BRAIN TUMOR CLASSIFICATION BASED ON CLUSTERED DISCRETE COSINE TRANSFORM IN COMPRESSED DOMAIN

[1]Anitha, V. and [2]S. Murugavalli

[1]Department of Information Technology, Sri Muthukumaran Institute of Technology,
Chikkarayapuram, Near Mangadu, Chennai-69, Tamilnadu, India
[2]Department of Computer Science and Engineering,
Panimalar Engineering College, Chennai-602103, Tamilnadu, India

ABSTRACT

This study presents a novel method to classify the brain tumors by means of efficient and integrated methods so as to increase the classification accuracy. In conventional systems, the problem being the same to extract the feature sets from the database and classify tumors based on the features sets. The main idea in plethora of earlier researches related to any classification method is to increase the classification accuracy.The actual need is to achieve a better accuracy in classification, by extracting more relevant feature sets after dimensionality reduction. There exists a trade-off between accuracy and the number of feature sets. Hence the focus in this study is to implement Discrete Cosine Transform (DCT) on the brain tumor images for various classes. Using DCT, by itself, it offers a fair dimension reduction in feature sets.Later on, sequentially K-means algorithm is applied on DCT coefficients to cluster the feature sets. These cluster information are considered as refined feature sets and classified using Support Vector Machine (SVM) is proposed in this study. This method of using DCT helps to adjust and vary the performance of classification based on the count of the DCT coefficients taken into account. There exists a good demand for an automatic classification of brain tumors which grealtly helps in the process of diagnosis. In this novel work, an average of 97% and a maximum of 100% classification accuracy has been achieved. This research is basically aiming and opening a new way of classification under compressed domain. Hence this study may be highly suitable for diagnosing under mobile computing and internet based medical diagnosis.

Keywords: Discrete Cosine Transform (DCT), Support Vector Machine (SVM), Magnetic Resonance Image (MRI)

1.INTRODUCTION

The major task being the classification, it is a foremost role to distinguish the images based on the image content. The differences among all types of healthy and tumor images would be very less and highly challenging to classify. Tumor images are acquired based on a very critical imaging technique subsequently followed by enormous mathematical calculations before it is formed as an image. Typically the images obtained are nothing but the energy emitted by the cells or tissues present in the exposed test area.

So abnormalities in the brain images would not reflect properly in the distinguished intensity levels.

Although MR Images can show the location and size of tumors, it may not be able to classify the types of tumors. So an improved wavelet based classification had been done in Meenakshi and Anandhakumar (2013) and the results were verified with different types of tumors.

The local binary patterns and gray level co-occurrence features, gray level and wavelet features are extracted and these features are trained and classified using Support vector machine classifier in the works done by Kumar and Kumar (2014).

Corresponding Author: Anitha, V., Department of Information Technology, Sri Muthukumaran Institute of Technology,
Anna University, Chikkarayapuram, Near Mangadu, Chennai-69, Tamilnadu, India

Principal component analysis had been done in Othman and Basri (2011), after which the classification was carried out using probabilistic neural network. In their work, it was observed that the PNN classifier showed good accuracy along with less training time. Moreover PNN offers good robustness for changes in weights andhas very negligible retraining time. Based on the spread values, accuracies were obtained in the range of 73% to a maximum of 100%. Magnetic Resonance Spectroscopy (MRS) is also a most relevant data through which tumors can be detected and classified. Arizmendi *et al.* (2011) methodology, they proved that MRS data could be used and a combination of DWT and PCA would give fair classification results.

Nuero fuzzy methods had been used by Murugavalli and Rajamani (2007) to detect the tumor through MRI data. In specific, hierarchical self organizing map and Fuzzy C menas had been used to classify the image layer by layer.

Nanthagopal and Sukanesh (2013) had proposed to use wavelet based statistical approach to extract the feature vectors and classify the brain tumor using MRI images with SVM. They had performed classification with a data base size of 108 slices.

