Using Data Mining Techniques for Children Brain Tumors Classification based on Magnetic Resonance Imaging

Eman M. Ali
Department of Computer Science
Helwan University
Cairo, Egypt

Ahmed F. Seddik
Dean of Faculty of Computer
Science,
Nahda University.
Professor at the Biomedical
Engineering Department, Helwan
University.
Cairo, Egypt

Mohamed H. Haggag
Vice Dean for Student Affairs
at Faculty of Computers and
Information
Professor at Computer Science
Department Faculty of
Computers and Information,
Helwan University
Cairo, Egypt

ABSTRACT

MRI (Magnetic resonance Imaging) is one source of brain tumors detection tools, but using MRI in children brain tumors classification is considered to be difficult process according to the variance and complexity of tumors. This paper presents a technique for the classification of brain tumors based on children MRI.

 The proposed system consists of four stages, namely, MRI preprocessing stage, Segmentation stage, Feature extraction, and Classification stage. In the first stage, the main task is to eliminate the medical resonance images (MRI) noise found in images due to light reflections or operator performance which may cause inaccuracies in the classification process. The second stage, which is the stage where ROI is extracted (tumor region). In the third stage, the features related with MRI images using Haar wavelet transform (HWT) will be obtained. The features of magnetic resonance images (MRI) have been decreased using (HWT) to essential features only. And finally the fourth stages, where new classifier will be presented and finally the result will compare the proposed classifier with six other classifiers have been used.

Image classification is an important task in the medical field and computer vision. Image classification refers to the process of labeling images into one of a number of predefined categories. In this survey, the test of various classification techniques against each other will be present. And it is expected that TANN will provide better results in terms of sensitivity, specificity, accuracy and overall running time.

Keywords
Brain Tumor, MRI, Image Classification, Naïve Bayes, Decision Tree, Support Vector Machine, k-Nearest Neighbor.

1. INTRODUCTION
Early detection and classification of brain tumors are very important in, clinical practice. Several researchers have proposed different techniques for the classification of brain tumors using different sources of information [1]. In this paper, a process for children brain tumor classification, depending on the analysis of Magnetic Resonance (MR) images and Magnetic Resonance Spectroscopy (MRS) data collected for patients with benign and malignant tumors will be proposed. The aim is to achieve a high accuracy in discriminating the two types of tumors through a combination of several techniques for image denoising, image segmentation, feature extraction and classification. The

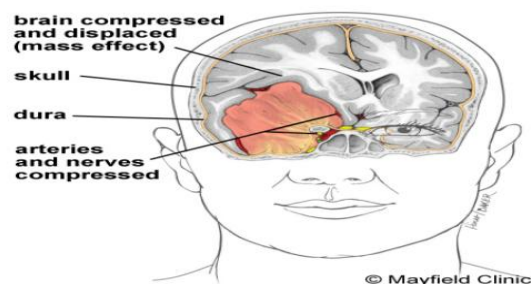proposed technique has the potential of assisting medical diagnosis.

Fig.1. Brain tumor compress and displace normal brain tissue, Increasing size, pressure and swelling cause symptoms like seizures or headaches.

Brain tumors are abnormal cells spread among the brain cells. Sometimes this malicious tissue replaces the original brain cells and causes what is called the primary tumors.

Also, these tumors may appear in other location in the body and may be transferred to the brain and this is called secondary tumors.

Primary tumors do not transfer from the brain region to any other organ in the body; these tumors may be malignant or benign tumors. But secondary brain tumors are always malignant tumors see Fig. 1[2].

Fig.2. Brain MRI with malignant tumor region [2].

These two types of tumors are life-threatening and cause death. The growth of the tumors affects the brain because space in the skull is limited. This growth causes the intracranial pressure to be increased and damage the other healthy tissues that perform important functions inside the brain [3].

Brain tumors are considered to be the second disease causing death in children and adults see Fig. 2.

According to the Central Brain Tumor Registry of the United States (CBTRUS) statistics, there are about 64.530 cases suffer from brain cancer among children by the end of this year [3].

