Brain Tumors Classification by Using Gray Level Co-occurrence Matrix, Genetic Algorithm and Probabilistic Neural Network

Raghad Majeed Azawi (MBChB)\textsuperscript{1}, Dhahir Abdulhade Abdulah (PhD)\textsuperscript{2}, Jamal Mustafa Abbas (PhD)\textsuperscript{3} and Ibrahim Tareq Ibrahim (MSc)\textsuperscript{4}

Abstract

Background: Brain tumors classification by MRI (Magnetic Resonance Imaging) is important in medical diagnosis because it provides information associated with anatomical structures as well as potential abnormal tissues necessary for treatment planning and patient's case follow-up. There are a number of techniques for medical image classification. In this paper brain tumors detection and classification system are developed into seven tumors types. The image processing techniques such as preprocessing by using a mean filter and feature extraction have been implemented for the detection of a brain tumor in the MRI images. In this paper, extraction of texture features using GLCM (Gray Level Co-occurrence Matrix). We used Probabilistic Neural Network Algorithm (PNNA) for image classification technique based on Genetic Algorithm (GA) and K-Nearest Neighbor (K-NN) classifier for feature selection is proposed in this paper.

Objective: MRI brain tumors detection and classification system by using GA and PNN which able to diagnose different types of tumors in human brain.

Patients and Methods: Medical image techniques are used to imaging the internal structures of the human body for medical diagnosis. Image processing is an effective field of research in the medical field. MRI dataset, obtained from the Atlas Website of Harvard University.

Results: Brain Tumors are classified by using the genetic algorithm where the total number of features (20 features) has been reduced to 10 features as the strongest features in the classification.

Conclusion: MRI brain image is one of the best methods in brain tumor detection and classification, by observing only MRI images the specialists are unable to keep up with diagnosing. Hence, the computer-based diagnosis is necessary for the correct brain tumor classification.

Key words: Brain tumors, MRI, GLCM, Classification accuracy, Genetic Algorithm (GA), K-NN and PNN.

Corresponding Author: raghanma2016@yahoo.com

Received: 6\textsuperscript{th} May 2018

Accepted: 24\textsuperscript{th} June 2018

https://doi.org/10.26505/DJM

\textsuperscript{1,4} College of Medicine- Diyala University - Diyala- Iraq.

\textsuperscript{2,3} Department of Computer Science- College of Science- Diyala University - Diyala-Iraq.
Introduction

The brain is the kernel part of the body. It has a very complex structure and hidden from direct view by the protective skull. This skull gives brain protection from injuries as well as it hinders the study of its function in both health and disease. The human body is made of many cells. Each cell has a specific job. The cells grow within the body and are divided to reproduce new cells. These divisions have certain functions in the body. But when each cell loses the ability to control its growth, these divisions are done without any limitations, and tumor consists. The brain is the central part of the human body responsible for coordinating and observing all other body organs, so if a tumor is present in any part of the brain then the activities controlled by this part of the nervous system are also affected. There are two types of brain tumors malignant tumor and benign tumor[1]. MRI is a medical imaging technique. Radiologist used it for the visualization of the internal structure of the body. It provides rich information about human soft tissues anatomy and helps for diagnosis of the brain tumor. Images obtained by the MRI are used for analyzing and studying the behavior of the brain [2].

Literature Review

Vinayadth V. Kohir and Sahebgoud H. Karradi, in 2015 [3] presented the artificial neural network approach, namely Back Propagation Networks algorithm (BPNs) and Probabilistic Neural Network algorithm (PNN) for classification of a brain tumor. Otsu's method of the threshold is used for detection of a tumor in the MRI brain images. Feature extraction stage is achieved using gray level co-occurrence matrix (GLCM). This proposed system (work) is aimed to design a system which is able to diagnose two types of tumors in a human brain (benign tumor and malignant tumor), using curvelet transform and PNN algorithm. Naveena H., Shreedhara K. and Mohamed R., in 2015 [4] the proposed system is to exploit the capability of ANN in the classification of MRI images to either cancerous or non-cancerous brain tumor. K-means clustering algorithm was used for segmentation. Then, gray level co-occurrence matrix (GLCM) was used for feature extraction of segmented image. Finally, Backpropagation Neural Network (BPN) and Probabilistic Neural Network (PNN) is used for the classification of brain tumors. The overall accuracy of the presented system is 79.02% in case of BPN and 97.25% in case of PNN.

