Classification of Brain Lesion using K-Nearest Neighbor technique and Texture Analysis

Sara Abd Al-hadi¹, Alyaa H.Ali², Maysaa Raba Naeemah³, Alaa Noori Mazher ⁴
¹,²,³Department of Physics/ College of science for women/ University of Baghdad/Iraq.
⁴Computer Department / University of Technology, Iraq
*Correspondence: aliahusain@ymail.com.

Abstract: Texture is an important property for analyzing many types of images. It provides a rich source of information about the image. In this paper, four statistical features (contrast, correlation, homogeneity and energy) are calculated from the gray level co-occurrence matrix in which the gray level of variation associated with gray levels of three samples from CT scan Images to patients with (hemorrhage, ischemic and cancer) in the brain, the region of interest has been obtained by classifying the images using the k-nearest neighbor (K-NN) five classes has been proposed for the classification purpose, the result shows two classes have infection region and the three other classes are healthy region. It has found that the cancer texture are highly correlated comparing with the stroke. The stroke texture is more homogenous than the cancer texture this make the distinguish between them more specific easy.

Keyword: Brain stroke, brain cancer, co-occurrence matrix, K-NN

1. Introduction

The medical image processing is one of the most important technique which is used to detect the lesion in the brain [1]. This work is based on detecting the cancer and stroke in its two cases (Hemorrhagic stroke, ischemic stroke ).The techniques which is used to detect the brain lesion depending on the statistical features and classification techniques, using digital image in medicine is very important because its helps in the correct diagnosis of the Brain Lesion, this is due to the computed tomography image having the ability to distinguish between any abnormal changes in the brain tissue. Early detection and correct treatment based on accurate diagnosis which are important steps to improve disease outcomes [1]. Image analysis techniques have played an important role in many medical applications. In general, applications automatically extract features from the image that is then used for a variety of classification tasks, such as distinguishing natural tissue from abnormal tissue [1],[ 2]. Among the traditional application areas for tissue analysis are biomedical image analysis. The tissue area can be distinguished in a non-uniform or spatial distribution of density, reflecting a difference in intensity in the scene [3]. The texture feature is a complex visual pattern consisting of entities or sub-patterns with specific properties (brightness, gradient, size) [4]. The textural features contain important information about the structural arrangement of surfaces and their relationship to the surrounding environment [5]. Simple texture features can be achieved by calculating statistical properties, such as the mean and contrast of the graph of the gray level of the image. However, the performance of these types of first-class statistics is usually weak. Haralik et al. (1973) [6] provide the Statistics of the second-grade scale using GLCM and defined fourteen statistical parameters for the tissue. Synchronous matrices give information about texture decoration, and are used to calculate structural properties. These features are sensitive to illumination, but they are very common in various texture analysis applications [7]. The supervised classification technique Nearest Neighbors is used to classify the brain infection which are the hemorrhage and ischemic stroke and the cancer case The nearest neighbor K, or K-NN, is identifies an object or test patterns for a class that relies on the majority of its closest K neighbors in the feature space. It uses training samples with category classifications, but does not use any form to match training data and does not explicitly specify row boundaries. The algorithm is very simple and depends on three operations: First, the distance between each training sample and the test sample is calculated.
Second, these distances are sorted to determine the nearest K training samples of the test pattern. Third, the test sample is assigned to the category that has the maximum number of representatives between the nearest neighbor and neighbor.

2. Methodology:

Material and methods:

The work includes some operations started by resizing the collected images to 512×512 and converted them to gray images, three CT images are used for each case in which (9) CT images is the study samples. A binary mask is created to remove the outer skull from the brain tissue to avoid the confusing which may happened in the region of interest isolation, the traditional segmentation method is the threshold is which is the first used in a try to separated the tumor from the rest images each image have its threshold depending on the type of tumor if it is cancer or stroke (Hemorrhage and Ischemic). The median filter is used as an enhancement technique to remove the unwanted information which may be consider as a noise with window size of 7×7. The images have been converted to the (HSV) color to make the separation of color layer easy and more accurate, this process is done for the reason of using K-NN classification process. After the region of interest produced the statistical features are calculated using the second order features co-occurrence.

Preprocessing: CT scans need pre-treatment because of the nature of irregular brain tissue
1-The images received from the CT scan are usually colored by RGB components (red, green, and blue). However, it is converted to a gray scale image by removing the brightness in its composition
2- The Pre-processing include resize the images to 512×512 and converted to gray images.
3- Mask creation which is binary mask to separate the outer skull from the brain tissue.
4- Filtering :The resultant image needs to be filtered using the median filter to remove the distortion in the CT scan image.
5- The first step in the segmentation process is the threshold, which helps to separate the area of strokes and cancer form the rest image.
6- The K-NN supervised classification process is applied to classify the brain tissue into normal region and abnormal.
7-After the region of interest is obtained the statistical features are calculated which are the second order co-occurrence (Homogeneity, contrast, correlation and energy). after extract some textural features for the segmented ischemic, hemorrhage stroke and cancer area and both have normal and abnormal brain parts.

