High Level Binary Feature Vector For Classification Of T1c+ Ring Enhancing Lesions

R. Anita Jasmine, P. Arockia Jansi Rani, Evangeline Ebenezer

Abstract: Human observations with limitations, quests for advance in techniques to assist automated diagnosis and further treatment. Predictive analytics encompasses several statistical data processing methods to interpret current data. This is based on the past facts serving as a in dispensable tool for radiology. Image processing tasks like segmentation, feature extraction, classification, and retrieval are used for computer aided diagnosis and treatment. The aim of this work is to analyze the soft tissue pattern in Ring Enhancing Lesion (REL) in MRI which is crucial for prediction whether the lesion is benign or malignant. A novel method to compute a High Level Binary Feature Vector (HLBFV) for heterogeneous, partial and homogeneous soft tissue classification is introduced to address the ambiguity of numerical feature values in pattern recognition. HLBFV is binary, comprehensible and supports very quick image retrieval. It is termed as high level feature as it is constructed from low level statistical features computed from Laws filter and GLCM features. After automatic segmentation, REL tissue is enhanced using selected Law’s masks for better discrimination of soft tissue pattern. The image is divided into non overlapping equal blocks for extracting several low level features to capture the intensity distribution, variation and correlation throughout the core of the lesion. The computed texture values are normalized to make it rotation invariant. From the low level values, the high level feature vector is constructed using the classification rules. Mean variance of intensity on different image blocks and experimentally verified threshold values on the dataset is used for classification. The performance of the classification is evaluated. The method achieves 100% accuracy for heterogeneous, 99% for partial and 90% for homogeneous lesions.

Index Terms: Feature vector, GLCM features, High Level Feature vector, Law’s filter, MRI, Ring Enhancing Lesion, Classification,T1C+.

1 INTRODUCTION

Ring Enhancing Lesion (REL) is an abnormal radiologic sign obtained using radio contrast medium to improve the visualization of anatomical internal structure. It is the most common radiological abnormality seen in young Indian patients with epilepsy [1]. REL are characterized by an area of decreased density surrounded by a bright rim from concentration of the enhancing dye. Contrast enhanced Magnetic Resonance Imaging (MRI) or CT images are used to diagnose REL. Basically there are eight different RELs. They are classified lesion pattern, completeness of the ring, smooth or irregular margins of the rim [2]. MRI images provide anatomical and physiological details in structure and function with 3D orientation, excellent soft tissues visualization and high spatial resolution [3]. Over the last two decades, MRI has emerged as the ideal modality for evaluating soft tissue lesions [4]. Advances in MRI technology improve the diagnostic accuracy of tumor, surgical planning and treatment. The role of MRI in differentiating benign and malignant tumor is based on signal intensity and pattern of tumor soft tissue. Usually, MRI contrast agents enhances the signal intensity of tumors like REL on T1-weighted spin-echo MR images, in some cases like FLAIR images, the demarcation between tumor and edema provides information on tumor vascularity [5][6]. The issue of distinguishing benign and malignant tumor can be successfully solved using MRI. Benign malignant differentiation in more than 90% cases can be predicted based on morphological features as suggested in [7]. The contrast enhancement pattern in tumor soft tissue is analyzed for tumor prediction.