Ata'a A. and Dhia A. in 2016 [5] this system is to detect and define tumor type in MRI brain images. The proposed system consists of multiple phases. The preprocessing stage the MRI image. Step two, transformations (features extraction algorithm based on using two level of 2-D discrete wavelet (DWT) and multiwavelet (DMWT) decomposition). Step three, the statistical measurements utilized to extract features from (GLCM). Step four, which deals with classification utilized (PNN) and the final Step, a proposed algorithm to segment, Superpixel Hexagonal Algorithm. segment, Superpixel Hexagonal Algorithm.
The accuracy of testing in DWT is 91% and in case DMWT is 97%.

Aswathy S., and et.al, in 2017 [6] designed a system for brain tumor segmentation using a genetic algorithm with SVM classifier. The proposed system is consisting of multiple phases. Step one is Pre-processing using the high pass, low pass and median filter for preprocessing. Step two, the segmentation by using a combination of Expectation Maximization Algorithm (EMA) and the level set method. Step three, feature extraction and selection using GA. Step four, classification MRI brain image to normal or abnormal by using SVM. The present work segments the tumor using Genetic Algorithm and classification of the tumor by using the SVM classifier.

Medical Image Analysis

In medical image analysis techniques have played a major role in many medical applications. In general, applications include the extraction of automatic features of the image that are subsequently used for a variety of classification tasks, such as the distinction between natural tissues from abnormal tissues. Depending on a particular classification job, extracted features may be shaped properties, color properties, or some formative properties of the image [7].

Magnetic Resonance Imaging (MRI)

MRI is a medical imaging technique viewed body's internal structure and gives high-quality images. MRI is safe for pregnant women and babies because of not using ionizing radiations in the imaging process [8]. Quick and reliable detection and classification of brain cancer are of significant technical and economic importance to the doctors.

The aim of this paper is to design an MRI brain tumor detection and classification system by using GA and PNN which able to diagnose different types of tumors in a human brain from tumor anatomical meaning and intensity. With the objective of utilizing more meaningful information to improve brain tumor and help doctors in the clinical diagnosis and accurate detection of the disease. The system supports physicians to prevent errors while identifying and classifying tumors. This is motivated by potential performance improvement in the general automatic and giving reliability in decision-making and rapid detection of brain cancer this technology is of great and economic significance to doctors.

The Proposed System

The proposed automated method is to classify the MRI brain images into two categories normal and abnormal, where the abnormal images are further classified into six categories namely Lymphoma, Glioblastoma Multiform, Cystic oligodendroglioma, Ependymoma, Meningioma and Anaplastic Astrocytoma. The automated method has five stages figure (1) show the details and block diagram of the proposed system. In the first, stage image acquisition. In the second stage, image preprocessing techniques such as noise removal and image enhancement. In the third stage, feature extraction by Gray-Level Co-occurrence Matrix (GLCM). In the fourth
stage, subset feature selection by Genetic Algorithm (GA). In the last stage, classification of MRI brain tumor in (PNN).

Figure (1): Block Diagram of the Proposed System.

**Image Acquisition Stage**

Medical image acquisition is very significant to the diagnosis of disease. Therefore speeding up of medical image acquisition is very important to the detection of disease. For the application of the proposed detection and classification system for magnetic resonance images of brain tumors, a dataset was collected from several sources for different categories of the most common brain tumors. As mentioned before, in this system six types of brain tumors and the normal case of the images of the MRI were used.

The classification system was implemented on MRI dataset, obtained from the Whole Brain of the Atlas Website of Harvard University [9], as there is in the database are available from these types of brain tumor. Figure (2) presents the samples of these types of brain MRI Image. The number of images in the dataset is 140 (20 images for each class) with 8 bit (pixel value from 0 to 255) and the type of them is BMP with various image sizes and its colored or gray level.
Brain Tumors Classification by Using Gray Level Co-occurrence Matrix, Genetic Algorithm and Probabilistic Neural Network

Raghe Majeed Azawi

Figure (2): Samples of Different Tumor Types and Normal of MRI Images.

Preprocessing Stage
This step is carried out to improve the quality of the image to make it ready for further processing. This improved and enhanced image will help in detecting edges and improving the quality of the overall image. Edge detection will lead to finding the exact location of the tumor [8]. Noise removal by using Mean Filter: in medical image processing, medical images are corrupted by a different type of noises. It is very important to obtain precise images to facilitate accurate observations for the given application. The Mean Filter is a linear filter and uses a mask over each pixel in the image. Each of the components of the pixels which fall under the mask is averaged together to form a single pixel. The mean filter, as shown in equation (1) is defined by:

\[
\text{Mean Filter } (x_1 \ldots x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i
\]

Where \((x_1\ldots x_N)\) is the image pixel range [10].