3. Conversion RGB to HSV:

The HSV stands for the Hue, Saturation and Value. The value represents intensity of a color, which is decoupled from the color information in the represented image. The hue and saturation components are intimately related to the way human eye perceives. The transformation equations for RGB to HSV color model conversion is given below [8].

\[ V = \max(R, G, B) \] ... ... (1)
\[ S = \frac{V - \min(R, G, B)}{V} \] ... ... (2)
\[ H = \begin{cases} \frac{6}{65} & , \text{if } V = R \end{cases} \] ... ... (3)
\[ H = \begin{cases} \frac{2}{6} + \frac{V - R}{65} & , \text{if } V = G \end{cases} \] ... ... (4)

4. Classification Process: K-NN is a method of supervised classification that classify objects based on nearest pixel. So it is proceed to distinguish the normal brain tissue from the abnormal area depending on the textural feature. For this reason, The image is colored after the removal of the skull by HSV color (for stroke and cancer). HSV is best way in our cases because, in the RGB layer, the color interfere with each other. As hue (H) Ranging from 0 to 1, the corresponding colors vary from red, during yellow, green, cyan, blue, and magenta, back to red. As saturation(S) ranging from 0 to 1, the corresponding colors vary from unsaturated (shades of gray) to fully saturated (There is no white component). As value (V), or brightness ranging from 0 to 1. The corresponding colors become brighter. The images are categorized into 5 classes, each class indicates a color layer. These classes are grouped in the following forms in the categories (stroke and cancer) and the normal tissue classes in the
brain. The (strokes, cancer) group may include one or two classes while the normal tissue group includes three or four classes. Figure (1,2,3) shows the hemorrhage stroke, figures (4,5,6) shows the ischemic stroke and figures (7,8,9) shows the cancer cases.

5. Co-occurrence Matrix: Synchronous matrices are important and powerful algorithms for statistical fabric. A common presence matrix is a two-dimensional graph, which indicates the number of times density pairs occur in a given spatial relationship [9].

Synchronous arrays are created by considering that each pixel contains eight neighbors (horizontally, vertically, and diagonally at 45 degrees). Furthermore, it is assumed that the relative frequency matrix for the occurrence of variation in grayscale levels can determine the texture context information. Many fabric measures can be obtained from these matrices such as homogeneity and contrast [10].

The texture is determined by the relative frequency matrix for frequency occurrence \( p(i,j) \), this matrix indicates how often each pixel is adjacent to an image, separated by a space \( i \) for one gray tone \( j \) for the other pixels. These matrices of gray spatial dependence frequencies are the function of the angle relationship between the adjacent pixels, in addition to the distance function between them. Synchronous arrays are based on the fabric. This composition is rapidly different in fine textures and more slow in coarse textures. The texture classification process is based on criteria (features) derived from these matrices [11].

- **Contrast**: Contrast is the measurement of intensity or differences in the gray level between the reference pixel and the surrounding pixel [12].

\[
C = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - j)^2 p(i,j) \quad (5)
\]

- **Correlation**: The linear dependency correlation of grayscale values is calculated in the common presence matrix. Demonstrates how the reference pixel is linked to the surrounding pixels[12].

\[
cor = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i,j)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \quad (6)
\]

Where:

\( \mu_x, \mu_y, \sigma_x, \text{ and } \sigma_y \) Are the means and standard deviation. The correlation is high when the values are uniformly distributed in the matrix and low otherwise.

- **Entropy**: Is a randomized measurement, is one of the statistical characteristics that determine the nature of fabric image input, randomized measurement of gray level distribution. The entropy is high if gray levels are randomly distributed around the image. Complex texture indicates that entropy is high. Or entropy are few if gray levels are distributed on a regular basis, which is inversely related to energy [13].

\[
H = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)log(p(i,j)) \quad (7)
\]
- Energy: The energy indicated how the gray levels are distributed in the image [14]

\[\text{Energy} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} [p(i,j)]^2 \quad \ldots \quad \ldots \ldots \ldots \quad (8)\]

\(p(i,j)\) is a histogram of a digital image with levels of intensity \([0, N_g-1]\). Energy measurement has a maximum value of 1 for an image of constant value and becomes increasingly smaller as pixel values are distributed across more gray level values, and all values of \(p(i,j)\) are less than or equal to 1. The greater the value, the easier it is to compress image data. If the power is high, the number of gray levels in the image is low, meaning that the distribution is concentrated in only a small number of different gray levels [12].