Sridhar and Krishna (2013), involves the usage of DCT to extract the features and probabilistic neural network to classify the images. Through the works done in, Zhang *et al.* (2011) relevant authors achieved 0.0451 seconds for extracting feature vectors from each image using wavelet transforms, PCA and neural network. Zhang *et al.* (2013), used wavelet transforms and SVM for an efficient classification of brain tumors. Ortiz *et al.* (2012), proposed Self Organized Mapping (SOM) based on neural networks and Genetic Algorithm (GA) had been used to select the features sets. Loukas *et al.* (2013) employed an integrated procedure consisting three pattern recognition algorithms (K-NN, SVM and PNN) for classifying and to characterize the breast cancer.Diabetes diagnosis had been with an accuracy of 93.2% as given by Purnami *et al.* (2009). wavelet features have been considered in the works done by Sebri *et al.* (2007). Through the earlier works discussed above, clearly emphasize process flow to obtain the results better as given in **Fig. 1**. Resizing the data base images to a common size is essential to fix the number of features obtained in all images to a common count irrespective of the size of the images. This common size will help while comparing the feature sets kept in feature data base one to one, otherwise there would be a dimensional mismatch.Anyhow, it is not advisable to resize the images to too less size. But, resizing allows a data loss, hence a trade off is always maintained to preserve the accuracy level.
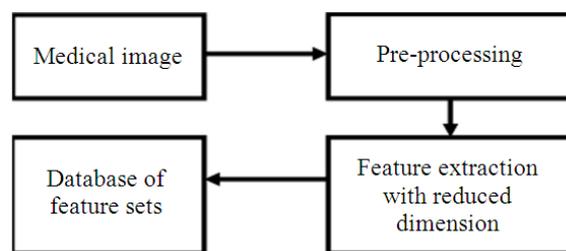
Fig. 1. General schematic of feature extraction

Reducing the dimensions of feature sets is also based on the eagerness to achieve a fair accuracy with less number of mathematical operations when a large data base of images is presented.

This study has been organized as follows. Section 2 describes the methodology, description of classification system and describes the Pseudo codes of classification system. In section 3 and 4 Results obtained and discussion has been presented.

2.MATERIALS AND METHODS

2.1. Discrete Cosine Transform (DCT)

A Discrete Cosine Transform (DCT) of a time domain signal is a sum of cosine functions oscillating at different frequencies. The authors Krishneswari and Arumugam (2012) hadexplained the method of applying DCT for feature extraction for biometric applications and essence has been presented below in order to obtain a good idea about DCT. The DCT is almost similar to the Discrete Fourier Transform (DFT), except that in DCT, more energy is concentrated in the lower order coefficients. For a 2D image data, the DCT is purely real (only magnitude). While performing a DCT operation, on a square matrix of pixels produces coefficients which are similar to the frequency domain coefficients produced by a DFT operation. An N-point DCT is closely related to a 2N-point DFT. The N frequencies of a 2N point DFT correspond to N points on the upper half of the unit circle in the complex plane. The DCT of an image is typically given by Equation 1:

$$X_{k1,k2} = \sum_{n_1=0}^{N_1=1} \left(\sum_{n_2=0}^{N_2=1} X_{k1,k2} \cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2 \right] \right)$$

$$\cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_1 \right] = \sum_{n_1=0}^{N_1=1}\sum_{n_2=0}^{N_2=1} X_{n1,n2} \quad (1)$$

$$\cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_2 \right]\cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2 \right]$$

where, $x_{n1, n2}$ represents the pixel value at n_1 and n_2. N_1 and N_2 represents the total number of rows and columns taken into consideration while applying DCT and $k_1 = 0...N_1.1$ and $k_2 = 0.N_2-1$.

For 2D images, after transformation, the major amount of signal energy is held by just a few lower order DCT coefficients. Quantizing these DCT coefficients would be based on the energy content in them. Such lower order coefficients can be more quantized with full more number of bits than the higher order coefficients. In addition to this dimension reduction, higher order coefficients after a threshold may be quantized to 0. When an input image is divided into several horizontal and vertical blocks, with each block sizing 8×8. So, in this case, N_1 and N_2 are typically 8 (after dividing the input image into several horizontal and vertical blocks) and the DCT is applied to each block of size 8×8. The result is an 8×8 matrix of transformed coefficients in which the top left element (1,1) shown in **Fig. 2** is the DC (Zero frequency) component and entries with increasing vertical and horizontal index values represent higher vertical and horizontal spatial frequencies. The DCT coefficients are reordered by the zig-zag scan. This method of scanning reorders all the DCT coefficients from low to high frequencies. In case of image processing, neglecting the higher order frequencies by quantizing with zero bits will not affect the appearance of the images. Human eyes are more perceivable to the low frequency components. Hence a good dimensionality reduction is achieved with DCT stage itself by discarding the higher order frequencies.

