Feature Extraction of Arterio-Venous Malformation Images using Grey Level Co-Occurrence Matrix

M. Ashwin Kumaar* and Palani Thanaraj

Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, Chennai – 600119, Tamil Nadu, India; ashwinbiomedic94@gmail.com, palanithanaraj.k@gmail.com

Abstract

Objective: Arterio-Venous Malformation is, mostly, as a result of incidental findings. This is a pioneering effort to incorporate advanced image segmentation techniques in order to improve its diagnosis. Method: Feature extraction of Arterio-Venous Malformation (AVM) brain images is attempted for automatic AVM recognition system using OTSU based Particle Swarm Optimisation (PSO) algorithm. Initially the input Magnetic Resonance Image (MRI) is segmented using PSO method, based on multiple threshold processes. The image features are then extracted from the partitioned regions using Gray Level Co-occurrence Matrix (GLCM) technique. The features obtained from AVM and normal brain images are compared using statistical measures. Findings: Our analysis suggests that the extracted GLCM features of AVM brain MRI images shows significant variation to normal brain MRI images. Out of the total 22 features extracted, 18 features shows a lesser feature extraction value for AVM affected brain image compared to the normal brain image. The other 4 features extracted shows a higher value for AVM affected brain image compared to the normal brain image. This pattern is the same for any AVM affected brain, in comparison to a normal brain. Applications: This work helps in the development of automatic recognition system for AVM, so that many cases can be identified in the preliminary stages and suitably treated.

Keywords: AVM, Feature Extraction, GLCM, MRI, PSO

1. Introduction

Automatic recognition system is the most challenging aspect of medical image processing. AVM is an abnormal connection between the arteries and the veins. High-pressure blood from the arteries flow into the low-pressure veins directly without capillaries and it prevents the consequent dampening effect. The functions of the capillaries such as taking up of oxygenated blood from the arteries to provide nutrients to the surrounding cells and feeding the de-oxygenated blood from the rest of the body to the veins are not performed.

As a result, blood vessels form a Nidus (meaning ‘nest’ in Latin). This abnormality can occur in Brain, Spinal cord, Lungs, Kidney, Liver, Spleen, Spermatic cord, Iris, Inter Coastal space, Arm and Shoulder. It occurs in the central nervous system. It is not hereditary, but it can also occur due to autosomal dominant disease. Its symptoms include headache, epilepsy, vertigo, difficulties in movement, coordination and speech, seizures and haemorrhage. They can cause intense pain, bleeding and serious medical symptoms. They are mostly incidental findings. Diagnosis is usually done through Computerised Tomography (CT) scan and Magnetic Resonance Imaging (MRI) scan. Automatic Recognition system is developed for AVM in this proposed work. The images are segmented using PSO algorithm. From the segmented image, texture features are extracted using GLCM. It provides the information about the structural arrangement of surfaces and their relationship with the surrounding environment. The above procedure is repeated for 256x256 MRI normal brain images. The results are compared to develop the automatic recognition system for AVM.

Much of research work has been carried out relating to this field. An automatic recognition system can reduce the burden of biopsy for a radiologist. R. Nithya et al.¹

*Author for correspondence
have developed a computer aided diagnosis system based on neural network. They proposed a feature selection method called Maximum Difference Feature Selection, which improved classification accuracy. David A. Clausi proposed a study on the effect of grey level quantization on the ability of co-occurrence probability statistics to classify natural textures like SAR-sea ice images. Nitish Zulpe et al. have extracted texture features from four classes of brain tumour using GLCM and fed it into Feed Forward Neural Network with Levenberg Marquart (LM) non-linear optimization algorithm which gives better recognition rate of 97.5%. Imad Zyout et al. have facilitated the diagnosis of Micro-Calcification (MC) clusters by introducing a new heuristic feature selection based on Particle Swarm Optimisation and KNN classifier (PSO-KNN) which was highly effective. Robert M. Haralick et al. have presented some easily computable texture features of images and demonstrated their application in category identification tasks of three different kinds of image data. Leen Kiat Soh et al. have defined the values of quantization, displacement and orientation that are best for SAR sea ice texture analysis using GLCM. A. Gebejes et al. have analysed the dependency of the features with the displacement considered for their computation and explored the possibility of features invariant under changes of the distance between the sample and observation position. Sasirekha N et al. have proposed a robust approach to estimate and remove the Richian noise of the 2D MRI images to improve their segmentation accuracy and for the detection of tumours.

