

# Real Brain Tumors Datasets Classification using TANNN

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

Cancerous tumors considered being one of the acute diseases that cause the human death especially brain cancers.

Many computer-aided diagnosis systems are now widely spread to aid in brain tumors diagnosis. Therefore, an automated and reliable computer-aided diagnostic system for diagnosing and classifying the brain tumor has been proposed [1].

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 survey of the most famous techniques used for the classification of brain tumors based on children MRI [2].

The brain tumors detection and classification systems consist of four stages, namely, MRI preprocessing, Segmentation, Feature extraction, and Classification stages respectively. In the first stage, the main task is to eliminate the medical resonance images (MRI) noise which may cause due to light reflections or any inaccuracies in the imaging medium.

The second stage, which is the stage where the region of interest is extracted (tumor region). In the third stage, the features related to MRI images will be obtained and stored in an image vector to be ready for the classification process. And finally the fourth stages, where classifier will take place to specify the tumor kinds.

TANNN is a new classification technique user to get a very high performance compared with other classification techniques such as KNN, SVM, DT, and Naïve Bayes.

Image classification is an important task in the image processing and especially in the medical diagnosis field. 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.

## Keywords

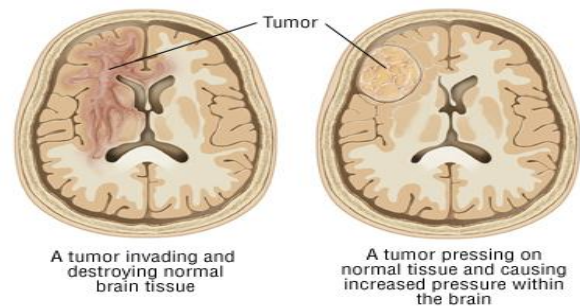
Brain Tumor, MRI, Image Classification, Naïve Bayes, Decision Tree, Support Vector Machine, k-Nearest Neighbor.

## 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, we test the TANNN technique upon three of the real brain tumors datasets. By analyzing the MR image and Magnetic Resonance Spectroscopy (MRS) a final decision can take place either the data collected from patient with benign or malignant tumors.

By using image processing techniques a tumor region can be obtained by using a combination of denoising, segmentation, feature extraction and classification techniques. The proposed technique has the potential of assisting medical diagnosis.



**Fig.1. The Brain tumor destroys the normal brain tissue and causing pressure within the brain**

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 [3].

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 brains 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 [4-6].

Computer technologies are widely used nowadays in medical diagnoses especially cancer-related researches such as the brain, breast, and liver. The usage of computer image processing technologies in medical diagnosis resulted in the use of some techniques for human body organs photos. MRI; is one of the most widely used technologies 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].

Various classification techniques are used in classifying tumors 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 [7-10].

The classification techniques used mainly on prospecting the MR image and extract some statistical or texture from the medical image to learn from how interestingly discern medical images to eventually develop “knowledgeable ” computer systems[11][12]. 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 4.

Section 2 present a brief description of the TANNN, section 3 show the result of applying TANNN on one of the real brain diseases datasets in the world and finally section 4 state the final conclusion.

## **2. APPLYING DATA MINING TECHNIQUES ON REAL BRAIN TUMORS DATASET**

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 [2].

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 [2].

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.

### **2.1 TANNN**

Most of Data mining techniques used mainly to kinds of classification methods which are:

1. Statistical methods
2. Data comparisons methods

Statistical methods such as Naïve Bayes, SVM and Discriminative Analysis; these methods mainly are very complex in applying and in computation complexity.

Data comparisons method such as Decision Tree, Nearest Neighbor, and Neural Network; these methods are considered to be simple but at the same time it's consumed a lot of time which increase the time complexity in these algorithms.

In comparison with these two methods it's clear that both of these methods are complex so to get a full classification technique a merge in both methods will be more benefit ; it will be faster (compared with the data comparisons methods) and also it will be lower in computation complexity (in comparison with the statistical methods).