Many kinds of research depending nowadays on using computer technology in medical diagnoses such as cancer-related research in brain, breast, and liver. MRI is one of the technologies used to take a photo of the body organs through the usage of the magnetic field. It has much higher features than other radiation tools such as x-ray and computed tomography (CT) [4]. The researcher had proposed various features for classifying tumor in MRI. The statistical, Intensity, Symmetry, Texture features etc., which utilize a gray value of tumors are used here for classifying the tumor. However, the gray values of MRI tend to change due to over – enhancement or in the presence of noise [5].

Though, the main objective of all proposed classification algorithms, whether it relies on characterizing texture to its statistics or modeling the medical image, aim to learn from how interestingly discern medical images to eventually develop "knowledgeable " computer systems. Thus, the objective of this paper presents an appraisal of the existing and conventional methods for the classification of medical images and based on these observations; propose a new framework for medical image classification. The rest of the paper is structured as Section 2 to Section 5.

Section 2 presents a survey of previous brain tumor identification and classification techniques. Section 3 illustrates the proposed framework followed by a comparative analysis of the presented classification techniques in Section 4. And finally the conclusion in Section 5.

2. SURVEY ON PREVIOUS TECHNIQUES

There are a large number of methodologies that present almost optimal results, but only for well-segmented images [6]. Some of these techniques achieved error rate above 10% in the segmentation phase only. In this section, some techniques that were concerned with tumor detection and classification will be described.

2.1 R. Mishra Tumor Identification System [4]

R. Mishra [4] uses the wavelet transform and NN (Neural network) to make a classification system to classify tumor region. Mishra makes use of the improvement of the wavelet packet, which is better in decomposing estimation as well as detail component every time as compared to wavelet transformation method. This technique achieved a good accuracy but have a long classification phase time according to the usage of the neural network which takes a very long time in the training and testing phase compared with other classification methods.

2.2 E. A. El-Dahshan Tumor Classification system [7]

E. A. El-Dahshan and T. Hosney [7] present a new system for tumor detection using MRI images. This technique uses a hybrid technique by using the artificial neural networks (ANN) and k-nearest neighbor (KNN).

This hybrid system gets a great performance by in classification, using these two kinds of classification techniques resulted in a very high classification time according to the usage of ANN and every time new information arrives needs new training and KNN precision can rapidly degrade when a number of features grow [7]. El-Dahshan's used also two kinds of feature extraction techniques, discrete wavelet transform (DWT) and then for selection of best-extracted features, the principle component analysis (PCA) method was used.

2.3 Selvaraj Tumor Classification System [8]

H. Selvaraj et.al [8] proposed a system for classification of MRI by also using wavelet transform to get the features which were the input to the classification phase. They make use also of two kinds of efficient classification techniques which were, support vector machine (SVM) and Artificial neural network (ANN).

In this technique neural network Self-Organizing Maps (SOM) were used as classifiers for tumor region which is used to capture nonlinear computation and SVM which, used to capture linear and nonlinear computation and this technique achieve a very high-performance rate 98% but still having the same classification time complexity due to the usage of the ANN (training and testing time).

2.4 Chaplot Tumor Identification System [9]

S. Chaplot and L.M. Patnaik [9] proposed a simple brain tumor identification system which is consisting of two phases only, feature extraction phase, and classification phase. They depend on using SVM to classify the brain as a normal or abnormal brain only.

This presented system makes use of the wavelet transformation and achieves a very good accuracy about 96% in classifying brain tumors.

2.5 Laskhari Tumor Detection System [10]

A.E. Laskhari [10] present a brain tumor detection in MR images based on Using neural network using geometric and Zernike moments. This presented technique consisted of 3 phases which are, an input image, feature extraction and finally the classification phase. In the first phase, they depending on using MRI images and Feature extraction phase occupies statistics features collection by mean, median, entropy and standard deviation as well as a non-statistic feature by geometric moment's invariants.

A selection of features was carried out by kernel F-score technique and given as input to Artificial Neural Network classifier. ANN classifies it into two classes as normal and cancer brain tissues.

A. E. Laskhari [11] proposed another technique based on the artificial neural network for brain tumor detection in MR images using Gabor Wavelets. This technique used PD, T1 and T2 weighted as input images. Preprocessing is performed for noise reduction to obtain excellent results. Feature extraction stage has two categories i.e. statistical features extraction by median, mean and variance wherever non-statistical feature extraction by Gabor wavelet technique. Feature selection stage is carried out by kernel F-score technique that is used to calculate the variation between two classes. Whilst features were chosen, these features were given as input to ANN for normal and abnormal brain tissues.