2 RELATED WORKS

As radiomics is gaining attention in recent years, several image processing techniques are being proposed to improve the accuracy of computer aided diagnosis in medical imaging. Feature extraction plays the core role in image segmentation, classification and retrieval. Feature extraction techniques capture the intensity information, texture distribution and shape variation, using statistical methods, wavelet based methods and structural methods. Chung et al [10] proposed a method for less experienced radiologist to identify benign or malignant lesion classification using systematic combination of depth, size and signal intensity. They focused on assessing three different benign and three different malignant tumors. The work concludes significant difference in signal heterogeneity between benign and malignant lesions. Hauptfleisch et al [11] assessed lesions in MRI according to
signal intensity, size, and anatomical position. The lesion boundary was classified into two classes i.e. thick walled and thin walled cystic masses. Krasnoff et al [12] considered signal intensity, homogeneity and mass margin for diagnosis. In addition with shape, margin, and intensity features, pattern of enhancement was also used for tumor prediction. Wu et al [13] used three layers feed forward neural networks with back propagation algorithm for predicting benign or malignant breast tumor by processing tumor features observed from well experienced radiologists. Janki et al [14] extracted histogram and GLCM based texture features to mine associating rules. Decision tree algorithm was used for tumor classification with accuracy around 96%. Arakere et al [15] implemented as two steps CBIR approach to retrieve similar images. The first step classifies the tumor as benign or malignant and the second step retrieves the related images. Wavelets and modified fuzzy c-means algorithm was applied for segmentation. PCA is used for selecting texture and shape features. ANN is used for classification and KD tree indexing for retrieval. Buciu et al [16] manually segmented image patches of tumor under the guidance of radiologist to extract directional features using gabor wavelets and classified them using proximal SVM. Ain et al [17] proposed ensemble classification for benign or malignant brain tumor which involves image preprocessing using Fast Discrete Curvelet Transform. Fuzzy anisotropic diffusion is used for segmentation. SVM classification using texture features from GLCM and histogram achieves 99% accuracy. Sivakumar et al [18] proposed a continuous Particle Swarm Optimization, to select optimal features by evaluating the classification accuracy of the selected features using First order statistics, GLCM and GLRLM. There was a significant improvement in classification accuracy by 12% using PSO for the three different feature selection methods. Zhou [19] et al proposed Posterior Acoustic Shadowing using half contour features for breast tumor segmentation and tumor classification. Six different shape features were analyzed and two out of them i.e. tumor irregularity which was measured using Standard Deviation and tumor circularity was found useful for benign, malignant differentiation. By exploring the then available methods Dahshan et al [20] proposed a hybrid intelligent machine learning approach for brain tumor detection. After tumor segmentation using Pulse Coupled Neural Network, feature extraction using Discrete Wavelet Transform, feature reduction using Principal component Analysis and feed forward back propagation neural network is adapted for classification. Classification accuracy is 99% and the result is significantly good. Rodríguez et al [21] proposed statistical analysis of brain tissue images in the wavelet domain to estimate inter group differences called Wavelet Based Morphometry. The work extracts wavelet features and selects the salient features using Minimum Description Length algorithm and mapping them to spatial domain. Wibmer et al [22] proposed Haralick texture analysis for differentiating non-cancerous prostate lesions from cancerous prostate lesions by T2 weighted MRI. They used five texture descriptors viz. energy, entropy, correlation, homogeneity, inertia and proved significant variation in the texture features between cancerous and non-cancerous tissue. Parveen et al [23] developed an effective hybrid technique for brain tumor classification using fuzzy c-means and SVM. The MRI brain image is enhanced using mid range double stretch method. The skull is stripped using double thresholding and morphological operations for classification based on the GLRM statistical texture features. Wu et al [24] proposed a non-invasive technique for brain tumor diagnosis. Statistical features from lesion area are represented sparsely to eliminate redundant features and image patch dictionary to learn texture information is constructed by dictionary training. The method performs well with accuracy of 98% to differentiate PCNSL and GBM tumors. Probabilistic collaborative representation based classifier was proposed by Cai et al [25] for pattern classification. This was based on probability association of a feature with the different classes rather than conventional distance based classifiers. Results over different data sets, prove its performance over many classifiers including linear SVM. Havaei et al [26] proposed Convolutional Neural Network for high speed automatic segmentation of glioblastoma lesion by isolating the tumor from white matter, gray matter and CSF. Local and global features were extracted from image patch using two path CNN to generate feature map. The method achieved excellent results in segmenting, enhancing and non enhancing tumors, necrosis and edema. Salim et al [27] proposed a method for glioma lesion detection after segmenting and analyzing the tumor features. The tumor was segmented using Particle Swarm Optimization technique and fourier spectrum features were extracted for multi scale analysis using Hurst Exponents. SVM classifier was trained using positive and negative glioma samples. This achieves the accuracy rate around 99%. Agliozzo et al [28] presented Computer aided diagnosis of DCE MRI breast lesion images for distinguishing benign and malignant tumors. They used a quantitative combination of morphological, spatiotemporal and kinetic features, for classification using SVM. The method was evaluated for different lesion sizes achieving an accuracy of around 93%. Computer assisted tool for grading Glioma using MRI was presented by Kevin et al [29]. They used image histograms depicting the gray level distributions of lesion area for extracting statistical features. The accuracy of the results was around 87% and suggested that with advance improvements CAD could be a better teammate for assisting radiologists. Based on the above research work, an algorithm to classify tumor tissue is presented, comprising of segmentation, feature extraction and classification using a binary feature vector.
3. ENHANCEMENT PATTERN OF A MASS
Mass enhancement occurs in different patterns like homogeneous, heterogeneous, rim enhancement or with bright or dark internal septations. In this proposed method, three types of mass enhancement are considered, homogeneous, heterogeneous and partially homogeneous or heterogeneous. Homogeneous enhancement is of uniform intensity appearing confluent throughout the mass highly specifying benign tumor. Heterogeneous enhancement is non uniform and varies in intensity throughout the mass and is highly specific for malignant tumor. Partial enhancement is partially homogeneous and partially heterogeneous.

