Intelligent Methodology for Brain Tumors Classification in Magnetic Resonance Images

Heba Mohsen, El-Sayed A. El-Dahshan, El-Sayed M. El-Horbaty and Abdel-Badeeh M. Salem

***Abstract*—**Recently, a lot of researches have been made in the area of automatic detection and diagnosing the brain tumor type based on different medical imaging techniques. This paper presents a new intelligent methodology applying k-means segmentation technique and a hybrid support vector machine (SVM) classifier based on Linear-SVM and Multi-SVM using two feature extraction techniques, namely : Gray level co-occurrence matrix (GLCM) and discrete wavelet transform (DWT) followed by Principle component analysis (PCA) to detect brain tumors in brain magnetic resonance images (MRIs) and differentiate between three types of malignant brain tumors: glioblastoma, sarcoma and metastatic bronchogenic carcinoma. The results of the two feature extraction techniques were compared according to their accuracy, sensitivity and specificity showing good results and high robustness.

***Keywords*—**machine learning, support vector machine, k-means, discrete wavelet transform, Principle component analysis, Gray level co-occurrence matrix, medical informatics.

I. Introduction

Brain tumor is the abnormal growth of cells that serves no purpose. Brain tumors are classified into primary tumors which starts in the brain and usually does not spread to other parts of the body and secondary (metastatic) tumors which is formed by cancer cells from a primary cancer elsewhere in the body that have spread to the brain. In fact, Primary brain tumors may be benign or malignant unlike secondary tumors that are always malignant [1], [2]. Generally, brain tumors are graded from 1 to 4, according to their behavior and malignant tumors are either grade 3 or 4 because they are relatively growing fast and early detection and diagnosing can help in the treatment plan or reducing the chances of the tumor regrowing after surgical removing.

According to the Central Brain Tumor Registry of the United States (CBTRUS), Brain tumors are considered to be the third most common cancer occurring and causing of death among the young adults (ages 15-39). In 2017, they are expecting to diagnose more than 79,000 new cases of primary malignant and non-malignant brain and other CNS tumors in the United States where this estimation expects around 26,000 primary malignant and 53,000 benign cases [3].

Heba Mohsen is with the Faculty of Computers and Information Technology, Future University, Cairo, Egypt.

El-Sayed A. El-Dahshan is with the Faculty of Science, Ain Shams University, Cairo, Egypt and Egyptian E-Learning University, Giza, Egypt.

El-Sayed M. El-Horbaty is with the Faculty of Computers and Information Sciences, Ain Shams University, Cairo, Egypt.

Abdel-Badeeh M. Salem is with the Faculty of Computers and Information Sciences, Ain Shams University, Cairo, Egypt.

Automated brain tumor detection and identification is the field of interest in research nowadays using medical imaging techniques like Magnetic resonance imaging (MRI). MRI has an advantage that it provide very detailed diagnostic image for the brain soft tissues with high contrast that allows to identify specifically the abnormalities if exist [4].

In this paper we proposed a new intelligent methodology based on hybrid support vector machine (SVM) classifier to detect abnormalities in brain MR images and differentiate between three types of malignant tumors using two different feature extraction techniques and compare the results for each of them.

The paper is organized as follows: Section 2 contains the proposed intelligent methodology which consists of data acquisition, image segmentation, feature extraction and reduction and classification stages. In section 3 the experimental results are given. The conclusion and future work are given in section 4.

II. Related Work

K.G. Khambhata and S.R. Panchal [1] proposed a system of multi-stage for brain tumor diagnosis based on multiclass SVM classifier to classify 5 types of tumors and gave a results of 76.14% accuracy in Astrocytoma, 76.65% in Glioblastoma, 86.60% in Medulloblastoma, 84.26% in Meningioma and 82.23% in Metastatic Melanoma. The system includes a preprocessing stage based on image denoising, skull stripping and image enhancement, a feature extraction stage based on extracting texture, color, shape and intensity features and segmentation stage based on region, intensity, and clustering techniques.

A.N. Pathak, and R.K. Sunkaria [5] presented a system for multiclass classification of brain tumors of four stages: pre-processing of MR images, feature extraction based on DWT using Haar wavelet and PCA technique for feature reduction and finally the classification stage based on multiclass SVM. The system Results show overall 100% accuracy for the individual classes.