2.6 Kharrat Brain Tumor Classification System [1,7]

A.Kharrat et.al [1, 7] proposed a system for brain tumor classification using genetic algorithm and support vector machine (SVM).

The features have been extracted by two methods. Firstly, extraction of features from both normal and abnormal images is carried out by spatial gray level dependence method (SGLDM). Secondly, the image is decomposed at the second level by applying Daubechies wavelet transform. The optimal features were selected by the genetic algorithm. Those optimal features were given as an input to classifiers SVM. The accuracy rate achieved by this technique was 96.36% to 97.59%.

3. PROPOSED TECHNIQUE

The proposed system has mainly four modules namely Pre-processing, segmentation using Contribution-Based Clustering Algorithm, Feature extraction, and disease classification. According to the need of the next level, the pre-processing step converts the image. It performs filtering of noise and other artifacts in the image and sharpening the edges in the image. RGB to gray conversion and reshaping also takes place here. It includes a median filter for noise removal. The feature extraction is extracting the cluster, which shows the predicted tumor at the Haar wavelet transform output. The extracted cluster is given to the threshold process. It applies a binary mask over the entire image.

In the disease detection and classification step, the tumor area is calculated using the binarization method making the dark pixel darker and white brighter. In threshold coding, each transform coefficient is compared with a threshold and if it's less than the threshold value, it is considered as zero or else one. In the approximate reasoning step, the tumor area is calculated using the binarization method. That is the image having only two values either black or white (0 or 1). Here 200x200 JPEG image is a maximum image size. The binary image can be represented as a summation of a total number of white and black pixels. Pre-processing is done by filtering.

Segmentation is carried out by Content-based Image Retrieval (CBIR) algorithm. The feature extraction is done by considering the threshold and finally, approximating the classification method to recognize the tumor shape and position in MRI image using edge detection method [15] see Fig. 4.

3.1 Image Preprocessing

In the preprocessing stage, first transform the brain MRI image from RGB mode to Grayscale level and from eight bit to double precision pattern to get a high-resolution image of the brain while at the same time being noninvasive. The aim of this paper is to detect, segment and classify the tumor cells, but for the complete stage it needs the process of noise removal.

To get the best MR image quality, a median filter was used to remove any noise from the original image see Fig. 3.
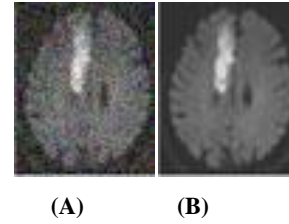

| (A) | (B) |

Fig.3. (A) : The Original MRI image with noise.

(B) : The same image after applying the median filter.

After applying the median filter to the noisy image now, the brain MR images are ready to go to the segmentation phase and isolate the tumor region fig. 4 state the block diagram of the proposed system.
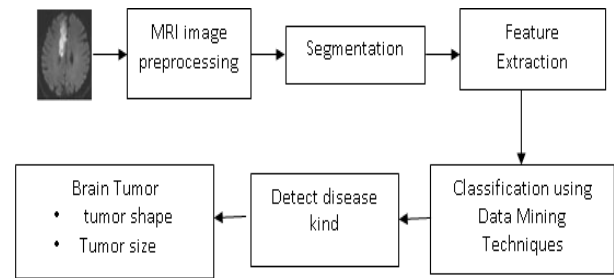

Fig.4. Block diagram of the proposed brain diseases classification system.

3.2 Segmentation Phase

In this paper, use one of the unsupervised classifications clustering forms, which are used to group the image pixels based on the similarity between this pixels. This partitioned clustering algorithm is based on the notion of 'contribution of a data point'. The content-based image retrieval algorithm will be applied and compare its performance with that of the k-means clustering algorithm. In comparison with the k-means algorithm, CBIR achieves a better result in measuring the similarity of both intra-cluster and inter-cluster.