![Homogeneous enhancement](a) Heterogeneous enhancement (b) Partial enhancement enhancement (c) Partial enhancement

4. PROPOSED METHOD
The proposed method for HLBFV construction from low level features, is shown in the Fig.2. It comprises of three phases for feature enhancement, feature extraction and soft tissue prediction.

4.1 Texture Enhancement using Law's mask
Intensity variation over the soft tissue is analyzed for classifying for homogeneous, heterogeneous or partial enhancing lesions. The soft tissue texture of the REL is enhanced using Law's filter convolution kernels [30] and the resultant images are fused to get two images as shown in the Fig. 2.

Let f(i,j) denote the image pixel value at the pixel co-ordinates (i,j) and m1, m2, m3, m4 be the convolution kernels based on Law's filter. The filtered images are obtained as follows,

\[ g1 = \sum_{s=-2}^{2} \sum_{t=-2}^{2} m1(s, t) f(i+s, j+t) \]  
\[ g2 = \sum_{s=-2}^{2} \sum_{t=-2}^{2} m2(s, t) f(i+s, j+t) \]  
\[ g3 = \sum_{s=-2}^{2} \sum_{t=-2}^{2} m3(s, t) f(i+s, j+t) \]  
\[ g4 = \sum_{s=-2}^{2} \sum_{t=-2}^{2} m4(s, t) f(i+s, j+t) \]

where M1, M2, M3 and M4 are the convolution kernels based on Law’s filter given by,

\[
m1=\begin{bmatrix}
-1 & -2 & 0 & 1 \\
-4 & -8 & 0 & 4 \\
-6 & -12 & 0 & 6 \\
-4 & -8 & 0 & 4 \\
-1 & -2 & 0 & 1
\end{bmatrix}
\]

\[
m2=\begin{bmatrix}
-1 & -4 & -6 & -4 & 1 \\
-2 & -8 & -12 & -8 & 2 \\
0 & 0 & 0 & 0 & 0 \\
2 & 8 & 12 & 8 & 2 \\
1 & 4 & 6 & 4 & 1
\end{bmatrix}
\]

\[
m3=\begin{bmatrix}
-1 & 0 & 2 & 0 & -1 \\
-4 & 0 & 8 & 0 & -4 \\
-6 & 0 & 12 & 0 & -6 \\
-4 & 0 & 8 & 0 & -4 \\
-1 & 0 & 2 & 0 & -1
\end{bmatrix}
\]

\[
m4=\begin{bmatrix}
-1 & -4 & -6 & -4 & 1 \\
0 & 0 & 0 & 0 & 0 \\
2 & 8 & 12 & 8 & 2 \\
0 & 0 & 0 & 0 & 0 \\
-1 & -4 & -6 & -4 & 1
\end{bmatrix}
\]

The convoluted image matrices are fused to get Y and Z using arithmetic mean values given by the equations

\[
Y = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} g1 2gij
\]

\[
Z = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} g3 3gij
\]

4.2. GLCM Feature computation
The success of classification and retrieval depends on feature extraction. To assess the variation of image intensity over the image, the original image matrix X, the fused matrices Y and Z are divided into n non overlapping image blocks of equal size. The original image matrix X, the fused matrices Y and Z are divided into n non overlapping image blocks of equal size. Where n=16 and i=1, 2, 3.