2. Particle Swarm Optimisation

PSO was conceived by Kennedy and Eberhart in the year 1995. It is an evolutionary technique, developed from the notion of the social activities in birds flocking and fishes schooling. At each step of the optimization, every particle evaluates its fitness and the neighboring particles fitness. It is used to find the global optimum solution in a complex search space. It is less dependent of a set of initial points than other optimization technique. It is a derivative-free algorithm. It has fewer parameters to adjust. The results produced are fast and cheap. The particles from the swarm, work together for finding an optimal solution to the parameter selection problem. It randomly initializes the particles in the parameter space. During the optimization process, the PSO algorithm stores the locations of the best fitness that has been achieved by each particle and the global best fitness achieved by all particles in the swarm. The information is then used to update the movement and the position of the particles in the parameter space.

The equation of velocity update and equation of position update in PSO algorithm are represented as;

\[ V_t(t+1) = W V_t + C_1 R_1 (P_t^i - X_t^i) + C_2 R_2 (G_t^i - X_t^i) \]  
\[ X_t(t+1) = X_t^i + V_t(t+1) \]

Where \( W \) is inertia weight assigned as 0.75, \( V_t^i \) is the current velocity of particle, \( V_t(t+1) \)-updated velocity of particle, \( X_t^i \)-current position of particle, \( X_t(t+1) \)-updated position of particle, \( R_1, R_2 \) are the random numbers \([0,1]\) and \( C_1 \) and \( C_2 = 2.1 \).

3. Gray Level Co-Occurrence Matrix

The Gray-Level Co-Occurrence Matrix (GLCM) method is used to extract second order texture features. A GLCM is created, which depends on, how the pixels pairs with the corresponding values, which are in a specified spatial relationship in an image. The function called gray comatrix in MATLAB creates a Gray-Level Co-Occurrence Matrix (GLCM) by calculating how often a pixel with the intensity (graylevel) value \( j \) occurs in a specific spatial relationship to a pixel with the value \( i \). In a GLCM, the number of rows and columns is equal to the number of gray levels, \( G \), in the image. The size of GLCM depends on the number of gray levels. GLCM is calculated in four angles \( (0^\circ, 45^\circ, 90^\circ, 135^\circ) \).

4. Feature Extraction

The number of different gray levels in an image is \( G \). GLCM is represented by \( P \), \( \mu \) is the mean value of \( P \) and \( \sigma \) are the means of \( P_x \) and \( P_y \). \( \sigma_x \) and \( \sigma_y \) are standard deviations of \( P_x \) and \( P_y \). \( P_x(i, j) \) is the \( i \)th entry in the matrix obtained by summing the rows of \( P \) (i, j).

\[ P_x(i) = \sum_{j=0}^{G-1} P(i, j) \]  
\[ P_y(j) = \sum_{i=0}^{G-1} P(i, j) \]  
\[ \mu = \sum_{i,j=0}^{G-1} iP(i, j) \]
\[ \mu_x = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) = \sum_{i=0}^{G-1} i P_x(i) \]  
\[ \mu_y = \sum_{j=0}^{G-1} \sum_{i=0}^{G-1} P(i, j) = \sum_{j=0}^{G-1} j P_y(j) \]  
\[ \sigma_x^2 = \sum_{i=0}^{G-1} (i - \mu_x)^2 \sum_{j=0}^{G-1} P(i, j) = \sum_{i=0}^{G-1} [P_x(i) - \mu_x]^2 \]  
\[ \sigma_y^2 = \sum_{j=0}^{G-1} (j - \mu_y)^2 \sum_{i=0}^{G-1} P(i, j) = \sum_{j=0}^{G-1} [P_y(j) - \mu_y]^2 \]  
\[ P_{x+y}(k) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \text{ for } k = 0, 1, \ldots, 2(G-1) \]  
\[ HXY = -\sum_{i,j=0}^{G-1} P(i, j) \log_2 P(i, j) \]  
\[ HXY1 = -\sum_{i,j=0}^{G-1} P(i, j) \log_2 \left( P_x(i) P_y(j) \right) \]  
\[ HXY2 = -\sum_{i,j=0}^{G-1} P_x(i) P_y(j) \log_2 \left( P_x(i) P_y(j) \right) \]  