The algorithm aims at partitioning a group of data points into disjoint clusters optimizing a specific criterion [2]. When the number of data points is large, a brute-force enumeration of all possible combinations would be computationally expensive. Instead, there are many algorithms are applied to find the optimal partitioning. The most popular criterion function used for partition clustering is the sum of squared error function given by:

$$E = \sum_{i=1}^{k} \sum_{x \in_i} (x - m_i)^2$$

(1)

Where k is the number of clusters, Ci is the ith cluster, x is a data point and mi is the centroid of the ith cluster. A widely used squared-error based algorithm is the k-means clustering algorithm [2]. Here, a clustering algorithm similar to the k-means algorithm will be used. The contribution of a data point belonging to a cluster is defined as the impact that it has on the quality of the cluster. This metric is then used to obtain an optimal number of clusters from the given set of data points.

While this work uses the concept of contribution to finding the optimal number of clusters, CBIR is used for optimal partitioning of the data points into a fixed number of clusters.

The proposed outline presents contribution-based clustering algorithm. It optimizes on two measures, namely the intra-cluster dispersion given by:

$$\alpha = \frac{1}{n} \sum_{x \in C_i} (x - m_i)^2$$

(2)

And the inter-cluster dispersion given by:

$$\beta = \frac{1}{k} \sum_{i=1}^{k} (m_i - \overline{m})^2$$

(3)

Where k is the number of clusters and $\overline{m}$ is the mean of all centroids. The algorithm tries to minimize α and maximize β.

After founding the tumor pattern region, a resize of the tumor region will take place to be 200 × 200 to have a suitable region to have enough features, to get a high performance in the next phase.

3.3 Feature Extractions

Once the local tumor pattern is found, which is a 200 × 200 image, now the local tumor features are found but, this tumor feature vector is too large to use it in the classification phase so, to generate a 50 ×50 feature vector for this pattern the Haar wavelet transform is used. Haar wavelet can be used to decompose the data in the tumor region into sub-components that appear in different resolution. It divides the tumor image into four sub-images. These resulted images consist of two high-resolution images one image that has been high pass in horizontal and vertical directions and one that has been low pass filtered in both directions.

In comparison with other methods by using the Haar wavelet transform, it successfully reduces the feature vector which affects the overall performance of the system and decrease the classification process overall time.

3.4 Classification Phase

Image classification refers to the labeling of images into one of a number of predefined categories. Classification system consists of a database that contains predefined patterns that compare with a detected object to classify into the proper category. Image classification is an important and challenging task in various application domains.

In this paper, a merge of two kinds of classification techniques to achieve a very high result which are, Tree Augmented Naïve Bayes (TAN) as a challenging classifier to improve the performance of the children brain tumors classifications process and nearest-neighbor classifier as very high classification accuracy and call the new algorithm as (TANNN).

3.4.1 *Tree Augmented Naïve Bayes Nearest Neighbor (TANNN)*

In many kinds of research, there are a lot of classification techniques applied and most of them based on detecting a sequence similarity between features extracted from the image and the pattern stored in the database.

Most of these classifiers use binary classifiers for classification of the image as a normal image or abnormal image only without detecting which kind of tumors the child

suffering from such as Neural Networks, Support Vector Machines (SVMs), etc.

This paper presents a framework using the Tree-Augmented Bayesian Networks Nearest Neighbor (TANNN) which performs multi-classification based on the theory of learning Bayesian Networks. In order to enhance TAN's performance, pre-processing of data is done by feature discretization and post-processing is done by using Mean Probability Voting (MPV) scheme. The advantage of using Bayesian approach over other learning methods is that the network structure is intuitive.

The TANNN algorithm is very simple, the task is to determine the probability that an image Q is belonging to a certain class C

The training data for a class is a collection of descriptors extracted from a set of labeled example images. These are stored in data structures that allow for efficient nearest neighbor searches (the nearest neighbor of descriptor di in class C is NNC (di)).

The aim of this new technique is to classify the image using the TAN as a normal or abnormal using the simple structure of using TAN and low time complexity using TAN.

After this process, the nearest neighbor is used to classify the tumor region to detect which kind of tumors the child suffers from using the MRI image database.

So, assume that TANNN will perform a great performance compared with all the previous techniques because it uses the two major advantages of TAN which is the classification speed and the accuracy according to the usage of nearest neighbor.

4. EXPREMENTAL RESULTS

In this section, the results obtained using a database of images is presented. Start by presenting the database with which conducted in tests, and then, present the results according to the used structure.

4.1 Database

The famous medical imaging is MRI. A magnetic resonance imaging (MRI) scanner uses powerful magnets to polarize and excite hydrogen nuclei (single proton) in human tissue, which produces a signal that can be detected and it is encoded spatially, resulting in images of the body.