D.R. Nayak, R. Dash, and B. Majhi [6] presented a system for brain MR image classification to normal and abnormal images that consists of three stages: feature extraction using level-3 2D DWT that is normalized before the feature dimensionality reduction using probabilistic PCA (PPCA) and the classification based on the AdaBoost algorithm with random forests (ADBRF). The presented system reaches 100% accuracy.

E.A. El-Dahshan., H. Mohsen, K. Revett, and A.M. Salem

[7] proposed a system of four stages that uses the median filter for pre-processing, the feedback pulse coupled neural network (FPCNN) for image segmentation, 2D DWT for features extraction and the wavelet coefficients was reduced using PCA and the classification stage based on feed forward back propagation neural network (FP-ANN) to classify brain MR images as normal and abnormal. This system reached 99% classification accuracy.

M. Alfonse, and A. M. Salem [8] developed a system based on SVM classifier to classify the brain MR images to normal and abnormal images with accuracy 98.9%. The system consists of four stages: Image enhancement and cropping in the preprocessing stage, segmentation stage based on Expectation Maximization (EM) algorithm and adaptive thresholding, feature extraction stage using Fast Fourier Transform (FFT) and the feature was reduced using Minimal-Redundancy-Maximal-Relevance criterion (MRMR) and finally the classification stage.

III. Related Work

The proposed intelligent methodology is illustrated in Fig.1. The methodology consists of three stages: image segmentation stage based on k-means clustering technique, feature extraction and reduction stage based on two techniques: (a) gray level co-occurrence matrix (GLCM) and (b) discrete wavelet transform (DWT) integrated with principle component analysis (PCA) techniques and finally the classification stage based on linear-SVM classifier to detect the abnormal images followed by multi-SVM classifier to classify the images into four classes: normal, glioblastoma, sarcoma and metastatic bronchogenic carcinoma.

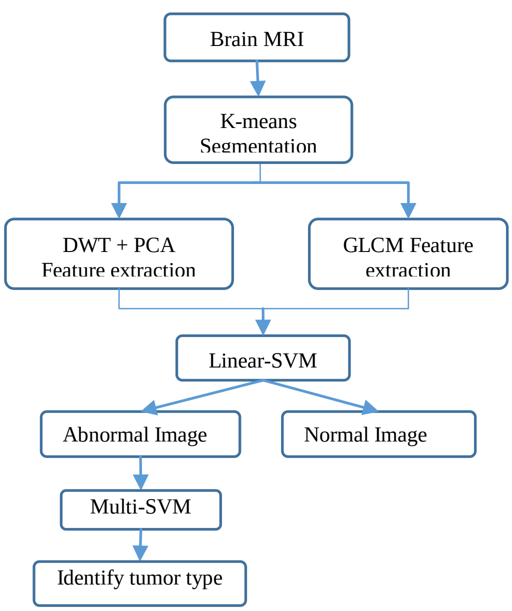


Fig. 1. The proposed intelligent methodology

A. Data Acquisition

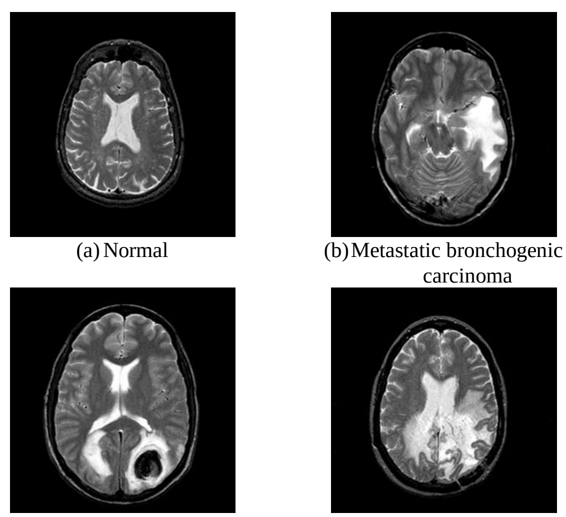
The dataset was collected from Harvard Medical School website (http:// med.harvard.edu/AANLIB/) [9]. It consists of 66 real human brain MRIs with 22 normal and 44 abnormal images which are glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumors. All the brain MRIs was in axial plane, T2-weighted and 256 × 256 pixel. A sample of the dataset is illustrated in Fig.2.