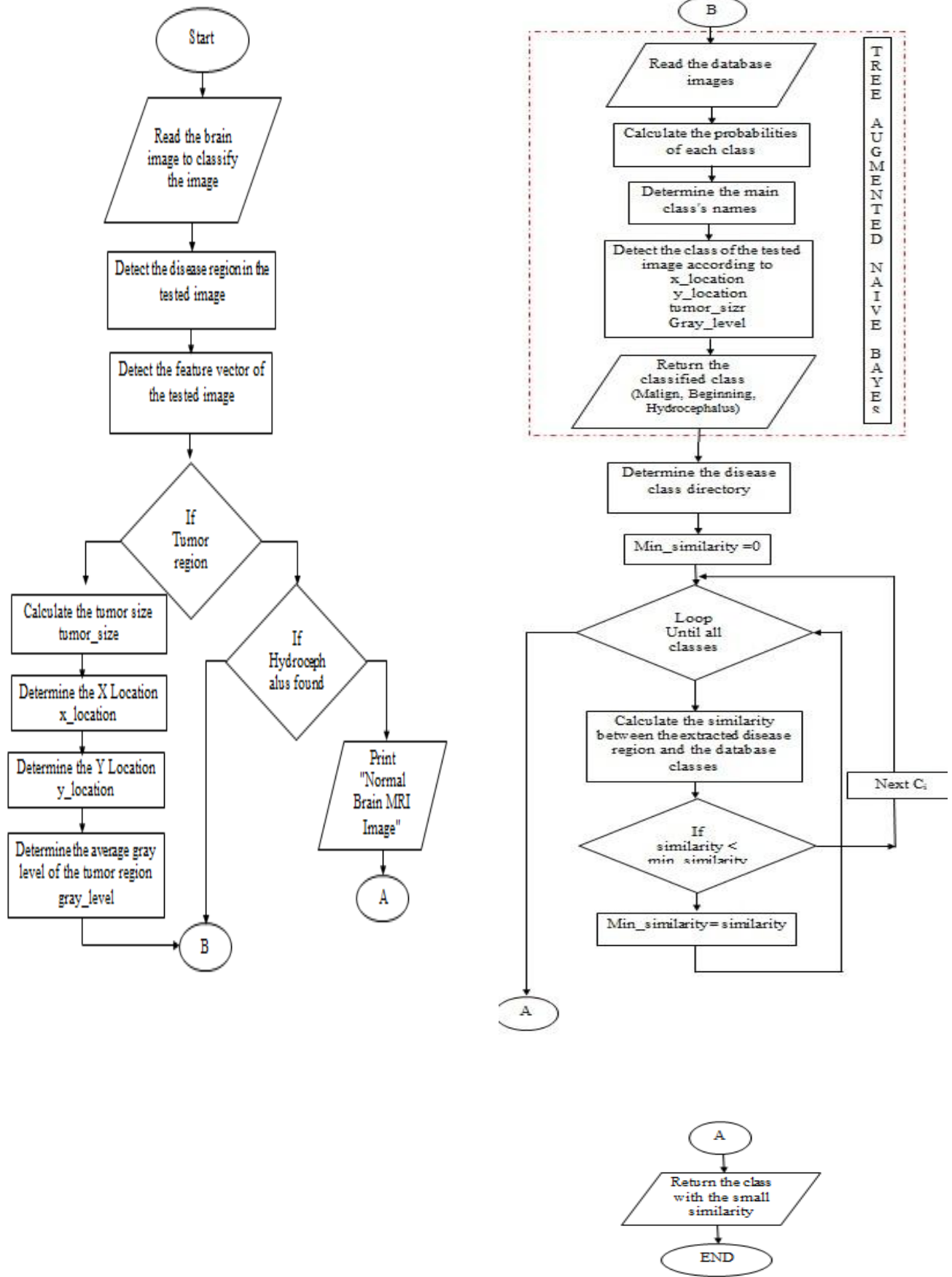


Fig 3 shows the flow chart of TANN algorithm.

### 3. 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.

#### 3.1 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.

Here we apply the proposed algorithm on one of the real MRI brain tumors datasets called BRATS []. Also, we can check the proposed system upon more than three datasets.