Zig-zag scanning process is performed on DCT coefficients shown in **Fig. 2** after zig-zag scanning, (1×64) vector is produced. This vector contains coefficients of DCT block from low frequency to high frequency. Low-pass, Band-pass, High-pass filters are applied to this vector for constructing a feature vector by considering the whole spectrum range of the sub block.

2.2. K-Means Algorithm

The simplest way to explain K-means algorithm is as follows. Let us assume that given a dataset of n data points $x_1, x_2,, ..., x_n$ such that each data point is in R_d. The major aim will be to group the given data into k clusters. Each of the K clusters will contain certain number of elements. The k points $\{m_j\}$ (j = 1, 2, ..., k) in R_d such d that is minimized, where $d(x_i, m_j)$ denotes the Euclidean distance between x_i and m_j. which should have the minimum variance. The points $\{m_j\}$ where j = 1, 2, ..., k are known as cluster centroids. The problem

in Equation 2 is to find k cluster centroids, such that the average squared euclidean distance or Mean Squared Error (MSE) between a data point and its nearest cluster centroid is minimized.

$$\frac{1}{n}\sum_{i=1}^{n}[\min_j d^2(x_i, m_j)] \qquad (2)$$

The K-means algorithm can be understood as a gradient descent procedure, updates the cluster centroids to minimize the objective function in Equation 2. The K-means always converge to a local minimum. Velmurugan and Santhanam (2010) found the particular local minimum found on the starting cluster centroids.

2.3. Support Vector Machines (SVM)

The basic idea behind the SVM classification technique is to identify the class of the input test vectors. This is a supervised learning algorithm, where the training vectors are used to train the system to map these training vectors in a space with clear gaps between them using some standard kernel functions and the input test vectors are mapped on to the same space to predicted as given in Bharathi and Natarajan (2011), the possible class. The linear kernel scenario is shown in **Fig. 3**.

Given some training data D, a set of n points of the form where the class y_i is either belonging to the class 1 or class-1, indicating the class to which the point Xi belongs. Each Xi is assumed to be a p-dimensional real vector. Here it is needed to find the maximum-margin hyper plane that divide the points having $y_i = 1$ from those having $y_i = -1$.
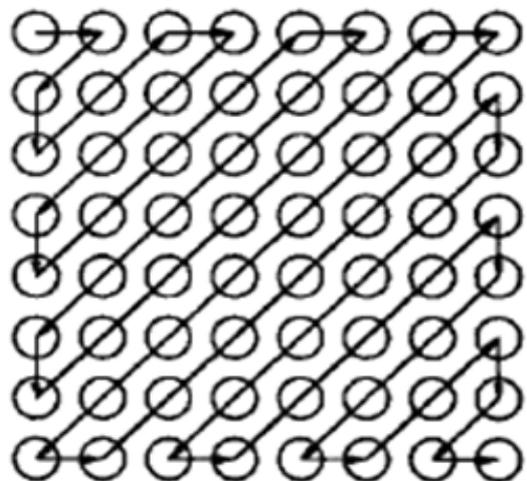
Fig. 2. 8×8 matrix DCT coefficients
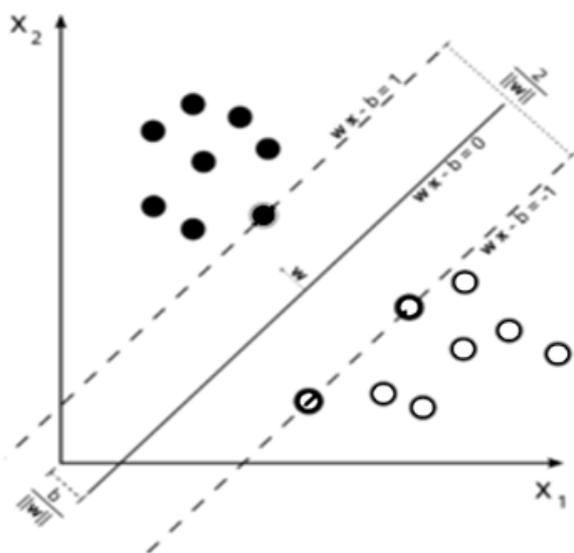
Fig. 3. SVM scenario