B. Image segmentation

The human brain segmentation consists of separating the different normal brain tissues such as gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) and the skull from the tumor tissues [10]. The clustering techniques suit the segmentation problem of brain MR images as it based on organizing the objects into groups of similar features, attributes and characteristics. Clustering techniques divided into supervised and unsupervised techniques which differ in the type of learning. In supervised techniques, the cluster criteria needed to be known where in the unsupervised techniques, only the number of clusters needed to be known to perform the clustering [7]-[9].

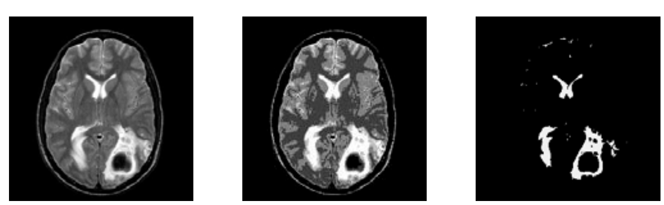
1) K-Means Clustering

K-means clustering is the simplest unsupervised clustering technique that can work for large number of variables and classifies the input data into multiple classes based on their inherent distance from each other. In k-means clustering, it clusters a given set of data using a certain number of classes based on the similarity between the given data and the classes' centers [7], [8], [10], [11]. Fig.3 illustrates brain MRI segmentation using k-means, the number of classes was given as 5.



(a) Normal (b) Metastatic bronchogenic carcinoma (c) Sarcoma (d) Glioblastoma

Fig. 2. Brain MRIs dataset sample



(a) Original image (b) Segmented Image using K-means (c) Segmented tumor using K-means

Fig. 3. Segmentation using K-means

C. Feature extraction and reduction

Features extraction is finding the most prominent features that can completely describe an image [16]. Feature extraction is a challenging task as the feature set extracted defines the accuracy of the classification. The proposed methodology is based on two different techniques: GLCM and DWT integrated with PCA to reduce the feature vector. The features extracted from each of the two techniques are then used in the training of the classifier separately in the classification stage.

1) Gray level co-occurrence matrix (GLCM)

Feature extraction using GLCM is a simple statistical method to extract textural features from the relationship of pixels compared to other techniques like wavelet transform. This statistical method proposed by Haralick et el. [17] uses the GLCM to extract certain features affecting to the spatial distribution of the grey levels in an image [13, 18, 19]. In this paper, the textural features of Haralick method calculated from the GLCM for each image after normalization in addition to some common intensity features used in researches [16], [19] for more accurate and robust classification.

The following 20 features extracted from the GLCM of each image was used to train the classifier in the classification stage: Angular second moment, Contrast, Correlation, Variance, Inverse Difference Moment (IDM), Sum variance, Sum average, Sum entropy, Entropy, Difference Variance, Difference entropy, Information measures of correlation, Homogeneity, Dissimilarity, Cluster Prominence , Cluster Shade, Smoothness, Mean, Skewness, Kurtosis.

2) Discrete wavelet transform (DWT)

The wavelet is a powerful mathematical tool for feature extraction, and has become the method of choice in many medical image analysis and classification problems. The main advantage of wavelets is extracting the image features at different directions and scales as they provide localized time-frequency information of a signal which is particularly beneficial for classification [6], [7], [20], [21]. The DWT is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length. It separates data into different frequency components, and then studies each component with resolution matched to its scale [21]. The DWT is implemented using cascaded filter banks in which the lowpass and highpass filters satisfy certain specific constraints. The basic scheme of DWT decomposition and its application to MR images is shown in Fig.4 where the functions h(n) and g(n) represent the coefficients of the high-pass and low-pass filters, respectively. As a result, there are four sub-band (LL, LH, HH, HL) images at each scale. The LL subband can be regarded as the approximation component of the image, while the LH, HL, HH subbands can be regarded as the detailed components of the image [7].

In our algorithm, level-3 approximation coefficient of Haar wavelet was utilized to extract the image features as illustrated in Fig.5. Haar transform has an advantage of high speed computation and efficient performance to analyze the local feature of an image [21], [22]. The resulted features extracted was 32 × 32=1024 features for a brain MRI that it requires to be reduced.