Expressions for GLCM features are as follows:

1) **Auto-Correlation:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left( P_x(i) - \mu_x \right) \left( P_y(j) - \mu_y \right) / \sigma_x \sigma_y \]  
It refers to correlation of a time series with its own past and future values.

2) **Contrast:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) (i - j)^2 \]  
It is the local intensity variation in the image. It is the difference in color that makes the object distinguishable.

3) **Correlation:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \times (i \times j) - \mu_x \times \mu_y / \sigma_x \sigma_y \]  
It is a measure of gray level linear dependence between the pixels at the specified positions about each other.

4) **Cluster Prominence:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)(i + j - \mu_x - \mu_y)^4 \]  

5) **Cluster Shade:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} ZP(i, j)(i + j - \mu_x - \mu_y)^3 \]  

6) **Dissimilarity:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i - j| P(i, j) \]  

7) **Energy:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)^2 \]  
It is a measure of the amount of disorder in the system.

8) **Entropy:**
\[ = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \log(P(i, j)) \]  

9) **Homogeneity:**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) / (1 + |i - j|) \]  

10) **Maximum Probability:**
\[ = \max (i, j) P(i, j) \]  

11) **Sum of Squares: Variance**
\[ = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)(i - \mu)^2 \]  

12) **Sum Average (sa):**
\[ = \sum_{i=0}^{2G-2} i P_{x+y}(i) \]  

13) **Sum Variance:**
\[ = \sum_{i=0}^{2G-2} (i - sa)^2 P_{x+y}(i) \]  

14) **Sum Entropy:**
\[ = -\sum_{i=0}^{2G-2} P_{x+y}(i) \log(P_{x+y}(i)) \]  

15) **Difference Entropy:**
\[ = \sum_{i=0}^{G-1} P_{x+y}(i) \log(P_{x+y}(i)) \]
16) **Information Measure of Correlation:**
\[
H_{XY} - H_{XY1} = \frac{H_{X,Y}}{\max(H_X,H_Y)}
\]  
(29)

17) **Information Measure of Correlation 2:**
\[
\sqrt{1 - \exp[-2.0(H_{XY2} - H_{XY})]}
\]  
(30)

18) **Inverse Difference Moment Normalized:**
\[
\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{P(i,j)}{1+|i-j|^2}
\]  
(31)

**Thresholded Images:**

![Thresholded Images](image)