Samples on the margin are called the support vectors. w.x-b = 0 wheredenotes the dot product andis the normal vector to the hyper plane. The parameter b/||w|| determines the offset of the hyper plane from the origin along the normal vector. If the training data are linearly separable, then two hyper planes can be selected in such a way that they separate the data and there are no points between them and then tried to maximize their distance. The region bounded by them is called "the margin". These hyper planes can be described by the equation:

$$w.x - b = \text{ and } w.x - b = -1$$

At the testing phase, the data points $x_i$are separated using the following constraints w.x$_i$-b≥1 for x_i of thefirst class or w.x$_i$-b≤1 for $x_i$of the second class.

The discussion is all about a two class SVM. In our proposedmethod, the target is a multiclass SVM where the number of classes is more than 2. A conventional two class SVMcan be extended into a multiclass SVM, by two methods. (1) By comparing one feature sets with all other remaining feature sets as in another class (one-versus-all). (2) Between every pair of classes (one-versus-one). SVM classification methods have been well discussed as given in Yogameena *et al*. (2010).

2.4. Description of Classification System

The Tumor images were obtained from Harvard Medical School (Web: http://www.med.harvard.edu/

aanlib/home.html) database. The brain datasets consists of 102 T2-weighted MR brain images in axial plane and 256×256 in-plane resolution, were downloaded. A total of 102 images from 5 different classes like Normal, Glioma, Meningioma, Alzhimer and Alzhimer with visual agnosia. The typical size of the resized images in our case is 150×150. Resizing is done to reduce the time processing **Fig. 4**.

The preprocessed images were decomposed into its discrete cosine terms. The number of coefficients were taken as option to decide the accuracy of theclassification system.The cluster counts and cluster centers are considered as feature elements representing an image.The training sets with such 40 vectors are chosen and trained with SVM. Reducing the dimension of the feature vectors is a common technique to reduce the processing time. Anyhow, in our classification, dimension reduction is achieved using DCT as well as using a novel method of feature vectors, being the element counts and centroids of each clusteras features.

The schematic shown in **Fig. 5** shows the typical image process for one image. The down arrows represents the contents of the corresponding block. All the processes are carried out by individual block of size 8×8 basis in our implementation.