3) Principal component analysis (PCA)

Excessive features increase the computation time and memory storage which sometimes causes some complications in the classification process (the curse of dimensionality), and so it is required to reduce the number of features [7]. Principal components are the projection of the original features onto the eigenvectors and correspond to the largest eigenvalues of the covariance matrix of the original feature set. PCA can be used to approximate the original data with lower dimensional feature vectors. The basic approach is to compute the eigenvectors of the covariance matrix of the original data, and approximate it by a linear combination of the leading eigenvectors [16]. We used the PCA to reduce the 1024 features extracted using DWT for each brain MR image.

D. Classification

In the classification stage the proposed intelligent methodology was able to differentiate between the normal and abnormal images which contain tumors. Supervised machine learning techniques have great performance in classification stage in medical image analysis problems. In this paper, linear support vector machine and multi-class support vector machine (multi-SVM) classifiers are used.

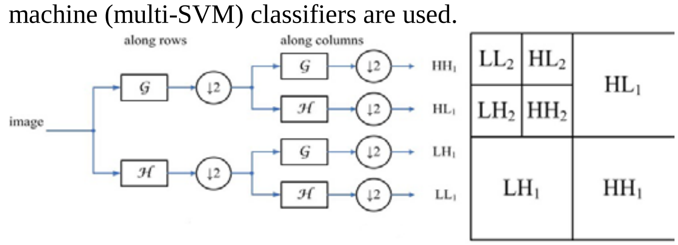
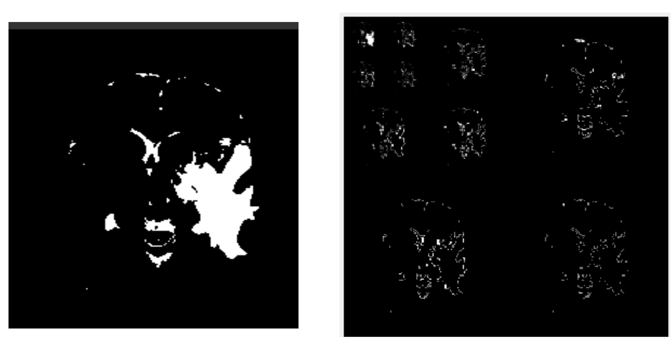


Fig. 4. DWT decomposition scheme



(a) Original segmented image (b) Level-3 Haar DWT decomposition

Fig. 5. applying Level-3 DWT decomposition on a segmented brain MRI

SVM is a supervised machine learning technique that is used to classify the input brain MR images into four classes: normal and three types of malignant brain tumors (glioblastoma, sarcoma and metastatic bronchogenic carcinoma). Table I shows the setting of the training and testing images from the dataset used.

1) Linear-SVM

SVM classifier is usually used in several research areas in image processing due to its high performance over other classifiers due to their capability of classifying classes that are linearly or non-linearly separable [5], [8]. Linear-SVM is a binary supervised classification method that can construct an optimal hyper plane in high dimensional feature space that maximize the space between classes [2], [13], [18], [23], [24], [25]. The Linear-SVM kernel classifier used to classify the input MRIs to normal and abnormal brain images where the brain has a tumor to be then identified by the multi-SVM.

2) Multi-SVM

Multi-SVM based on one-versus-one approach is an extended version of the binary SVM that can efficiently classify more than two classes [26], [27]. This approach construct different hyper planes using combinations of different classifiers for all possible combinations of classes with training a separate classifier for each different pair of classes [1], [28], [29]. In this paper, Multi-Svm is used to identify the tumor type if a tumor was detected in an image. It can classify three types of malignant tumors which are glioblastoma, sarcoma and metastatic bronchogenic carcinoma. Multi-class one-versus-one approach SVM was implemented using statistics and machine learning toolbox in MATLAB.

IV. Experimental results

The results of the proposed intelligent methodology was obtained using MATLAB R2015a under windows 8.1 operating system forming a friendly GUI to perform the automated process of analyzing the brain MRIs. The performance of the systems was measured in terms of sensitivity, Specificity and Accuracy that were defined as follows:

- Sensitivity (true positive rate) measures the proportion of abnormal MR images which are correctly identified

$$\text{Sensitivity}=\frac{TP}{(TP+FN)}$$

- Specificity (true negative rate) measures the proportion of normal MR images which are correctly identified

$$\text{Specificity}=\frac{TN}{(TN+FP)}$$

TABLE I. Setting of training and test images

Total No. of images	No. of Images in training set (39)		No. of Images in testing set (27)	
	Normal	Abnormal	Normal	Abnormal
66	14	25	8	19