Feature Extraction Stage
Haralick has proposed textural features contain information about image contextual characteristics such as contrast, homogeneity, gray-tone linear dependencies, the complexity of the image and the number and nature of boundaries present. Texture features contain information come from blocks of image data surrounding the area being analyzed. Haralick used the co-occurrence probabilities using GLCM for extracting various texture features for the first time.
Brain Tumors Classification by Using Gray Level Co-occurrence Matrix, Genetic Algorithm and Probabilistic Neural Network
Ragheed Majeed Azawi

GLCM is also called as Gray Level Dependency Matrix (GLDM). GLDM is defined as "A two-dimensional histogram of gray levels for a pair of pixels, which are separated by a fixed spatial relationship". Following equations define some of the features proposed by Haralick et. al and notations are used to explain the various features. Table (1) show the number of features used from GLCM Matrix. For getting more reliable texture feature multiple GLCMs are computed for different directions at (0°, 45°, 90° and 135°) which can give the spatial relationship between neighboring pixels. This method reduces the computational complexity. After calculation for GLCMs of images, it is used to calculate features of the image which uniquely describes the images.

Table (1): The Number of Features from GLCM Matrix.

| Gray Level Co-occurrence Matrix–Clausi | Inverse difference normalized, Inverse difference moment normalized. |

1. Energy = \[ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P^2(i,j) \] (2)
2. Entropy = \[ -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j) \log_2 (P_d (i,j)) \] (3)
3. Contrast = \[ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j) \times (i-j)^2 \] (4)
4. Homogeneity = \[ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{P(i,j)}{1 + (i-j)^2} \] (5)
5. Variance (v) = \[ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - \mu_x)^2 p(i,j) \] (6)
6. Dissimilarity = \[ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} |i - j| \times P (i, j) \] (7)
7. Maximum Probability = max \{ p(i,j) \} (8)
8. Sum Entropy (SE) = \[ -\sum_{i=2}^{2N_g} p_{x+y}(i) log p_{x+y}(i) \] (9)
9. Sum Variance (SV) = \[ \sum_{i=2}^{2N_g} (i - SA)^2 p_{x+y}(i) \] (10)
10. Sum Average (SA) = $\sum_{i=2}^{2Ng} p_{x+y}(i)$

11. Difference Variance (DV) = $\sum_{k=0}^{Ng-1} [k - \sum_{i=0}^{Ng-1} p_{x+y}(i)]^2 p_{x+y}(i)$

12. Difference Entropy = $-\sum_{i=0}^{Ng-1} p_{x-y}(i) \log [p_{x-y}(i)]$

13. Cluster Shade = $\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} [i + j - \mu_x - \mu_y]^3 p(i,j)$

14. Cluster Prominence = $\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} [i + j - \mu_x - \mu_y]^4 p(i,j)$

15. Autocorrelation = $\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} ij p(i,j)$

16. Inverse Difference Moment (IDM) = $\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{p(i,j)}{1+(i-j)^2}$

17. Inverse Difference Normalized (IDN) = $\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{p(i,j)}{1+(i-j)}$

18. Correlation = $\frac{\sum_{i,j}(i)(j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$

*Where $\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$ are the means and standard deviations of $p_x$ and $p_y$.

$\mu_x = \sum_{i=0}^{ng-1} i \sum_{j=0}^{ng-1} p(i,j)$

$\mu_x = \sum_{j=0}^{ng-1} j \sum_{i=0}^{ng-1} p(i,j)$

$\sigma_x = \sum_a (a - \mu_x)^2 \sum_b p(a - b)$

$\sigma_y = \sum_b (b - \mu_y)^2 \sum_a p(a - b)$

19. Information Measure Correlation 1 (IMC 1) = $\frac{H_{xy} - H_{xy1}}{\text{Max}(H_x,H_y)}$

20. Information Measure Correlation 2 (IMC 2) = $\sqrt{1 - \exp(-2(H_{xy2} - H_{xy}))}$

*Where, $H_x$ & $H_y$ are the entropies of $p_x$ & $p_y$, respectively. While:

$H_{xy1} = -\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i,j) \log 2 [p(i) p(j)]$

$H_{xy2} = -\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i,j) \log 2 [p(i) p(j)]$
Feature Selection Stage

Our proposed approach combines genetic algorithm and K-NN to improve the classification accuracy of MRI brain tumor data set. We used the KNN as an assistant classifier (tool) for the genetic algorithm to calculate the classification accuracy of the population (the first step of the genetic operation) and use this accuracy to calculate the fitness function for the individual (second step of the genetic operation). The genetic algorithm search works as a goodness measure to prune redundant and irrelevant attributes. The classification algorithm is built based on evaluated attributes and least ranked attributes are removed.