- Homogeneity: homogeneity image is characterized only by a few gray levels and gives only a few relatively high values. Thus, the sum of the squares too, would be of high measures, and homogeneous (Hom) are [13]:

\[\text{Hom} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{p(i,j)}{1 + |i-j|} \quad \ldots \quad \ldots \ldots \ldots \quad (9)\]

Table (1,2,3) gives information about the statistical feature which are calculated from co-occurrence matrix. Tables and graphs show the contrast, correlation, energy and homogeneity for ischemic, hemorrhage stroke and cancer. Table(4) gives the average value of each table.

6. Discussion: from table(4) which is the textural features for average value for all cases and figure(10) shows that the maximum value of the contrast is for the Ischemic and the minimum value is for the hemorrhage, it determines the intensity difference between a pixel and its neighborhood, when the differences is high this means the deformation in the texture is high. As the ischemic has the highest value of the contrast it has the highest value of homogeneity and the energy, since the homogeneity measuring the purity of the image texture and the energy feature measure the regularity of intensity level distribution, when the distribution of intensities number is small the energy value is high, the cancer has the maximum value of correlation and the ischemic has the minimum value, it measure the pixel correlation and how it correlated with its neighborhood.

7. Conclusion: The stroke refer to the block of the blood vessel or to the bleeding occur in the blood vessel Hemorrhage stroke, the blood leaks into the brain tissue, so it homogeneity is lower than the ischemic, so is the energy and the contrast while, the ischemic the blood clots stop the flow of the blood to the area of the brain. Cancerous brain tumors can spread within the brain to other parts of the body. They lack distinct boundaries because of their tendency to send "roots" to nearby natural tissues it cell are highly correlated comparing with the stroke.
References

[1] Tuceryan, M. and Jain A. 1998. Texture analysis, in The Handbook of Pattern Recognition and Computer Vision, 2nd Edition, World Scientific Publishing Co., pp:207-248.

[2] Hatam, H. 2009. Breast Cancer Tissues Recognition Using Fractal Dimension, Fuzzy Logic and Neural Network, M.Sc. Thesis, Department of Computer Science, College of Science, Al-Nahrain University, Iraq.

[3] Bhiwani, R.J., Khan M.A., and Agrawal, S.M. 2010. Texture Based Pattern Classification, International Journal of Computer Applications. 1(1):54-56.

[4] Dalia N. Abdul Wadood. 2014. Detection and Localization of Non-Melanoma Skin Cancer Using Texture Analysis, M.Sc. Thesis, Department of Computer Science, College of Science, Baghdad University, Iraq.

[5] Suraksha Ransingh and Manjusha Singh. 2015. A Survey Paper on Image Retrieval Based on Colour Feature, International Journal on Recent and Innovation Trends in Computing and Communication, 3(2):96-98.

[6] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein. 1973. Textural Features for Image Classification, IEEE Transactions On Systems, Man and Cybernetics, SMC-3(6):610-21.

[7] Markus Turtinen. 2007. Learning and Recognizing Texture Characteristics Using Local Binary Patterns, Academic dissertation, University of Oulu, Finland, C 278.

[8] Neelamma, K. P. and Ravi, M.Y. 2011. "Comparison between HSV and YCbCr Color Model Color-Texture based Classification of the Food Grains". International Journal of Computer Application. Vol.34, No.4, PP:51-57.

Information Technologies, National Engineering School of S-fax, Tunisia.

[9] Mathiapietikainen and Rosenfeld, A. 1981. "Image segmentation by Texture Using Pyramid Node Linking", IEEE Transactions on System, Man, and Cybernetics, Vol. SMC-11, No. 12.

[10] Clangphukhieo, B.; Aimmanee, P. and Uyyanonvara, B. 2014. "Segmenting the Ventricle from CT Brain Image using Gray-Level Co-occurrence Matrices (GLCMs)". Proceedings of the World Congress on Engineering, Vol.I, London, U.K.

[11] Kharrat, A.; Gasmi, K.; Messaoud, B.; Benamrane, N. and Abid, M. 2010. "A Hybrid Approach for Automatic Classification of Brain MRI Using Genetic Algorithm and support Vector Machine"

[12] Alyaa Hussein Ali, Shahad Imad Abdulsalam, Ihssan Subhi Nema, 2015. "Detection and Segmentation of Hemorrhage Stroke using Textural Analysis on Brain CT Images " International Journal of Soft Computing and Engineering (IJSCE), Volume-5 Issue-1, March 2015.