**Figure 1.** (a) AVM Brain Threshold. (b) Normal Brain Threshold.

| S.NO | GLCM FEATURES | AVM BRAIN THRESHOLD VALUE | NORMAL BRAIN THRESHOLD VALUE |
|------|---------------|--------------------------|-----------------------------|
| 1    | AUTO CORRELATION | 5.1093 | 7.6979 |
| 2    | CONTRAST | 0.1381 | 0.2347 |
| 3    | CORRELATION (mat lab) | 0.9529 | 0.9537 |
| 4    | CORRELATION | 0.9529 | 0.9537 |
| 5    | CLUSTER PROMINENCE | 152.3126 | 343.2286 |
| 6    | CLUSTER SHADE | 17.6990 | 31.0359 |
| 7    | DISSIMILARITY | 0.1100 | 0.1613 |
| 8    | ENERGY | 0.3581 | 0.3201 |
| 9    | ENTROPY | 1.5273 | 1.7639 |
| 10   | HOMOGENEITY (mat lab) | 0.9493 | 0.9300 |
| 11   | HOMOGENEITY | 0.9478 | 0.9266 |
| 12   | MAXIMUM PROBABLITY | 0.5385 | 0.5170 |
| 13   | SUM OF SQUARES: VARIANCE | 5.1184 | 7.7437 |
| 14   | SUM AVERAGE | 3.8540 | 4.5950 |
| 15   | SUM VARIANCE | 11.6466 | 18.9225 |
| 16   | SUM ENTROPY | 1.4200 | 1.5933 |
| 17   | DIFFERENCE VARIANCE | 0.1381 | 0.2347 |
| 18   | DIFFERENCE ENTROPY | 0.3585 | 0.4567 |
| 19   | INFORMATION MEASURE OF CORRELATION 1 | -0.7246 | -0.6859 |
| 20   | INFORMATION MEASURE OF CORRELATION 2 | 0.9076 | 0.9173 |
| 21   | INVERSE DIFFERENCE NORMALIZED | 0.9881 | 0.9829 |
| 22   | INVERSE DIFFERENCE MOMENT NORMALIZED | 0.9979 | 0.9965 |
5. Result and Discussion

Here GLCM feature extraction procedure is followed for both a normal brain image and an AVM affected brain image (Figures a, b). The various feature values obtained are tabulated and compared. We can see that the values obtained for the AVM brain image is, in general, a lower value compared to that of the normal brain. However, the exception cases include the values of energy, homogeneity, inverse difference moment normalized and maximum probability, out of the total 22 features extracted. This trend is common for all the AVM affected brain images and henceforth, it can be a useful tool in its diagnosis.

6. Conclusion and Future Scope

This work has many applications, especially in Medical Imaging field. This work can be used to develop an Automatic Recognition System for AVM, as we can distinguish between AVM brain and a normal brain using their GLCM features obtained. The future scope includes the development of AVM grading that is helpful to a doctor to assess its severity and to prescribe drugs for this cause suitably. This work is a pioneering effort to apply Image Processing and GLCM feature extraction in the diagnosis of AVM. As this AVM can occur not just in Brain but in many parts of the body, as mentioned in the introduction, the same procedure can be applied, with the corresponding MRI data.

7. Acknowledgement

I would like to thank Bharath Scans, Royapettah, Chennai-600114, for providing us with the necessary MRI images including the AVM MRI scan results, with patient’s consent and approval. This work would not have been possible without this source and we are deeply indebted.

8. References

1. Nithya R, Santhi B. Mammmogram classification using maximum difference feature selection method. Journal of Theoretical and Applied Information Technology. 2011 Nov; 33(2):197–204.
2. Clausi DA. An analysis of co-occurrence texture statistics as a function of gray level quantization. Can J Remote Sensing. 2002; 28(1):45–62.
3. Zulpe N, Pawar V. GLCM texture features for brain tumour classification. International Journal of Computer Science Issues. 2012 May; 9(3(3)):354–9.
4. Zyout I, Qader IA, Jacobs C. Classification of MC clusters in digital mammography via haralick descriptors and heuristic embedded feature selection method. Ubiquitous Computing and Communication Journal. 2011 Oct; 31(2):34–9.
5. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics. 1973 Nov; SMC-3(6): 610–21.
6. Soh LK, Tsatsoulis C. Texture analysis of SAR-Sea ice imagery using gray level co-occurrence matrices. IEEE Transactions on Geoscience and Remote Sensing. 1999 Mar; 37(2):780–95.
7. Gebejes A, Huertas R. Texture characterization based on gray level co-occurrence matrix. Conference of Informatics and Management Sciences. 2013 Mar; 2(1):375–8.
8. Kennedy J, Eberhart R. Particle swarm optimisation. Neural Networks. IEEE International Conference. 1995 Nov-Dec; 4(1):1942–8.
9. Sasirekha N, Kashwan KR. Improved segmentation of MRI brain images by denoising and contrast enhancement. Indian Journal of Science and Technology. 2015 Sep; 8(22).