##### 3.1.1 BRATS Dataset [13]

MRI dataset consists of 811 Children MRI jpg format images, with 750 images containing malign tumor and 50 images containing benign tumor and 11 normal images. These images are divided into three categories, 70% for training, 15% for testing and 15% for validation.

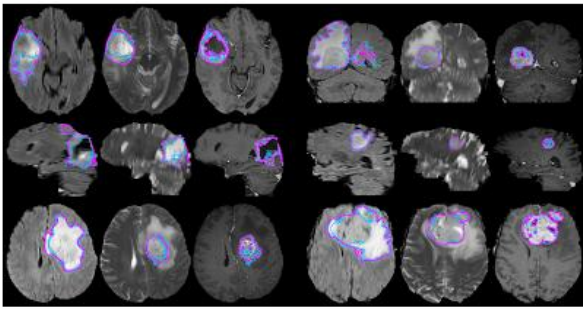


Fig.4. Samples of children brain dataset containing tumors BRATS dataset images [13]

Label	Project	Date	Subject	W/F	Age	Type	Scanner	Scans
1	IGT_OLIGOMA	11/01/2005	case01	M	7	IGT_OLIGOMA		3D SPGR VOLUME(4), HEME SUSCEPTIBILITY(1), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(3), T1 SPIN ECHO(2)
2	IGT_OLIGOMA	14/01/2005	case02	M	12	IGT_OLIGOMA		3D SPGR VOLUME(4), HEME SUSCEPTIBILITY(1), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(4), T1 SPIN ECHO(8)
3	IGT_OLIGOMA	18/01/2005	case03	F	8	IGT_OLIGOMA		3D SPGR VOLUME(1), HEME SUSCEPTIBILITY(1), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(3), T1 SPIN ECHO(7)
4	IGT_OLIGOMA	07/02/2005	case04	F	14	IGT_OLIGOMA		3D SPGR POST(1), T1 SAGITTAL(1)
5	IGT_OLIGOMA	11/02/2005	case05	M	16	IGT_OLIGOMA		3D SPGR VOLUME(4), HEME SUSCEPTIBILITY(1), LOCALIZER(2), SCAN 1 T2 SCAN 2 PD(3), T1 SPIN ECHO(1)
6	IGT_OLIGOMA	24/05/2005	case06	M	9	IGT_OLIGOMA		3D SPGR VOLUME(4), HEME SUSCEPTIBILITY(1), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(8), T1 SPIN ECHO(2)
7	IGT_OLIGOMA	10/06/2005	case07	F		IGT_OLIGOMA		3D SPGR VOLUME(4), HEME SUSCEPTIBILITY(1), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(7), T1 SPIN ECHO(8)
8	IGT_OLIGOMA	22/07/2005	case08	M		IGT_OLIGOMA		HEME SUSCEPTIBILITY(1), LOCALIZER(2), SCAN 1 T2 SCAN 2 PD(7), T1 SPIN ECHO(3)
9	IGT_OLIGOMA	25/07/2005	case09	M		IGT_OLIGOMA		HEME SUSCEPTIBILITY(1), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(8), T1 SPIN ECHO(2)
10	IGT_OLIGOMA	16/08/2005	case10	F		IGT_OLIGOMA		3D SPGR VOLUME(4), HEME SUSCEPTIBILITY(1), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(10), T1 SPIN ECHO(2)
11	IGT_OLIGOMA	19/09/2005	case11	F		IGT_OLIGOMA		3D SPGR VOLUME(4), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(4), T1 SPIN ECHO(1)
12	IGT_OLIGOMA	07/10/2005	case12	F		IGT_OLIGOMA		3D SPGR VOLUME(4), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(4), T1 SPIN ECHO(1)
13	IGT_OLIGOMA	07/10/2005	case13	F		IGT_OLIGOMA		3D SPGR VOLUME(4), HEME SUSCEPTIBILITY(2), LOCALIZER(1), SCAN 1 T2 SCAN 2 PD(3), T1 SPIN ECHO(2)

Fig.5. Samples of BRATS dataset data [13]

##### 3.1.2 NBTR Dataset [14]

The information collected by the NBTR covers both benign and malignant primary brain cancer, arising in the central nervous system (CNS), the skull base and pituitary gland. Data is collected under the permissions granted to the English Cancer Registries by the National Information Governance Board.