MRI dataset consists of 191 Children MRI jpg format images, with 100 images containing malign tumor and 51 images containing benign tumor and 40 normal images. These images are divided into three categories, 70% for training, 15% for testing and 15% for validation.
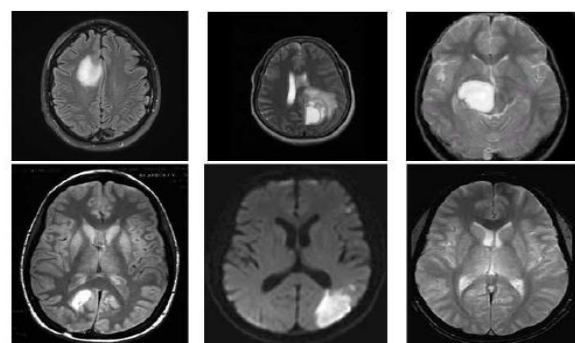

Fig.5. Samples of children brain dataset containing tumors.

4.2 Comparative Analysis

Next figures show the images as an output. I.e. gray scale image, median filter image, segmented image, finally extracted tumor from MRI image. For this purpose, real-time patient data is taken for analysis. As tumor in MRI image have an intensity more than that of its background so it become very easy locate it and extract it from MRI image.
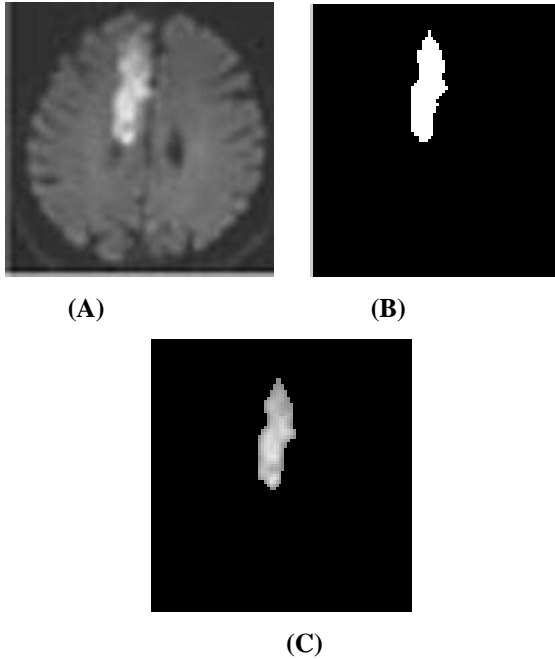

(A)

(B)

(C)

Fig.6. (A): MRI image of tumor affected brain gray scale image.

(B): Threshold segmented image.

(C): Final Extracted brain tumor from MRI image.

After using two kinds of segmentation techniques on these images, contribution based information retrieval achieving better performance compared with K-Means algorithm see Fig. 7 and Fig. 8.
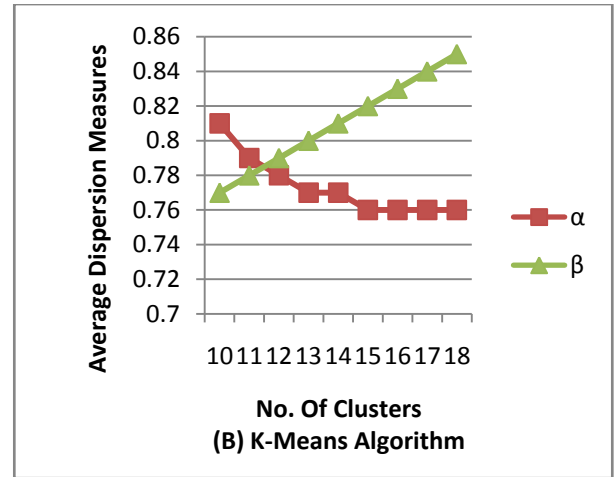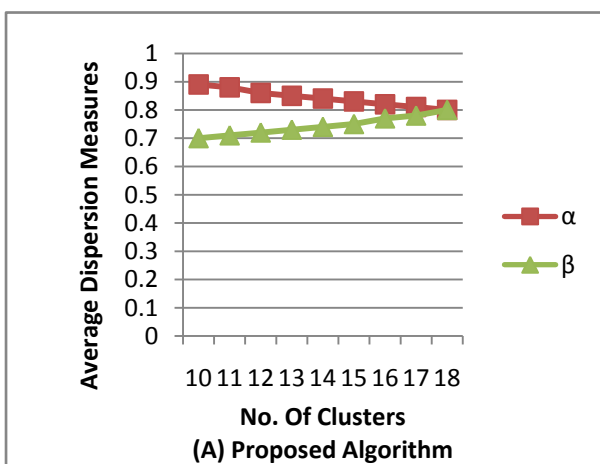