A. K-Nearest Neighbor Classifier

K-Nearest neighbor is based on supervised learning. The aim is to find nearest k sample from the existing training data when new samples appear and classify the appeared sample according to most similar class. Generally, closeness is defined with Euclidean distance. Euclidean distance between any two samples or vectors Xi and Xj is given in (28).

\[
D (X_i, X_j) = \sqrt{\sum_{i=1}^{N} (X_i - X_j)^2}
\]  

B. Genetic Algorithm

Genetic algorithms are search and maximization methods that work similar to the evolutionary continuum at nature. It searches the best solution for multi-dimensional search space according to "best live" principle. Genetic algorithms produce a solution set that includes different solutions instead of producing only one solution. Hereby, many points at search space are evaluated at the same time and probability of reaching a total solution is increasing. To evaluate each matrix in this population, the input patterns of wavelet transform (M) are multiplied by the matrix (Gi) and producing a set of transformed patterns (N) which are then sent to a classifier (KNN) as in the equation [29]. This is illustrated in figure (3).

\[
N=M \times Gi
\]  

Where N is the transformed patterns, M is the input patterns. Each feature is encoded into a vector called a chromosome. A fitness value will be used to measure the fitness of a chromosome as in the function (30) and decides whether a chromosome is good or not.

\[
\text{Fitness} = W_A \times \text{Accuracy of K-NN} + \frac{W_{nb}}{N}
\]  

Where \( W_A \) is the weight of accuracy, and it's can be set from 0.75 to 100% according to user’s requirements and according to accuracy value and \( W_{nb} \) is the weight of N features participated in classification where \( N \neq 0 \). The figure (4) shows the weight of accuracy. Iteration number of (50, 100 and 500) is used as stopping criterion of the algorithm. Then we take the highest accuracy rate of the population and mean is calculated from the equation (31) for all features. Then detection powerful features in the classification of the true chromosomes (the mean is more than 50%). Table (2) shows the classification accuracy (Accuracy) versus the accuracy weight (\( W_A \)).
\[
\text{Mean} = \frac{1}{N} \sum P(i) \tag{31}
\]

*Where, Mean is the mean value, and N is total no. of chromosomes.

**Figure (3):** Classification Accuracy Using a GA-Based Features Extractor.

**Figure (4):** The Weight of Accuracy (WA).

**Table (2):** The Classification Accuracy (Accuracy) Versus Weight (WA).

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>WA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>0.75</td>
</tr>
<tr>
<td>20%</td>
<td>0.80</td>
</tr>
<tr>
<td>40%</td>
<td>0.85</td>
</tr>
<tr>
<td>60%</td>
<td>0.90</td>
</tr>
<tr>
<td>80%</td>
<td>0.95</td>
</tr>
<tr>
<td>100%</td>
<td>1</td>
</tr>
</tbody>
</table>
In the proposed method, an initial population of 280 chromosomes is randomly created and then ranked-based roulette wheel selection method is used to select the fittest chromosomes to generate next generation. The genetic operators, single point crossover, and mutation are used. The crossover rate is 0.7 and mutation rate is 0.01. Finally, 10 features are selected as the best features: Energy, Entropy, Variance, Contrast, Sum Entropy, Difference Entropy, Homogeneity, Cluster Prominence, Cluster Shade and Dissimilarity. These features are used as the input vector of PNN classifier for classifying the MRI images into 7 classes.

C. Proposed Algorithm

Begin
Step (1): Input patterns (M).

Step (2): Apply genetic search to generate the random population (Gi).

Step (3): Compute the transformed patterns (N) by applying the equation (29).

Step (4): Calculates the accuracy of the classifier (K-NN) and returns to GA.

\[
\text{Accuracy} = \frac{\text{no. of samples correctly classified in test data}}{\text{Total no. of samples in the test data}}
\]  

(32)

Step (5): Calculate the fitness value of the population by applying the function (30).

Step (6): Select the subset of higher fitness features.

Step (7): Crossover is done between the fittest individual.

Step (8): Mutation is done between the fittest individual.

Step (9): New population is created.

Step (10): If the generation is not ended, it will calculate fitness value.

Step (11): Select the generation for the highest accuracy of classification.

Step (12): Extract the correct chromosomes.