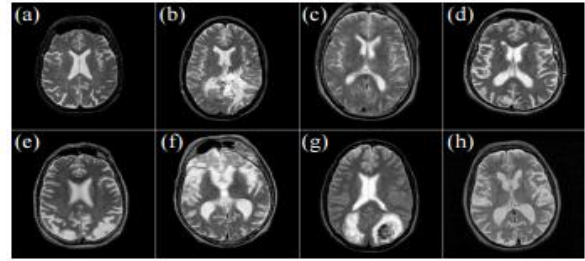


Fig.6. Different Samples of brain MRI with different kinds of tumors [14]

##### 3.1.3 OASIS Dataset [15]

OASIS is made available by Dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN). OASIS dataset contains 416 images from infants to 18 years old ages.

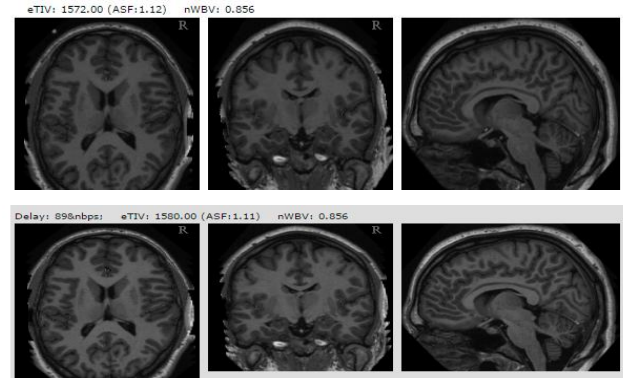


Fig.7. Samples of children brain dataset containing tumors OASIS dataset images [15]

OASIS Home | Data access tools | Contact | About

MRI Reliability data across the adult lifespan

Search >>>

MR Session	Subject	W/F	Hand	Age	eTIV	nWBV	ASF	Delay	Test/Retest	Scans	
1	OAS1_0061_MR1	OAS1_0061	F	R	20	1768	0.86	1.4434	TEST	repr-1 repr-2 repr-3 repr-4	
2	OAS1_0061_MR2	OAS1_0061	F	R	20	1767	0.86	0.999	1	RETEST	repr-1 repr-2 repr-3 repr-4
3	OAS1_0080_MR1	OAS1_0080	F	R	25	2620	0.87	1.0777	TEST	repr-1 repr-2 repr-3 repr-4	
4	OAS1_0080_MR2	OAS1_0080	F	R	25	2625	0.86	1.203	20	RETEST	repr-1 repr-2 repr-3 repr-4
5	OAS1_0092_MR1	OAS1_0092	M	R	22	1442	0.84	1.2169	TEST	repr-1 repr-2 repr-3 repr-4	
6	OAS1_0092_MR2	OAS1_0092	M	R	22	1457	0.85	1.1205	5	RETEST	repr-1 repr-2 repr-3 repr-4
7	OAS1_0101_MR1	OAS1_0101	M	R	29	2486	0.80	1.0307	TEST	repr-1 repr-2 repr-3 repr-4	
8	OAS1_0101_MR2	OAS1_0101	M	R	29	2502	0.85	1.169	64	RETEST	repr-1 repr-2 repr-3 repr-4
9	OAS1_0111_MR1	OAS1_0111	M	R	23	1711	0.85	1.0205	TEST	repr-1 repr-2 repr-3 repr-4	
10	OAS1_0111_MR2	OAS1_0111	M	R	23	1714	0.81	1.024	2	RETEST	repr-1 repr-2 repr-3 repr-4
11	OAS1_0117_MR1	OAS1_0117	M	R	25	1769	0.83	0.996	TEST	repr-1 repr-2 repr-3 repr-4	
12	OAS1_0117_MR2	OAS1_0117	M	R	25	1763	0.82	1.021	5	RETEST	repr-1 repr-2 repr-3 repr-4
13	OAS1_0145_MR1	OAS1_0145	M	R	34	1832	0.82	1.0619	TEST	repr-1 repr-2 repr-3 repr-4	
14	OAS1_0145_MR2	OAS1_0145	M	R	34	1834	0.82	1.061	10	RETEST	repr-1 repr-2 repr-3 repr-4
15	OAS1_0156_MR1	OAS1_0156	F	R	20	1510	0.79	1.042	TEST	repr-1 repr-2 repr-3 repr-4	
16	OAS1_0156_MR2	OAS1_0156	F	R	20	1506	0.76	1.045	1	RETEST	repr-1 repr-2 repr-3 repr-4
17	OAS1_0156_MR1	OAS1_0156	F	R	20	1591	0.84	1.0302	TEST	repr-1 repr-2 repr-3 repr-4	
18	OAS1_0156_MR2	OAS1_0156	F	R	20	1577	0.82	1.113	12	RETEST	repr-1 repr-2 repr-3 repr-4
19	OAS1_0191_MR1	OAS1_0191	F	R	21	1421	0.85	0.9246	TEST	repr-1 repr-2 repr-3 repr-4	
20	OAS1_0191_MR2	OAS1_0191	F	R	21	1426	0.87	1.219	23	RETEST	repr-1 repr-2 repr-3 repr-4
21	OAS1_0202_MR1	OAS1_0202	F	R	23	1574	0.85	1.1151	TEST	repr-1 repr-2 repr-3 repr-4	
22	OAS1_0202_MR2	OAS1_0202	F	R	23	1540	0.81	1.124	21	RETEST	repr-1 repr-2 repr-3 repr-4

