Detection and Classification of Brain Lesion Depending on Statistical Features Textural Analysis

Alaa N. Mazher¹, Maysaa. R.Naeemah², Alyaa H.Ali²
¹Computer Science Department, University of Technology, Baghdad, Iraq.
²Department of Physics, College of Science for Women, University of Baghdad, Baghdad,

Abstract. The early detection of brain lesion which includes the stroke (Hemorrhage and Ischemic) and cancer helps the doctors to overcome the health problem in the future. The correct diagnose may save many people from death. the medical image processing including the supervised classification "support vector machine", "the gray level neighbors matrix", "Fisher’s Discrimination Ratio(FDR)" and the Accuracy matrix to detect and classify the brain lesion, twenty two computed tomography (CT) brain images have been taken with size 512×512. This methods provides very good diagnosis of the brain lesion, the accuracy for cancer detection is 98%, for the Hemorrhage is 96%, for the ischemic 99% and normal 90%.

Keywords Brain lesion, Support Vector Machine, Confusion matrix, Gray level neighbors matrix, Fisher’s Discrimination Ratio.

1. Introduction
The brain lesion is one of most causes of death; the early diagnosis is essentially to detect the type of lesion especially the stroke. If the stroke is detected early in the first three hours, it will save the life of the patients when the stroke is in its initial stage [1]. The use of computer system in the detection and analysis of medical image CT is very useful to the radiologists. The application of Computerized Tomography is noninvasive and low radioactive, this made the use of CT in medical very important [2]. The image processing textural analysis is used with the CT image to give the best result which can be obtained to diagnose the brain lesion. The support vector machine which is supervised classification with the textural statistical analysis (“Gray level neighbors matrix”) and more statistic features helps to give the accurate detection of the brain cancer and stroke. In 2014 Abraham Chandy et al proposed a statistical textural matrix it depend on the occurrence of neighbors gray level pixels [3].

2. Methodology
The search methodology based on several stages, the first one is collecting the data images which are (22) Brain lesion CT images of size 512×512, figure(1) shows the test images. The images have been converted to gray scale images and treated to classify them. The support vector machine is used to classify the images into cancer, Hemorrhage and Ischemic stroke, the gray level matrix has been used as a textural analysis matrix in which four features are calculated and used as a decision features. The confusion matrix is used to predicate the accuracy of the classification.
3. Radial Basis Support Vector Machine

The support vector machine can be regarded as a supervised classifier, it’s a binary technique, its operation depend on defining the maximum separation hyper plane two classes of cases, it create the largest distance between the hyper plane separation and cases on each side of it by maximize the edges\[4,5\]. The algorithm can be summarized as follows[6].

The right hyper plane is finding to maximize the margin and minimizes the miss classification

\[ \sum_{i=1}^{L} X_i Y_i \propto_i \]

where \(X_i\) and \(Y_i\) are point of hyper plane \(X,Y\) coordinate and \(\propto_i\) is the vector parameter, \(L\) is the number of training cases.

The number of binary classification can be finding as follows

\[ X.W+b=0 \]

Let us consider a supervised binary classification problem. The training data can be representing as \([X_i, Y_i]\), where the value of \(i=1,2,...,N\), \(Y_i\) has the value in the rang (1,-1) and \(N\) is the number of training cases, the (+1) is for \(W_1\) and the (-1) for class \(W_2\), linear separation are consider for the two classes. the hyper plane which separated the classes [7].

Is \( F(x)=W.X_i +W_0 =0 \)

the value of \(W\) and \(W_0\) is obtained in a manner that \(Y_i (W.X_i +W_0) \geq +1\), class \(W_1\) for which \(Y_i\) is +1 and \(Y_i (W.X_i +W_0) \leq -1\) class \(W_2\) in which \(Y_i\) is -1 these can be combine together to obtined the following equation [14].

\[ Y_i (W.X_i +W_0)-1 \geq 0 \]

The super vector mechanism aims to find the suitable hyper plane that have the best merge between the classes, the super vector machine are on the hyper plane which are parallel it gives by [7].

\[ W.X_i +W_0=-1, +1 \]

the hyper plane can be represented by solving the following equation [7].

