity is 1.0, whereas the very worst is 0.0.

$$SN = \frac{TP}{TP+FN} \qquad (7)$$

Specificity (True negative rate)

Specificity (SP) is found as the fraction of number of perfect negative predictions to the total number of negatives. The finest possible specificity is 1.0, whereas the very worst is 0.0.

$$SP = \frac{TN}{TN+FP} \qquad (8)$$

Precision (Positive predictive value)

Precision (PREC) is computed as the fraction of perfect positive predictions to the total number of positive predictions. The finest possible precision is 1.0, whereas the very worst is 0.0.

$$PREC = \frac{TP}{TP+FP} \qquad (9)$$

False positive rate

False positive rate (FPR) is found using a fraction of imperfect negative predictions to the total number of negatives. The finest possible false positive rate is 0.0 whereas the very worst is 1.0. It can also be computed as 1 – specificity.

$$FPR = \frac{FP}{TN+FP} \qquad (10)$$

The accuracy of KNN classifier will be checked for each K value and optimum value of K is found iteratively to get the maximum accuracy as shown in fig.4. The maximum value of K is chosen as 13 at which maximum accuracy is achieved for KNN classifier with GLCM.

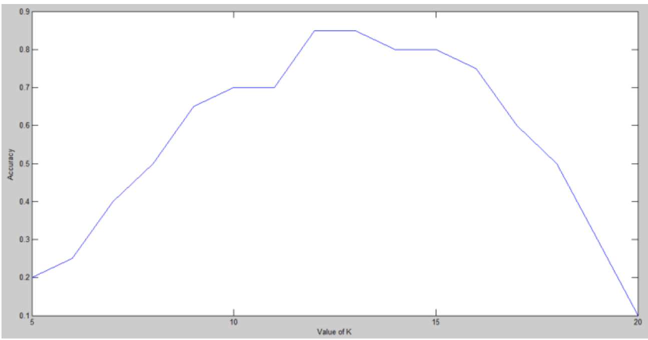

Fig.4. Tuning of K value to get maximum accuracy

After classification using KNN, confusion matrix is constructed and performance measures are calculated using equations (5)-(10). The confusion matrix for both cases and comparison table for the performance measures of KNN classifier with GLCM and without GLCM is shown below.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 2: Confusion Matrix of KNN Classifier without GLCM

n=20	Predicted as tumor	Predicted as non-tumor
Actual tumor	5	5
Actual non-tumor	3	7

Table 3: Confusion Matrix of KNN Classifier with GLCM

n=20	Predicted as tumor	Predicted as non-tumor
Actual tumor	8	2
Actual non-tumor	1	9

Table 4: Performance Comparison of KNN Classfier with and Without GLCM

Parameters	KNN WITHOUT GLCM	KNN WITH GLCM
Error Rate	0.4	0.15
Accuracy	0.6	0.85
Sensitivity	0.5	0.8
Specificity	0.7	0.9
Precision	0.625	0.888
False Positive Rate	0.3	0.1

As explained in table 4, the KNN classifier combined with GLCM feature extraction technique outperforms the performance of KNN classifier without GLCM in terms of all the six parameters. This improvement in accuracy is mainly due to GLCM usage along with energy, homogeneity, dissimilarity and contrast features instead of conventional statistical features like mean, variance, kurtosis and skewness.

VI. CONCLUSION

The performance of KNN classifier with and without GLCM feature extraction is analysed for brain tumor detection using MATLAB software. The performance measures of both the cases are calculated based on Confusion matrix. It is observed that accuracy of KNN classifier with GLCM features is much improved (1.4 times) over that of KNN classifier without GLCM. KNN classifier combined with GLCM features has the advantages of improved accuracy and less time consuming. Hence this technique can be used for rapid detection of brain tumor for a large quantity of datasets.

REFERENCES

1. Louis D.N., Ohgaki H., Wiestler O.D, Cavenee W.K. (2017), “WHO Classification of Tumors of the Central Nervous System”, International Agency for Research on Cancer (IARC), Lyon, France
2. Ritu Rana, Parvinder Singh (2015), “Brain Tumor Detection through MR Images: A Review of Literature”, IOSR Journal of Computer Engineering (IOSR-JCE), Volume 17, Issue 5, PP 07-18
3. P A McKinney (2017), ”Brain Tumours: Incidence, Survival, and Aetiology”, Neurology in Practice, doi: 10.1136/jnnp.2004.040741
4. Rodriguez (2004), Principles of Magnetic Resonance Imaging, Revista Mexicana de Fisica, Vol. 50, No. 3, pp. 272-286.
5. Riddhi.S.Kapse , Dr. S.S. Salankar , Madhuri.Babar (2015), ”Literature Survey on Detection of Brain Tumor from MRI Images”, IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), Volume 10, Issue 1, PP 80-86.
6. Cover, T.M. (1968) "Rates of convergence for nearest neighbor procedures", In Proceedings of the Hawaii International Conference on System Sciences, Univ. Hawaii Press, Honolulu, 413–415.
7. Cover, T.M. & Hart, P.E. (1967) "Nearest neighbor pattern classification", IEEE Trans. Inf. Theory, 13: 21–27.
8. Dudani, S.A. (1976) "The distance-weighted k-nearest neighbor rule", IEEE Transactions on System, Man, and Cybernetics, 6: 325-327.
9. Guo, G., Wang, H., Bell, D., Bi, Y. & Greer, K., (2006) "Using KNN Model for Automatic Text Categorization", Soft Computing –A Fusion of Foundations, Methodologies and Applications 10(5): 423–430.
10. Gou, J., Du, L. Zhang, Y. & Xiong, T. (2012) "A New Distance-weighted k-nearest Neighbor Classifier", Journal of Information & Computational Science, 9(6): 1429-1436.