Fig.8. Samples of OASIS dataset data [15]

### 3.2 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 becomes very easy locate it and extract it from MRI image.

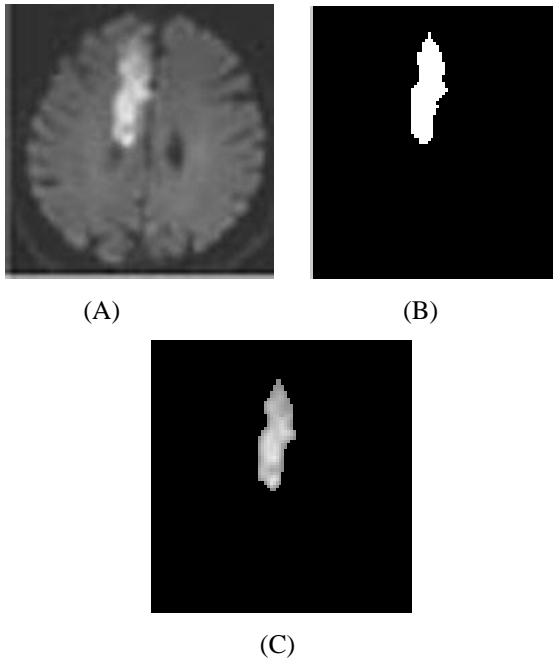


Fig.9. (A): MRI image of tumor affected brain gray scale image. (B): Threshold segmented image. (C): Final Extracted brain tumor from MRI image.

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.



Fig.7. Average value of  $\alpha$  and  $\beta$  against the number of clusters.

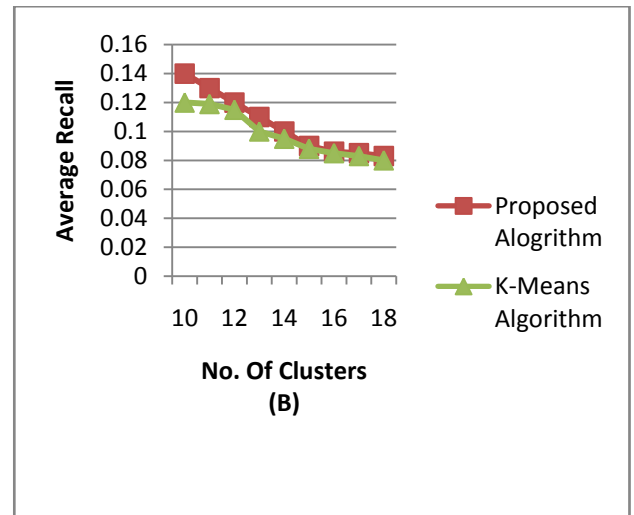


Fig.8. Average Precision and Recall against the number of clusters.