2.5. Pseudo CodesofClassification System

Step1:   Collecting the data base.
Step2:   Preprocessing (This includes resizing the image and noise elimination)
Step3:   Dividing the images into horizontal and vertical blocks of size 8×8. And applying pass band DCT for every 8×8 window
Step4:   DCT coefficients are clustered using k-means algorithm
Step4.1: Initial group centroids are placed into the2D space
Step4.2: Each object of the data set is placed into the group that has the closest centroid
Step4.3: Recalculate the positions of the centroids
Step4.4: If the positions of the centroids didn't change, advance to the next step, else go to Step4.2
Step4.5: End
Step5:   Constructing SVM using a default linear kernel
Step6:   Input the test MRI to the trained SVM and outputting the prediction
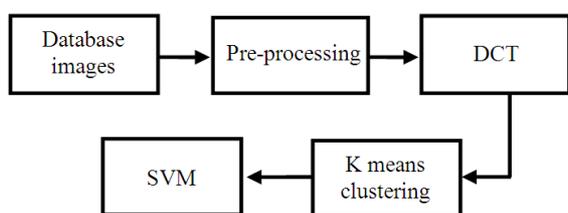
Fig. 4. Implementation of classification system
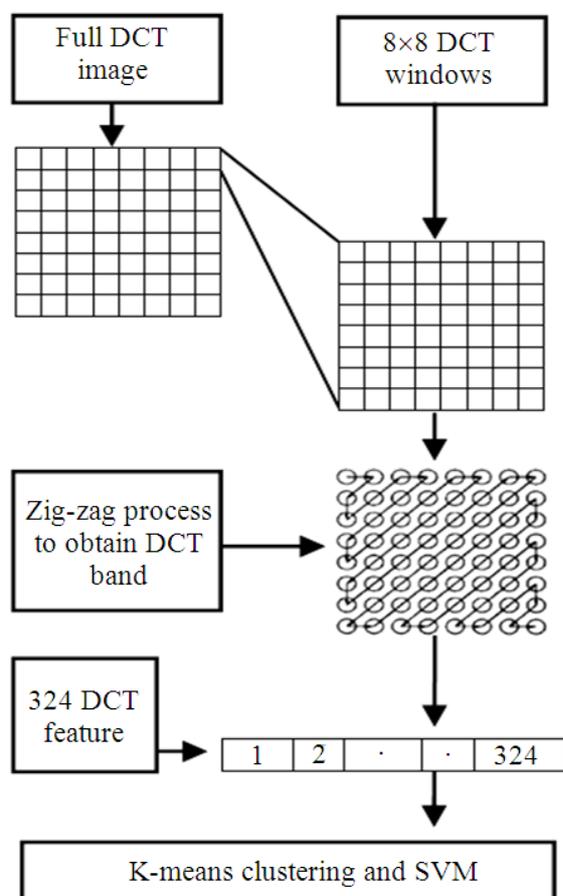
Fig. 5. Schematic of classification for single image

3. RESULTS

The experiments were conducted on the hardware platform of P4 IBM with 3.3 GHz processor and 2 GB RAM, running under Windows 7 operating system. The algorithm was developed as Matlab script using Image processing and Mathematical tool box. The test sets are chosen and classification accuracy is obtained as given in the **Table 1**. When a resized image of size 150×150 is divided into 8×8 blocks, we obtain 324 blocks (18 horizontal and 18 vertical blocks) which is a rounded number through matlab coding. So, 64 coefficients from each window are mapped into only one DCT coefficient, which leads to 324 coefficients for the whole image. Anyhow, number of coefficients taken from each window can be increased when the accuracy obtained is less.

MR images with various diseases as named by medical experts were framed as a final image data base for our classification. It is well known that, any individual brain disease type reflects in the appearance of image. Hence it is worth to comment, that the classification process would highly help in identifying the disease concerned with the input MR image **Fig. 6**.

Thus the classification had been done with a total of 40 features, out of 40 features, 20 features are centroids and 20 features are the number of elements in each cluster. The classification accuracy is obtained as 97.05%. As this is a new method of taking cluster centroids and number of elements in each cluster as features of each DCT image, there is no motive to compare with earlier methods. The values presented in **Table 1** are just the accuracy values obtained through our method which is fair.

4. DISCUSSION

The following **Table 2** shows the accuracy variations for different number of DCT coefficient and number of cluster.

From **Table 2** and **Fig. 7** it is known that, the classification accuracy is insensitive to the number of DCT coefficients. But it depends on the number of features, in turn it depends on number of clusters. It may be thought thathow classification accuracy is dependent of number of DCT coefficients, but a minimum accuracy of 96% is achieved with only one DCT coefficient per block against the earlier methods discussed in previous sections.

Figure 8 shows the performance increases linearly with an increase in the number of cluster features. But a better performance is arrived while a higher number of DCT coefficients are considered for classification. **Table 3** shows the comparison of classification accuracies of existing methods and our proposed methods. Our proposed method yields 100% accuracy while calculating at compressed domain. Hence this research work opens a new way of classificaion method for medical images.