Minimize \( \frac{1}{2} ||w||^2 \)

subject to

\[ Y_i (W.X_i +W_0)-1 \geq 0 \]

\(i=0,1,...,N\).

the accuracy can be measured as follows

\[ accuracy = \frac{T_P+T_N}{Totalnumberofclasses} \times 100\% \]

\(T_P\) is the true positive class which detect the tumor type ("stroke and cancer")

Figure 1. Shows the tests images.

Normal

Ischemic

Cancer

Hemorrhage
$T_N$ is the negative class
To evaluate the proposed method the false positive $F_p$ must be obtained [1,8].

**Neighbor's matrix**: the gray level neighbors matrix idea based on extracting textural feature basing on the gray level occurrence of gray level of neighbors with special neighbor. The procedure of this matrix is the central pixel in which the elements is $(i,j)$ [3].

$$F(i,j)=\text{No.}\{(x,y)|(x,y)\}=i,\text{ No.}\{(p,q)|N_{xy}(p,q)=i\}=j$$

No. is the number of elements in $N_{xy}(p,q)$ neighborhood in the given image the images has been quantized to 8 gray level, four features can be calculated from this matrix

**Feature1** = $\sum_{i=1}^{N} \sum_{j=1}^{N} F(i,j)^2$  
the N is the number of gray level which is 255.

$$\text{Feature2} = \sum_{i=0}^{N-1} n^2 \sum_{j=0}^{n} F(i,j)$$

$$\text{Feature3} = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{F(i,j)}{1 + (i - j)^2}$$

$$\text{Feature4} = \sum_{i=1}^{N} \sum_{j=1}^{N} (\min_{j} - \mu)^2 F(i,j)$$

$$\mu = \frac{1}{N \times N} \sum_{i=1}^{N} \sum_{j=1}^{N} F(i,j)$$

4. Discriminate Ration calculation
The discrimination ratio based on calculating the mean and variance of two classes, this ratio is called "Fisher Discriminate Ration (FDR)"[8].

$$FDR = \frac{(\text{mean}1 - \text{mean}2)^2}{\text{variance}1^2 - \text{variance}2^2}$$

mean1, mean2, variance1 and variance2 are the mean and variance of class1 and class2. The FDR is high the distance b between classes is large and low with in the class variance i-e the result is better, the FDR ratio is higher [1].

| Table 1. Shows the textural features for the stroke (Ischemic). |
|----------------------|-----------------|----------------|
| features           | mean       | variance   | FDR     |
| Feature1           | 0.191      | 0.00385    | 0.987   |
| Feature2           | 1.074      | 0.0108     | 0.876   |
| Feature3           | 1.985      | 370.25     | 0.672   |
| Feature4           | 0.917      | 0.0056     | 0.765   |

| Table 2. The textural features for the stroke (Hemorrhage). |
|----------------------|-----------------|----------------|
| features           | mean       | variance   | FDR     |
| Feature1           | 0.232      | 0.083      | 0.985   |
| Feature2           | 1.989      | 325.26     | 0.916   |
| Feature3           | 2.292      | 8.674      | 1.254   |
| Feature4           | 1.987      | 3.876      | 0.547   |

| Table 3. The textural features for the Cancer. |
|----------------------|-----------------|----------------|
| features           | Mean       | variance   | FDR     |
| Feature1           | 0.331      | 0.213      | 1.821   |
Table 4. The features for the Normal case.

| features | mean  | variance | FD R  |
|----------|-------|----------|-------|
| Feature1 | 0.100 | 0.213    | 0.241 |
| Feature2 | 0.138 | 3.712    | 0.312 |
| Feature3 | 0.201 | 1.082    | 0.120 |
| Feature4 | 0.103 | 5.101    | 0.202 |

Table 5. Represent the Confusion matrix.