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

Table1. The Accuracy of the Classification Techniques

Segmentation	DA	NN	NB	SVM	DT	KN N	TANN N
BRATS	75.93	91.44	76.08	92.59	87.04	82.3	99.4
NBTR	90.12	96.1	93.52	92.59	96.19	93.7	99.8
OASIS	94.4	93.6	95.2	92.5	96.4	96.6	99.2

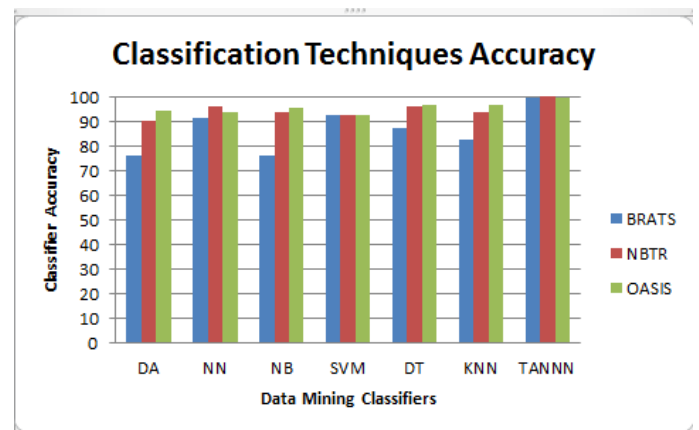


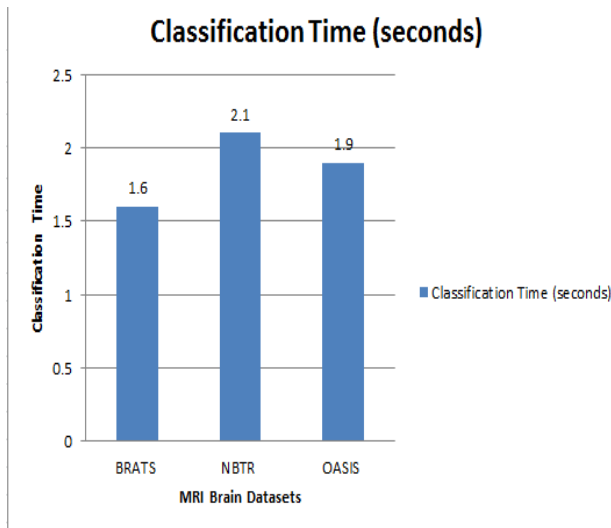
Fig.9. The accuracy of classification techniques.



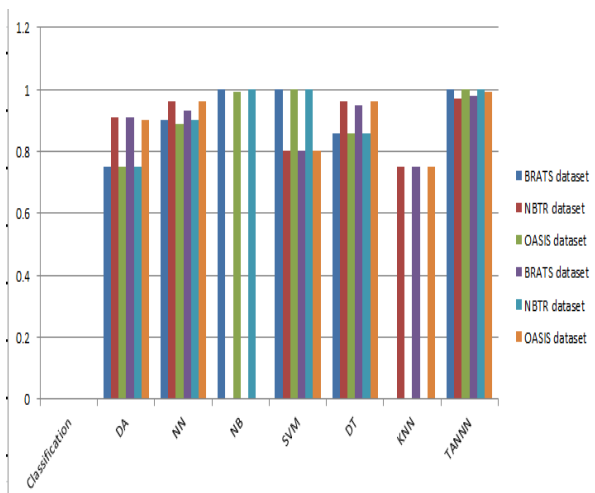
The next table states a comparison between the classification techniques according to the classification time see table 2.