Table 1. Classification accuracy against total number of images and number of images classified

Class number and nature	Total number of images	Number of images classified	Classification accuracy(%)
Class 1 (Normal)	23	22	95.65
Class 2 (Glioma)	18	17	94.40
Class 3 (Meningioma)	21	20	95.20
Class 4 (Alzheimer)	18	18	100.00
Class5 (Alzheimer plus agnosia)	21	21	100.00
All classes	102(when test set is equal to train set)	99	97.05
Average accuracy	-	-	97.00
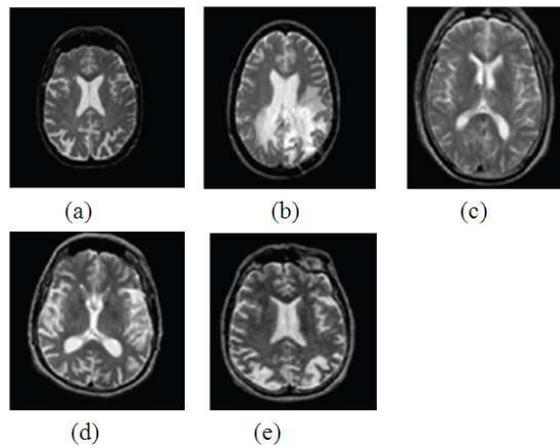
Fig. 6. Sample of brain MRIs,(a) Normal brain (b) Glioma (c) Meningioma (d) Alzheimer(e) Alzhiemer plus agnosia
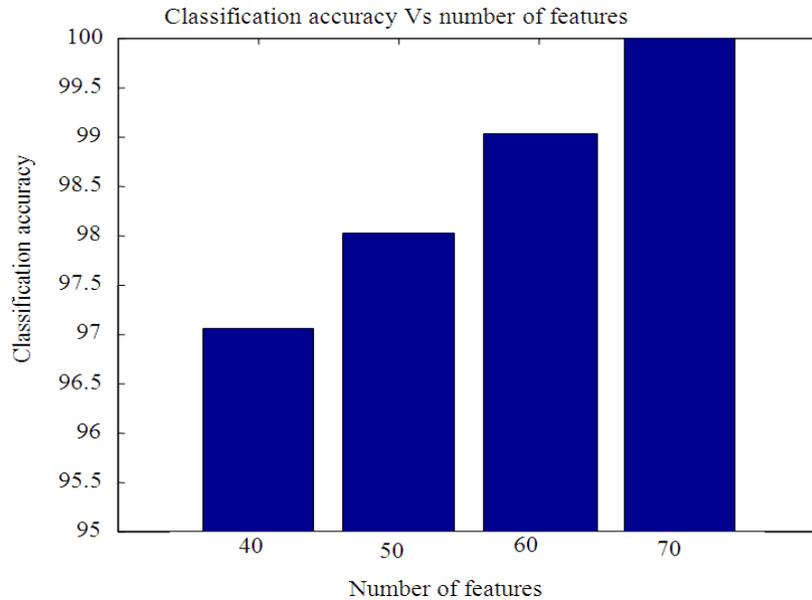
Fig. 7. Classification accuracy Vs Number of features
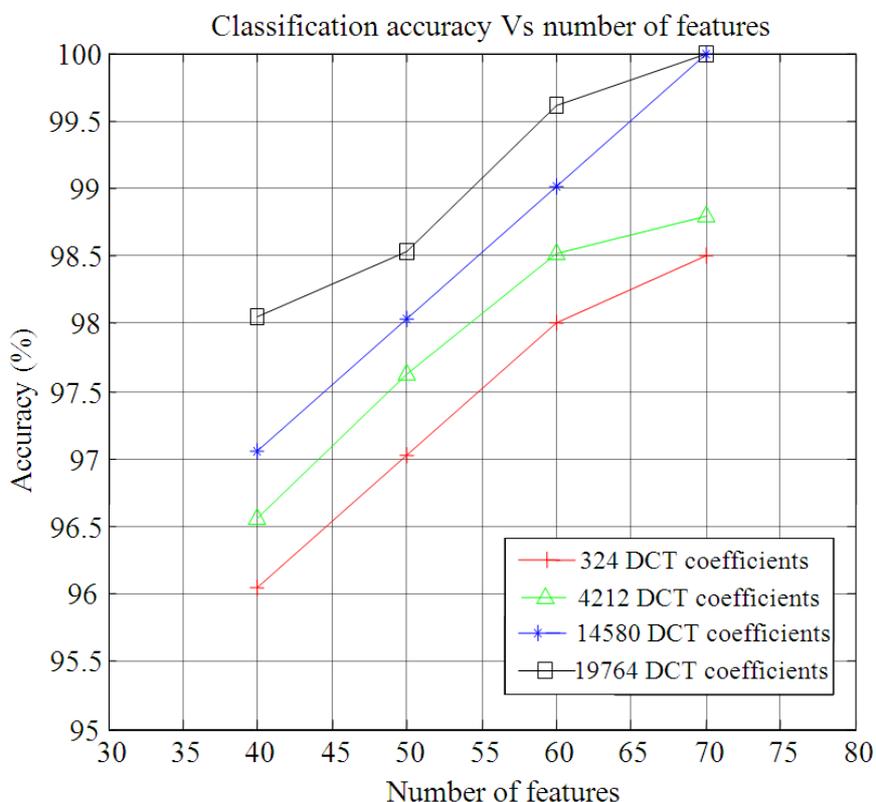
Fig. 8. Shows the performance increases linearly with an increase in the number of cluster features. But a better performance is arrived while a higher number of DCT coefficients are considered for classification