| Cancer    | Hemorrhage | Ischemic | Normal | case       |
|-----------|------------|----------|--------|------------|
| 5         | 0          | 0        | 3      | Cancer     |
| 1         | 5          | 0        | 0      | Hemorrhage |
| 0         | 0          | 6        | 0      | ischemic   |
| 1         | 0          | 0        | 6      | normal     |

Table 6. Represented the $T_P$, $T_N$ and the accuracy.

|          | $T_P$ | $T_N$ | accuracy |
|----------|-------|-------|----------|
| Cancer   | 0.978 | 0     | 98%      |
| Hemorrhage| 0.963 | 0.003 | 96%      |
| Ischemic | 0.952 | 0.045 | 99%      |
| Normal   | 0.899 | 0.001 | 90%      |

5. Discussion

The gray level neighbors matrix and the support vector machine can be used to classify and detect the brain lesion, the 22 CT images used as a test images six is classified as normal, six as ischemic, five as hemorrhage and five as cancer cases, table (1,2,3,4). Shows the mean, variance and Fisher discriminate ration for the four cases, when the value of fisher discriminate ration between the selected features is high, good the detection is. Table 5. Shows the confusion matrix, the diagonal value represent the classify cases, as the diagonal value have higher value than the other elements the classification is better. Table 6 represent the accuracy value, the best detect one is the ischemic case, it gives about 99%, then the cancer case which is 98%, the hemorrhage 96% and the normal case is 90%.

6. Conclusion

It is a good idea to use the textural features from the Neighbors matrix with the "SVM" classifier to obtain a good detection and classification of the brain lesion, the higher value of the fisher discriminate ration the large distance between the classes and the variance value with in the class is small, this features amalgamation are chosen, it gives potential use of the classifier, the confusion matrix helps in given the best performance of the cases, as our used test images are 22 it classified as four cases and the accuracy percentage result shows that the hybrid technique used is good in detecting and classifying the brain lesion. The suggested system much the medical report decision which belong to each selected test images.

Reference

[1] P. Bhat, M. Singh," Classification of Stroke using texture Analysis on CT images". (IJSCE). 2(3): 527-530, 2012.
[2] H. A. Elnemr, "Statistical Analysis of Law’s Mask Texture Features for Cancer and Water Lung Detection", 10(6):196-202, 2013.

[3] S. N. Mazhir, A. H. Ali, A. H; Abdalameer, N. K and Hadi, F. W. Studying the effect of Cold Plasma on the Blood Using Digital Image Processing and Images Texture analysis, International conference on Signal Processing, Communication, Power and Embedded System (SCOPES), IEEE Xplore Digital Library, 904-914 (2016).

[4] S. N. Mazhir, F. W. Hadi, A. N. Mazher, and L. H. Alobaidy, Texture Analysis of smear of Leukemia Blood Cells after Exposing to Cold Plasma. Baghdad Science Journal, 14(2):403-410 (2017).

[5] J. K. Annavarapu. "Statistical Feature Selection for Image Texture Analysis". 2 (5): 546-550, 2015.

[6] A. H. Ali, N. H. Ghaeb, Z. M. Musa. "Support Vector Machine for Keratoconus Detection by Using Topographic Maps with the Help of Image Processing Techniques" IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS). 12( 6) VI : 50-58, 2017.

[7] P. Smitha, L Shaji, G.M. Mini. "A Review of Medical Image Classification Techniques." International Conference on VLSI, Communication and Instrumentation, International Journal of Computer Application. 34-48, 2011.

[8] A. Ben-hur, J. Weston. "A User Guide to support Vector Machines Preliminaries: Linear Classifiers", In Methods in Molecular Biology; Humana Press, 609:223-239, 2010.

[9] S. B. Kotsiantis, I.D.Zaharakise, P. E. Pintelas. "Machine Learning: Areview of Classification and Combining Techniques". Spring Sci. Media 26, 159-190, 2007.

[10] A. M. VibhaVyas, and P. Rege. "Inverse Fisher discriminant Ratio based Training Set Selection for Optimal Classification Accuracy". 12(20):9535-9541, 2017.