**Table2. The classification Techniques Classification Time.**

Classification Technique	Classification Time (seconds)
BRATS dataset	1.6
NBTR dataset	2.1
OASIS dataset	1.9



**Fig.11. A comparison between classification techniques according to classification time.**



**Fig.12. Classification results in terms of specificity and sensitivity.**

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, TANNN appears to be the best suited for real-time classification tasks due to its relatively fast classification speed and high detection rate.

#### 4. 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.

#### 5. ACKNOWLEDGMENT

I am very grateful and would like to thank my guides Prof. Mohammed Haggag and Prof. Ahmed Farag for their advice and continued support. Without them, it would not have been possible for me to complete this paper. I would like to thank my husband for the thoughtful and mind stimulating discussion we had, which prompted us to think beyond the obvious.

#### 6. REFERENCES

- [1] E. M. Ali, A. F. Seddik. M. H. Haggag, "Using Data Mining Techniques for children Brain Tumors classification based on MRI", International Journal of Computer applications , Pp. 36-42, Vol. 131 , No. 2 , December 2015.
- [2] E. M. Ali, A. F. Seddik. M. H. Haggag, "Classification of Hydrocephalus using TAN", International Journal of Advanced Research in Computer Science and Software Engineering, Pp. 90-97, Vol. 5 , Issue. 11, November 2015.
- [3] A. Rajkumar, "A Multi- Stage Hybrid , CAD Approach for MRI Brain Tumor Recognition and Classification",

The IIOAB Journal School of Computing Science and Engineering VIT University, Januar 2016, India.

- [4] S. Ganesh, "A Comparative Study on Various Brain Tumor Classification Methods", *India Journal of Engineering* , Vol. 13 , Pp. 27-33, January 2016.
- [5] [5] K. Sakthivel, B. Swathi, S. Vishnu, "Analysis of Medical Image Processing and it's Application in healthcare", *International Journal od Advanced Engineering Research and Science (IJAERS)*, Vol. 3 Issue 2, Feb. 2016.
- [6] Y. Li , Y. Mingquan, Z. Hao, "Tumor Diagnosis Based on the GMM Feature Decision Classification of Brain MR Images", *International journal of Multimedia and Ubiquitous Engineering* , Vol. 11 No. 3 , Pp. 37- 44, 2016.
- [7] K. Mayuri, R. Khode, S. Salwe, "A review on efficient Brain Tumor Detection Using Various Methods", *International Journal of Research in Advent Technology* , Vol. 4 , No. 2 , February 2016.
- [8] L. Hou, D. Samaras, T. Kurc , " Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification", *ARXIV* , March 2016.
- [9] N. Kasat, S. Thepade, "Novel Content Based Image Classification Method Using LBG Vector Quantization Method with Bayes and Lazy Family Data Mining Classifier", 7<sup>th</sup> international conference on Communication , Computing and Virtualization, Pp. 483 – 489, 2016.
- [10] G. Santhosh , K. Sivanaruleselvan, P. Betty, " Survey on Brain Tumour Detection and Classification Using Image Processing", *ELK ASIA Pacific Journal of Computer Science and Information Systems*, Vol. 2, Issue. 1, 2016.
- [11] A. Mali, S. Pawar, "Detection & Classification of Brain Tumor", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol 4, Issue 1 , January 2016.
- [12] K. Selva , P. Geetha, "Semantic Feature Based Classification of Brain MRI using PCA and PNN", *International Conference on Electrical , Electronics, and Optimization Techniques (ICEEOT)*, 2016.
- [13] BRATS brain tumor dataset, [www.brain-tumor-segmentation.org](http://www.brain-tumor-segmentation.org)
- [14] NBTR brain tumor dataset , [www.nbtr.nhs.uk](http://www.nbtr.nhs.uk)
- [15] OASIS brain tumor dataset, [www.oasis-brains.org/](http://www.oasis-brains.org/)
- [16] ADNI brain tumor dataset, [www.adni.loni.ucla.edu/](http://www.adni.loni.ucla.edu/)