Table 2. Accuracy of proposed system

No of DCT coefficients per 8×8 window	Total number of DCT coefficients for one image	No of clusters (K)	Number of features	Over all classification accuracy (%)
1	324	20	40	96.0
		25	50	97.0
		30	60	98.0
		35	70	98.5
13	4212	20	40	96.5
		25	50	97.6
		30	60	98.5
		35	70	98.8
45	14580	20	40	97.0
		25	50	98.0
		30	60	99.0
		35	70	100.0
61	19764	20	40	98.0
		25	50	98.5
		30	60	99.6
		35	70	100.0

Table 3. Accuracy comparison

Methods classification accuracy (%)	
Co-occurrence texture features using bidirectional associative memory.	93.75%
Optimal dominant grey-level run length texture features using SVM classifier	95.80%
2D DWT and SVM	100.00%
Proposed method (DCT, K Means and SVM)	100.00%

5. CONCLUSION

In this study, a new brain tumor classification work had been done with dimensionality reduction with pass band DCT and k-means clustering combined with SVM classification. As discussed in previous sections many advantages are reaped. While designing a classfication algorithm, earlier works have never focused on the memory utilization. Though earlier works focus on speed of computations, it makes use of the full dimension of the actual image or their coefficients. In our study, an attempt has been done, to reduce the memory usage and hence the feature vectors are extracted in compressed domain. By reducing the number of DCT coefficients and k-means clustering, a two stage dimension reduction has been achieved. In our work the linear kernel has been chosen for SVM classification. The cluster centroids and number of elements obtained through the k-means fully depends on initial random seeds of centroid values. Besides this, while using k-means algorithm it introduces Not A Number (NAN) while clustering due to convergence problem. This can be avoided by setting input argument options for "kmeans" function given in matlab, while running the algorithm. As result of this NAN problem and random seed values, the number of elements in the same cluster and centroid values may differ when clustering is done at a different instant for the same DCT transformed image. But this randomness is bound to less than 5% noted from our observations and do not affect our classification results is our justification. In future, our work will be focused towards other clustering techniques and optimizing SVM parameters while using different kernel functions. Using such different kernel functions may accomadate even noisy images too. In such case, it is expected that classification performance is still improved even in case of visual and non visual changes in image characteristics. It is strongly believed by the authors of this study, that this method of classification would still become popular for any kind automatic classification applicable to other image data bases as well.