[13] Alyaa H. Ali, Kawther.A.K. Ihssan S.Nema 2014. "Detection of Brain Tumor for MRI using Hybrid MethodWavelet and Clustering Algorithm ". International Journal of Applied Information Systems (IJAIS), Volume 6—No. 7, January, p.21.

[14] Alyaa H. Ali, Entethar M. Hadi, 2015. "Diagnosis of Liver Tumor from CT Images using Digital Images Processing" International Journal of Scientific & Engineering Research, Volume 6, Issue 1, January-2015.
Class 1 | Class 2 | Class 3 | Class 4 | Class 5
---|---|---|---|---

**Image 1** shows Image No.1 with K-NN classification classes.

**Image 2** shows Image No.2 with K-NN classification classes.

**Image 3** shows Image No.3 with K-NN classification classes.

**Image 4** shows Image No.4 with K-NN classification classes.
| Image5 after remove skull | Color image | Class1 | Class2 | Class3 | Class4 | Class5 |
|--------------------------|-------------|--------|--------|--------|--------|--------|
| ![Image5 after remove skull](image1) | ![Color image](image2) | ![Class1](image3) | ![Class2](image4) | ![Class3](image5) | ![Class4](image6) | ![Class5](image7) |

stroke | Normal brain tissue

Figure 5. shows Image No.5 with K-NN classification classes.

| Image6 after remove skull | Color image | Class1 | Class2 | Class3 | Class4 | Class5 |
|--------------------------|-------------|--------|--------|--------|--------|--------|
| ![Image6 after remove skull](image1) | ![Color image](image2) | ![Class1](image3) | ![Class2](image4) | ![Class3](image5) | ![Class4](image6) | ![Class5](image7) |

stroke | Normal brain tissue

Figure 6. shows Image No.6 with K-NN classification classes.

| Image7 after remove skull | Color image | Class1 | Class2 | Class3 | Class4 | Class5 |
|--------------------------|-------------|--------|--------|--------|--------|--------|
| ![Image7 after remove skull](image1) | ![Color image](image2) | ![Class1](image3) | ![Class2](image4) | ![Class3](image5) | ![Class4](image6) | ![Class5](image7) |

cancer | Normal brain tissue

Figure 7. shows Image No.7 with K-NN classification classes.

| Image8 after remove skull | Color image | Class1 | Class2 | Class3 | Class4 | Class5 |
|--------------------------|-------------|--------|--------|--------|--------|--------|
| ![Image8 after remove skull](image1) | ![Color image](image2) | ![Class1](image3) | ![Class2](image4) | ![Class3](image5) | ![Class4](image6) | ![Class5](image7) |

cancer | Normal brain tissue

Figure 8. shows Image No.8 with K-NN classification classes.
Figure 9. shows Image No. 9 with K-NN classification classes.

| Table 1. Shows the textural features for hemorrhage stroke. |
|----------------------------------------------------------|
| Region of interest | Class3 | Class1 | Class2 | Class4 | Class5 |
|---------------------|--------|--------|--------|--------|--------|
| Contrast            | 0.0029 | 0.16785| 0.0901 |
| Correlation         | 0.89205| 0.97625| 0.96535|
| Energy              | 0.97645| 0.7778 | 0.87905|
| Homogeneity         | 0.9988 | 0.98925| 0.9926 |

| Table 2. Shows the textural features for ischemic stroke. |
|----------------------------------------------------------|
| Region of interest | Class3 | Class1 | Class2 |
|---------------------|--------|--------|--------|
| Contrast            | 0.06815| 0.03535| 0.03255|
| Correlation         | 0.98435| 0.92855| 0.931  |
| Energy              | 0.481  | 0.471  | 0.4971 |
| Homogeneity         | 0.9847 | 0.9823 | 0.98375|
Table 3. Shows the textural features for cancer case.

| Region of interest | 0.10175 | 0.03245 | 0.10175 |
|--------------------|---------|---------|---------|
| Contrast           | 0.9528  | 0.9688  | 0.9836  |
| Correlation        | 0.9597  | 0.88305 | 0.52435 |
| Energy             | 0.99545 | 0.99465 | 0.9664  |
| Homogeneity        |         |         |         |

Table 4. Shows the average value for the statistical features for the Stroke and cancer case.

| Infection   | Contrast   | Correlation | Energy    | Homogeneity |
|-------------|------------|-------------|-----------|-------------|
| Cancer      | 0.063533   | 0.9684      | 0.789033  | 0.9855      |
| Ischemic    | 0.08695    | 0.94455     | 0.877767  | 0.99355     |
| Hemorrhage  | 0.04535    | 0.947967    | 0.483033  | 0.983583    |

Figure 10. shows the textural features for co-occurrence matrix.