GLCM, the statistical texture analysis for examining texture based on spatial association between pixels is used for computing after Xi normalization. The feature vector comprising of Contrast, Correlation, Energy and Homogeneity are computed for all the individual blocks to create three GLCM matrices each of size 16X4.

\[
F = \begin{bmatrix}
f_{1} & f_{1} & \cdots & f_{1} & 4 \\
\vdots & \ddots & \vdots & \vdots & \ddots \\
F_{1} & F_{1} & \cdots & F_{1} & 4
\end{bmatrix}
\]

The variance vector of Contrast, Correlation, Energy, Homogeneity for all the 16 blocks of X, Y and Z are calculated

\[
\mu_{k} = \frac{1}{n} \sum_{j=1}^{2} \sum_{k=1}^{2} F_{j} k
\]

The mean of variance is calculated by

\[
\sigma_{k}^{2} = \frac{1}{n} \sum_{k=1}^{2} \mu_{k}
\]

4.3 Classification Rules
Let F1, F2, and F3 be the three high level features initialized to zero represents heterogeneous, partial and homogeneous lesions. The high level feature vectors are computed based on the μ values and threshold th (th=0.006). The value of th is determined experimentally by analyzing all the dataset images. The pattern is heterogeneous if either of three μ values are greater than the threshold th. A non-homogeneous tissue with μ1 and μ 3 greater than th can be partial or homogeneous. The tissue is partially homogeneous if μ2<(th-0.002) and homogeneous if μ2>(th-0.002). The HLBFV feature vector is given by 

ft=[ F1, F2, F3 ]

4.4 Dataset
In this study, the 115 T1C+ MRI dataset for ring enhancing lesion is used. The dataset is obtained from Radiopaedia, S.P Scans and Pranav Scans Nagercoil. Since the focus is on soft
tissue classification, the MRI is segmented using automatic region growing algorithm [31]. The segmented soft tissue in the dataset consists of three classes heterogeneous (52), partial (31) and homogeneous (32).

5. RESULTS AND DISCUSSION

The proposed HLBV feature extraction method is simulated using Matlab R2015a version with intel dual core i3-370M with 2.4 GHz and 4GB internal RAM for three different RELs. The confusion matrix is computed to evaluate the classification performance. It depicts the accuracy and correctness of the classification model. Table I illustrates the three possible feature vectors for the three different classes. Table II shows the confusion matrix of the proposed method. The method is experimentally verified using 81 samples for training, 17 samples for validation and 17 samples for testing phases. For example, in heterogeneous class, TP is the number of heterogeneous instances that are classified as heterogeneous. TN is the number of non-heterogeneous instances that are not classified as heterogeneous. FP is the number of non-heterogeneous instances that are classified as heterogeneous. FN is the number of heterogeneous instances that are not classified as heterogeneous. Similarly TP, TN, FP and FN are calculated for the other two classes. Performance Evaluation is done using seven standard benchmark measures like accuracy, prediction, recall, sensitivity, specificity, positive predicted and negative predicted values and the results are presented in Table III. Classification accuracy is directly proportional to the seven performance measures. The results for heterogeneous class are 100% for all the evaluations which proves the efficacy to discriminate and identify heterogeneous lesions. Precision, specificity, positive predicted value is 100% for partial classes as it can strongly identify partial tissue class. In case of homogeneous class, accuracy, prediction, specificity, positive predicted and negative predicted values are high but recall and sensitivity goes down due to 30% misclassification. Receiver Operating Characteristic (ROC) curve is the plot of TP rate against the FP rate for different cut-off points. The closer the curve is to the upper left border indicates the accuracy of the method. Fig. 3 illustrates the plot for the three classes, training, testing, validation and the entire dataset. Visual analysis depicts the closeness of the curves towards the upper left corner for the heterogeneous and partial classes and proves the high performance for the first two classes when compared to the third class.

6. CONCLUSION

The volume of medical images is ever increasing and its role is vital to improve the health care and treatment. Medical Image Processing is a boon for radiologists to visualize, analyze and interpret medical images. Feature extraction plays a main role in segmentation, classification, pattern recognition and retrieval. In this work, a binary comprehensible feature vector is proposed for characterizing the soft tissue in RELs using MRI. This new concept for feature vector is introduced to solve the ambiguity of low level features. Once computed it is easy to store, match and retrieve related samples from the feature database of heavy volume. This feature vector supports very quick image retrieval. The system can serve as an efficient diagnostic tool for evaluating the ring enhancing lesions. In future, the pathological report of the lesion can be linked to improve the tumor assessment. The experimental results reveal that the method works excellent in differentiating heterogeneous tumor tissue. The result can be further improved by increasing the sample dataset.

REFERENCES


unpublished. (Unpublished manuscript)