- Accuracy measures the total number of MR images which are correctly identified

$$\text{Accuracy}=\frac{(TP+TN)}{(TP+TN+FP+FN)}$$

Where:

TP (True Positives) is the correctly classified positive cases

TN (True Negative) is the correctly classified negative cases

FP (False Positives) is the incorrectly classified negative cases

FN (False Negative) is the incorrectly classified positive cases

Table II and III shows the experimental results of the proposed intelligent methodology. The Linear-SVM classifier gives 93% accuracy using GLCM feature extraction technique and 97% accuracy using the level-3 Haar DWT feature extraction technique to classify the input testing images into normal and abnormal brain MRIs over the dataset used. However, the Multiclass SVM one-versus-one approach that have been used to classify the abnormal images into the tumor classes reaches 100% accuracy over the brain tumor classes (glioblastoma, sarcoma and metastatic bronchogenic carcinoma) with both of the feature extraction techniques.

TABLE II. Classification rates

The proposed intelligent methodology	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)
K-means + GLCM + Linear-SVM	18	7	1	1	95%	89%	93%
K-means + DWT +PCA + Linear-SVM	19	7	0	1	95%	100%	97%

TABLE III. Tumor classes accuracy

The proposed intelligent methodology	Tumor Class	Accuracy (%)
K-means + GLCM + Linear-SVM + Multi-SVM	Glioblastoma	100%
	Sarcoma	100%
	Metastatic bronchogenic carcinoma	75%
K-means + DWT +PCA + Linear-SVM +Multi-SVM	Glioblastoma	100%
	Sarcoma	100%
	Metastatic bronchogenic carcinoma	100%

V. Conclusion and Future Work

In this paper a hybrid classification technique based on support vector machine was proposed to define the abnormal MR images and then differentiate between 3 types of malignant brain tumors: glioblastoma, sarcoma and metastatic bronchogenic carcinoma. Also, a two feature extraction techniques was presented: GLCM along with some additional intensity features and DWT integrated with PCA that gave more accuracy for the classifier and robustness. In our system, K-means technique was used for brain MRI segmentation. The system evaluation shows high accuracy and robustness over the dataset used where Linear-SVM classifier for detecting abnormal images reaches 93% accuracy with GLCM feature extraction technique and 97% with DWT integrated with PCA feature extraction technique. On the other hand Multi-SVM classifier for identifying the tumor type in the abnormal images reaches 100% accuracy for the three types of malignant tumors.

Our intelligent methodology to be modified with other segmentation techniques which may result in higher accuracy for the system detection of abnormalities in brain MR images.

References

[1] K.G. Khambhata and S.R. Panchal, "Multiclass Classification of brain tumor in MR Images", IJIRCCE, Vol. 4, No. 5, pp. 8982- 8992, May 2016.

[2] A. Chinnu, "MRI Brain Tumor Classification Using SVM and Histogram Based Image Segmentation", International Journal of Computer Science and Information Technologies, Vol. 6, 2015, pp.1505-1508.

[3] http://www.cbtrus.org/factsheet/factsheet.html

[4] https://radiology.ucsf.edu/blog/neuroradiology/exploring-the-brain-is-ct-or-mri-better-for-brain-imaging

[5] A.N. Pathak, and R.K. Sunkaria, "Multiclass Brain Tumor Classification using SVM", International Journal of Computer Applications, Vol. 97, No.23, pp. 34-38, July 2014.

[6] D.R. Nayak, R. Dash, and B. Majhi, "Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests", Neurocomputing, Vol. 177, 2016, pp.188–197.

[7] E.A. El-Dahshan., H. Mohsen, K. Revett, and A.M. Salem, "Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm", Expert Syst. Appl. Vol. 41, 2014, pp. 5526–5545.

[8] M. Alfonse, and A. M. Salem, An Automatic Classification of Brain Tumors through MRI Using Support Vector Machine", Egyptian Computer Science Journal, Vol. 40, No. 3, pp. 11-21, September 2016.

[9] http:// med.harvard.edu/AANLIB/

[10] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation", Mag. Resonance Imag., Vol. 31, 2013, pp. 1426-1438.

[11] N. Dhanachandra, K. Manglem, and Y.J. Chanu. "Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm", Procedia Computer Science, Vol. 54, 2015, pp. 764-771.