(A) Proposed Algorithm

(B) K-Means Algorithm

Fig.7. Average value of α and β against the number of clusters.
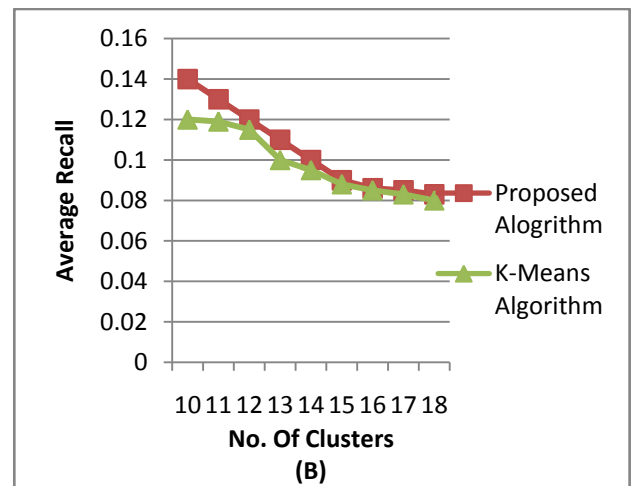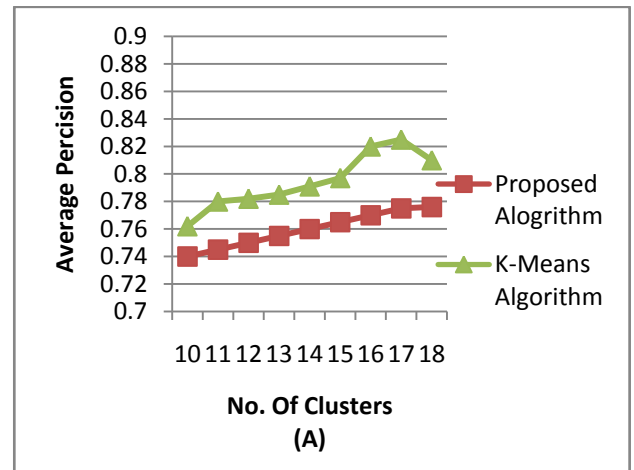

(A)

(B)

Fig.8. Average Precision and Recall against the number of clusters.

The next table shows the classification accuracy using CBIR and k-Means segmentation techniques.

Table1. The accuracy of the classification Techniques using K-means and CBIR segmentation algorithms

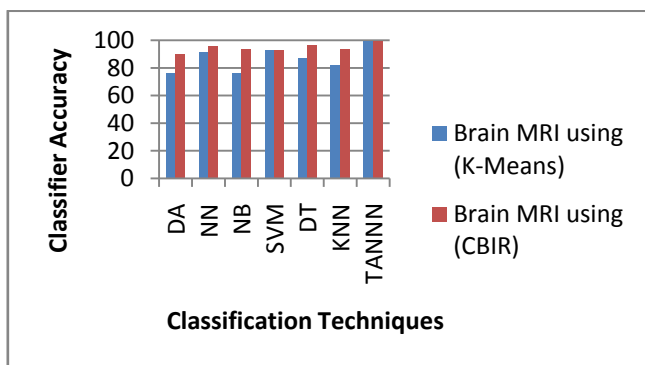| Segment ation	DA	NN	NB	SV M	DT	KN N	TAN NN
Brain MRI using (K-Means)	75.93	91.4 4	76.0 8	92.5 9	87.0 4	82.3	99.4
Brain MRI using (CBIR)	90.12	96.1 0	93.5 2	92.5 9	96.1 9	93.7	99.8