6. ACKNOWLEDGEMENT

7. REFERENCES

Arizmendi, C., A. Vellido and E. Romero, 2011. Binary classification of brain tumours using a discrete wavelet transform and energy criteria. Proceedings of the 2nd Latin American SymposiumCircuits and Systems, Feb. 23-25, IEEE Xplore Press, Bogata, pp: 1-4. DOI: 10.1109/LASCAS.2011.5750304

Bharathi, A. and A.M. Natarajan, 2011. Cancer classification using support vector machines and relevance vector machine based on analysis of variance features. J. Comput. Sci., 7: 1393-1399. DOI: 10.3844/jcssp.2011.1393.1399

Krishneswari, K. and S. Arumugam, 2012. Multimodal biometrics using feature fusion. J. Comput. Sci., 8: 431-435. DOI: 10.3844/jcssp.2012.431.435

Kumar, P.S. and P.G. Kumar, 2014. Performance analysis of brain tumor diagnosis based on soft computing techniques. Am. J. Applied Sci., 11: 329-336. DOI: 10.3844/ajassp.2014.329.33611

Loukas, C., S. Kostopoulos, A. Tanoglidi, D. Glotsos and C. Sfikas et al., 2013. Breast cancer characterization based on image classification of tissue sections visualized under low magnification. Comput. Math. Methods Med., 2013: 1-7. DOI: 10.1155/2013/829461

Meenakshi, R. and P. Anandhakumar, 2013. Wavelet statistical texture features with orthogonal operators tumour classification in magnetic resonance imaging brain. Am. J. Applied Sci., 10: 1154-1159. DOI: 10.3844/ajassp.2013.1154.1159

Murugavalli, S. and V. Rajamani, 2007. An improved implementation of brain tumor detection using segmentation based on neuro fuzzy technique. J. Comput. Sci., 3: 841-846. DOI: 10.3844/jcssp.2007.841.846

Nanthagopal, P. and A. Sukanesh, 2013. Wavelet statistical texture features-based segmentation and classification of brain computed tomography images. Proc. IET Image, 7: 25-32. DOI: 10.1049/iet-ipr.2012.0073

Ortiz, A., J.M. Gorriz, J. Ramirez and D. Salas-Gonzalez, 2012. Unsupervised neural techniques applied to MR brain image segmentation. Adv. Artificial Neural Syst., 2012: 1-7. DOI: 10.1155/2013/187074

Othman, M.F. and M.A.M. Basri, 2011. Probabilistic neural network for brain tumor classification. Proceedings of the 2nd International Conference Intelligent Systems Modelling Simulation, Jan. 25-27, IEEE Xplore Press, Kuala Lumpur, pp: 136-138. DOI: 10.1109/ISMS.2011.32

Purnami, S.W., A. Embong, J.M. Zain and S.P. Rahayu, 2009. A new smooth support vector machine and its applications in diabetes disease diagnosis. J. Comput. Sci., 5: 1003-1008. DOI: 10.3844/jcssp.2009.1003.1008

Sebri, A., J. Malek and R. Tourki, 2007. Automated breast cancer diagnosis based on GVF-snake segmentation, wavelet features extraction and neural network classification. J. Comput. Sci., 3: 600-607. DOI: 10.3844/jcssp.2007.600.607

Sridhar, D. and N.M. Krishna, 2013. Brain tumor classification using discrete cosine transform and probabilistic neural network. Proceedings of the International Conference Signal Processing Image Processing and Pattern Recognition, Feb. 7-8, IEEE Xplore Press, Coimbatore, pp: 1-5. DOI: 10.1109/ICSIPR.2013.6497966

Velmurugan, T. and T. Santhanam, 2010. Computational complexity between k-means and k-medoids clustering algorithms for normal and uniform distributions of data points. J. Comput. Sci., 6: 363-368. DOI: 10.3844/jcssp.2010.363.368

Yogameena, B., E. Komagal, M. Archana and S. Raju Abhaikumar, 2010. Support vector machine-based human behavior classification in crowd through projection and star skeletonization. J. Comput. Sci., 6: 1008-1013. DOI: 10.3844/jcssp.2010.1008.1013.

Zhang, Y., S. Wang, G. Ji and Z. Dong, 2013. An MR brain images classifier system via particle swarm optimization and kernel support vector machine. Sci. World J., 2013: 1-9. DOI: 10.1155/2013/130134

Zhang, Y., Z. Dong, L. Wu and S. Wang, 2011. A hybrid method for MRI brain image classification. Expert Syst. Applic., 38: 10049-10053. DOI: 10.1016/j.eswa.2011.02.01