[12] R.D.Deshmukh and C.Jadhav, "Study of Different Brain Tumor MRI Image Segmentation Techniques", International Journal of Computer Science Engineering and Technology, Vol. 4, No.4, 2014, pp. 133-136.

[13] S. Suhag, and L.M. Saini, "Automatic brain tumor detection and classification using svm classifier", Proceedings of ISER 2nd International Conference, Singapore, pp. 55-59, July 2015.

[14] H. Mohsen, E.A. El-Dahshan, E. El-Horbaty, and A.M. Salem, "A comparative study of segmentation techniques for brain Magnetic Resonance Images", Athens: ATINER'S Conference Paper Series, No: COM2016-1994, 2016.

[15] S.Tatiraju, and A. Mehta,"Image Segmentation using k-means clustering, EM and Normalized Cuts", Department of EECS, pp. 1-7.

[16] P. Rathi, and S. Palani, "Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis", International Journal of InformationSciences and Techniques, Vol.2, No.4, July 2012.

[17] R.M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification", IEEE Trans. Syst. Man Cybern., SMC-3, 1973, pp. 610-621.

[18] A.R. Kavitha, L.Chitra, and R.kanaga, "Brain Tumor Segmentation using Genetic Algorithm with SVM Classifier", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 5, No. 3, pp. 1468-1471, March 2016.

[19] S. Jafarpour, Z. Sedghi, and M.C. Amirani, "A robust brain MRI classification with GLCM features", International Journal of Computer Applications, Vol. 37, No.12, January 2012.

[20] L. Singh, and G. Chetty , "A Comparative Study of MRI Data using Various Machine Learning and Pattern Recognition Algorithms to Detect Brain Abnormalities", in Proceedings of the Tenth Australasian Data Mining Conference (AusDM 2012), Sydney, Australia, 2012.

[21] A.B. AL-Khafaji, "Classification of MRI Brain Images using Discrete Wavelet Transform And K-NN", International Journal Of Engineering Sciences & Research Technology, Vol. 4, No. 11, November 2015.

[22] S. Arora, Y.S. Brar, and S. Kumar, "Haar Wavelet Transform For Solution Of Image Retrieval", International Journal of Advanced Computer and Mathematical Sciences, Vol. 5, No. 2, 2014, pp. 27-31.

[23] M. Ahmad, M. Hassan, I. Shafi and A. Osman, " Classification of Tumors in Human Brain MRI using Wavelet and Support Vector Machine", IOSR Journal of Computer Engineering, Vol.8, No. 2, 2012, pp. 25-31.

[24] U. Ali, A. Shaukat , M. Hussain , J. Ali , K. Khan, M.B. Khan, and M.A. Shah, "Automatic Cancerous Tissue Classification using Discrete Wavelet Transformation and Support Vector Machine", J. Basic. Appl. Sci. Res., Vol. 6, No.7, 2016, pp. 15-23.

[25] J. Dolz, A. Laprie, S. Ken, H.A. Leroy, N. Reyns, L. Massoptier, and M. Vermandel, "Supervised machine learning-based classification scheme to segment the brainstem on mri in multicenter brain tumor treatment context", International journal of computer assisted radiology and surgery, Vol. 11, No.1, 2016, pp. 43–51.

[26] J. Xie, P.K. Douglas, Y. N. Wu, A.L. Brody, and A.E. Anderson, "Decoding the Encoding of Functional Brain Networks: an fMRI Classification Comparison of Non-negative Matrix Factorization (NMF), Independent Component Analysis (ICA), and Sparse Coding Algorithms", in press.

[27] T.C. Emmerling, J. Zimmermann, B. Sorger, M. Frost, and R. Goebel, "Time-resolved searchlight analysis of imagined visual motion using 7 T ultra-high field fMRI: Data on interindividual differences", Data in brief, 7, 2016, 468-471.

[28] G. Anthony, H. Gregg, and M. Tshilidzi, "Image Classification using SVMs: One-against-One Vs One-againstAll", In: Proc. 28th Asian Conference on Remote Sensing Kuala Lumpur, Malaysia, November 2007.

[29] G. Singh, M. Vadera, L. Samavedham, and E. Lim, "Machine Learning-Based Framework for Multi-Class Diagnosis of Neurodegenerative Diseases: A Study on Parkinson's Disease", IFAC-PapersOnLine, Vol. 49, No. 7, 2016, pp. 990–995.