Fig.9. The accuracy of classification techniques using CBIR and K-mean Segmentation techniques.
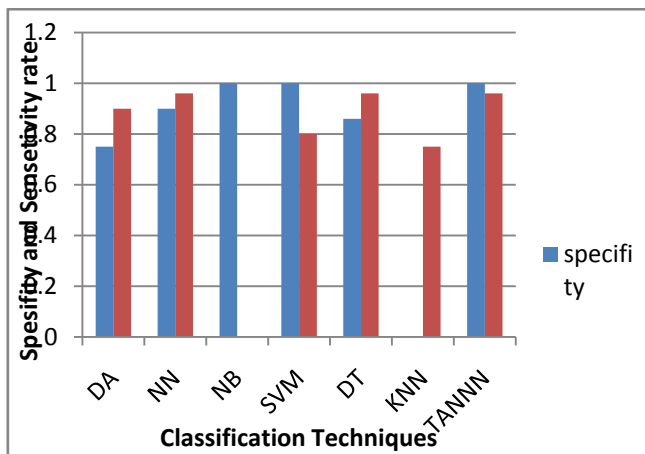

Fig.10. Classification results in terms of specificity and sensitivity.

The next table states a comparison between the classification techniques according to the classification time see table 2.

Table2. the classification Techniques Classification time.

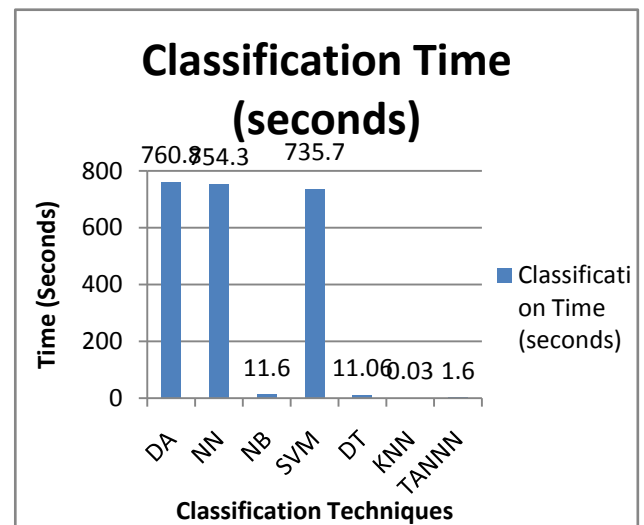| Classification Technique	Classification Time (seconds)
DA	760.8
NN	754.3
NB	11.6
SVM	735.7
DT	11.06
KNN	0.03
TANNN	1.6

Fig.11. A comparsion between classification techniques according to classification time.

From the results of experiments, the decision tree algorithm gave the best detection rate. It, achieving a classification rate of 96.19 %. But when considering computational performance, however, K-Nearest Neighbor algorithm proved to have a faster build time (Time it takes to build a model on network training data) at 0.03 seconds while having a detection rate of 93.7% as shown in table 1. Naïve Bayes had the second best build time at 1.6 seconds but a detection rate of 93.52%. Computational performance is particularly important when considering the real-time classification of potentially thousands of simultaneous networks traffic. From experiments, Naïve Bayes appears to be the best suited for real-time classification tasks due to its relatively fast classification speed and high detection rate.

5. CONCLUSION

This paper presents a survey on various image mining techniques that was proposed earlier by researchers for the better development in the field of content-based image retrieval. The purpose of the mining is to produce all considerable patterns without prior knowledge of the patterns. Important information can be hidden in images, conversely, few research talks about data mining on them. Image segmentation is the primary phase in image mining. In other words, image mining is simply an expansion of data mining in the field of image processing. Image mining handles with the hidden knowledge extraction, image data association and additional patterns which are not clearly accumulated in the images. Also, this paper provides a marginal overview for future research and improvements. Certain possible future investigations that are discussed may be done in the area of image mining which included the experimentations on other image elements such as textures, shape, etc.

In future, this program can be done more advanced so that tumor can be classified according to its type. Also, tumor growth can be analyzed by plotting the graph which can be obtained by studying sequential images of tumor affected patient.

The future research work may include the implementation of the Bayesian networks for relevance feedbacks and more extensive tests with other examples of image forensic work. It is also envisaged that subjective testing will be performed with input from forensic experts.