Step (13): Calculate the mean of all features for the correct chromosomes in eq. (31).

Step (14): Extract features that are higher than mean=50 %.  // Subset Feature.

Return V= Subset Feature.

END
D. Setting GA Parameters

The optimal solution of the proposed system is achieved when setting the GA parameters to the values in a table (3).

<table>
<thead>
<tr>
<th>NO.</th>
<th>Parameter Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Population Size</td>
<td>280 chromosome</td>
</tr>
<tr>
<td>2</td>
<td>Chromosome Length</td>
<td>20 genes</td>
</tr>
<tr>
<td>3</td>
<td>Number of Iteration</td>
<td>12 iteration</td>
</tr>
<tr>
<td>4</td>
<td>Number of Generation</td>
<td>50,100,500 by experiment</td>
</tr>
<tr>
<td>5</td>
<td>Crossover Rate</td>
<td>0.7</td>
</tr>
<tr>
<td>6</td>
<td>Mutation Rate</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Classification of Probabilistic Neural Network

In 1990, Donald F. Specht proposed a method to formulate the weighted-neighbor method in the form of a neural network. He called this a Probabilistic Neural Network, this is illustrated in figure (4). A probabilistic neural network (PNN) is a feed-forward neural network, which was derived from the Bayesian network and a statistical algorithm called Kernel Fisher discriminate analysis. It was introduced by D.F. Specht in the early 1990s. In a PNN, the operations are organized into a multilayered feed-forward network with four layers:

1. Input Layer
2. Hidden Layer
3. Pattern Layer / Summation Layer
4. Output Layer

PNN is often used in classification problems. When an input is present, the first layer computes the distance from the input vector to the training input vectors. This produces a vector where its elements indicate how close the input is to the training input. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities (calculate the PDF for the equation (33)). Finally, a
complete transfer function on the output of the second layer picks the maximum of these probabilities and produces a 1 (positive identification) for that class and a 0 (negative identification) for non-targeted classes [4].

\[ f_k(x) = \left( \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} \right)^{\frac{1}{N}} \sum_{i=1}^{N_k} \exp\left[-\frac{(x - x_{ki})^T(x - x_{ki})}{2\sigma^2}\right] \] (33)

Where \( d \) = denotes the dimension of the pattern vector(\( x \)). \( i \) = pattern number, \( N \) = denotes the total number of samples in class, \( X_{ki} \) = vector of \( i \)-th training pattern from class 1, \( T \) = vector transpose. The \( \sigma \) is the smoothing parameter which represents the single free parameter for this algorithm. The best value for the \( \sigma \) is obtained from the following equation (34).

\[ \sigma_j = \text{STD} (X_i) \] (34)

* Where \( X_i \) is the vector in training data and \( j \) number of classes.

**Results**

The proposed system includes the results of the classification methods. In this paper, an automatic brain tumor classifier was proposed. The proposed technique was implemented on MRI dataset (these are Lymphoma, Cystic oligodendroglioma, Glioblastoma multiform, Meningioma, Ependymoma and Anaplastic astrocytoma). The numbers of collected images are 140. The algorithm described in this paper is developed and successfully trained in Visual Basic.Net.2013 using a combination of image processing and neural network toolbox. The remaining 70 MRI brain images from different types will be utilized as testing data phase. The result represents that 70 images are classified correctly. The proposed system is to classify the MRI images brain with 10 GLCM features using the genetic algorithm. Classification rates of 4 cases (direction =0˚, 45˚, 90˚ and 135˚) are 98.57 %, 100%, 97.14% and 98.57 % respectively. The maximum classification rate of testing is 100% in case=45˚. Figure (5) illustrates a flowchart of the classification rate of the proposed system.
Discussion

Results obtained from several experiments and compared with other related works. The suitable population size for the GA in the proposed system is 100 to get the minimum number of features and the maximum value of accuracy to the population. The optimal solution of the proposed system is achieved in case (k=7) of the K-NN classifier. When calculating the PDF for the equation (33) in PNN classification, the best value for the σ is obtained from the equation (34).

Conclusion

MRI brain image is one of the best methods in brain tumor detection and classification, by observing only MRI images the specialists are unable to keep up with diagnosing. Hence, the computer-based diagnosis is necessary for the correct brain tumor classification. The new method is a combination of texture features extracted by using GLCM, Genetic Algorithm, K-NN to select subset features and Probabilistic Neural Network. By using this algorithm, an efficient brain tumor classification method is been constructed with a maximum classification rate of 100% in direction 45°. This method could serve inaccurate classification of Brain Tumor diagnosis.

References