Some possible future studies that may be conducted in the area of image mining include the experimentations on other image elements such as textures, shape, and so forth. It will also be interesting to investigate hidden relationships among images. For example, intensive and extensive exploratory pattern analysis involved in the existing systems in the database can be very useful.

6. ACKNOWLEDGMENT

7. REFERENCES

[1] K. Kharat, P.Kulkarni, M.B.Nagori," Brain Tumor Classification Using Neural Network Based Methods ", International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012.

[2] Brain MRI dataset, www.CEwebsource.com, 2014.

[3] V. Gladis, P. Rathi, S. Palani, " BRAIN TUMOR MRI IMAGE CLASSIFICATION WITH FEATURE SELECTION AND EXTRACTION USING LINEAR DISCRIMINANT ANALYSIS ", Sudharsan Engineering College Sathiyamangalam, Pudukkottai , India,2014.

[4] Y. Jung , Y. Choi," Using Augmented Bayesian Networks to Compare Preference of Performance ",International Journal of Bio-Science and Bio-Technology Vol. 5, No. 1, February, 2013.

[5] A. Gale, S. Salankar," A Review On Advance Methods Of Feature Extraction In Iris Recognition System " , International Conference on Advances in Engineering & Technology – 2014 (ICAET-2014), PP 65-70,2014.

[6] D. Bheda, V.Nagar, "A Study on Features Extraction Techniques for Image Mosaicing" , International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 3, March 2014.

[7] K.Kharat , P.Kulkarni , "Brain Tumor Classification Using Neural Network Based Methods" , International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012.

[8] M. Nagori, S. Mutkule , "Detection of Brain Tumor by Mining fMRI Images", International Journal of Advanced Research in Computer and Communication Engineering , Vol. 2, Issue 4, January 2013.

[9] A. Jose, S.Ravi, M.Sambath , "Brain Tumor Segmentation Using K-Means Clustering And Fuzzy C-Means Algorithms And Its Area Calculation" , International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 3, March 2014.

[10] S. Anwar , R. Ibrahim, "A Framework for Medical Images Classification Using Soft Set", The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013), Malaysia.,2013.

[11] P. Kamavisdar, S. Saluja, S. Agrawal, "A Survey on Image Classification Approaches and Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2013.

[12] H. Aslam, T. Ramashri, M. Ahsan, "A New Approach to Image Segmentation for Brain Tumor detection using Pillar K-means Algorithm" , International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 3, March 2013.

[13] K.Kothavari, R.Keerthana, M.Mariselvam,S.Kaveya, L.Mekala, "A Hybrid approach for PNN-Based MRI Brain Tumor Classification and Patient Detail Authentication Using Separable Reversible Hiding" , International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 4, April 2013.

[14] V. Rekha, S. Anantharajan, "C.A.D.S., for Classification of MRI Brain Tumor Using Decision Tree", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 3, March 2014.

[15] M. Suganya, M. Menaka, "Various Segmentation Techniques in Image Processing: A Survey", International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014.

[16] C. P. Loizou, M. Pantziaris, C. S. Pattichis, I. Seimenis, "Brain MR image normalization in texture analysis of multiple sclerosis", Journal of Biomedical Graphics and Computing, Vol. 3, No. 1, 2013.

[17] P. Loizoua, C. Kyriacou, "Brain white matter lesion classification in multiple sclerosis subjects for the prognosis of future disability", Intelligent Decision Technologies, Pp. 3–10, 2013.

[18] Dr S. McKay, Dr R. Hadfield, "Current Knowledge in Brain Cancer Research", Developed for Cure Brain Cancer Foundation, Pp. 1-33, May 2014.

[19] A. Morgado, L. Caldeira, N. Silva, "Dynamic analysis of MR-PET data on brain tumors", 3rd Conference in PET/MR and SPECT/MR , Kos Island, Greece. 19-21 May 2014.

[20] B. Wei, Z. Yu, G. Yang, "A New Multistage Medical Segmentation Method Based on Superpixel and Fuzzy Clustering" , Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, Volume 2014, Article ID 747549, 13 pages, 2014 .

[21] Ms. P. Jayashri, Mrs. D.Gandhimathi, "COMPARATIVE STUDY ON DIFFERENT BRAIN TUMOR SEGMENTATION IN MRI IMAGES", International Journal of Advanced Technology in Engineering and Science, Volume No.02, Issue No. 09, September 2014.