J. Appl. Environ. Biol. Sci., 7(4)230-236, 2017
© 2017, TextRoad Publication

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Magnetic Resonance Images (MRI) Classificatıon through Support Vector Machine

Zia Ur Rahman*[1], Izaz Ahmad Khan[1], Safyan Mukhtar[2], Raees Khan[2], Sajjad Khan[2], Reena Rashid[3], Waqas Ahmad[4]

Department of Computer Science, Bacha Khan University, Charsadda, KPK, Pakistan[1]
Department of Mathematics &Statistics, Bacha Khan University, Charsadda, KPK, Pakistan[2]
Department of Computer Science, Shaheed Benzir Butto Women University, Peshawar, KPK, Pakistan[3]
Department of Computer Science, University of South Asia, Lahore, Punjab, Pakistan[4]

Received: December 3, 2016
Accepted: February 12, 2017

ABSTRACT

The Magnetic Resonance (MR) images are important for the credentials of diseases of human brain. The accurate dissection of the magnetic resonance images into several group tissues, for example, white matter, cerebrospinal fluid and gray matter is useful in analysis of brain. Support Vector Machines (SVMs) offer a number of advantages over other classification techniques but it has some shortcomings because it makes a liberal use of the kernel function which requires more computational time, estimation of error/trade off parameter is a across validation process which is wastage of both data and computation and the kernel function must satisfy the Mercer's condition which means it will be a positive integral operator of continuous symmetry. In SVM training involves non-linear optimization. The objective function is convex due to which solution is relatively straight forward as compared to other methods. The number of basic functions in the resulting model is smaller than the number of training points but still they are large and directly proportional to the size of training set. SVM determines model parameters, which corresponds to a convex optimization problem due to which any local solution is a global optimum. SVM classification classifies the data with the objective of minimizing error on the training set but also maximizes the margin between the two classes due to which it avoids over fitting that produces good generalization. It also leads to sparse model dependent only on subset of kernel function. The criteria for comparison are classification error rate and total time taken in training and testing the data is taken from the McConnell Brain Image Center, Montreal Neurological Institute, McGill University. The size of image volume and thickness is 181x 217x60 pixels and 3mm respectively. The data taken has 3% noise level and 40% intensity non-uniformity.

KEYWORDS: Magnetic Resonance Image, Support Vector Machines (SVM), LIBSVM, Gray Matter (GM), White Matter (WM), Cerebrospinal Fluid (CSF),SVMLITE.

1. INTRODUCTION

Magnetic Resonance (MR) imaging is a significant technique for the findings of the diseases of the human brain. Magnetic resonance image segmentation has been recycled for many detecting determinations but spontaneous segmentation of magnetic resonance examinations is of important value in neurological pathology medical study and exploration. The accurate dissection of the magnetic resonance images into several tissues clutch, for example, cerebrospinal fluid (CSF), white matter (WM) and gray matter (GM) is helpful in identification and prediction of various diseases of brain which consist of neurodegenerative syndromes such as movement disorders and alzheimer diseases. Image segmentation is mostly done by edge detection, histogram thresholding, feature space clustering, region-based method, neural networks and some other techniques but these techniques have some shortcomings due to which SVM is used. Edge detection method suffers from difficulties when an input image has too many edges [1], where recognition of a closed curve or boundary is not possible. Histogram thresholding technique has two major disadvantages [2], first is that it cannot produce satisfactory results when the input image has no obvious gray level tip, and second is that the connection of the segmented areas is not guaranteed. Feature based clustering method [3] cannot easily identify the existence of the number of clusters and obtaining the suitable feature to get the necessary information. While neural networks [4] and [5] are not assured to produce the intended result as its performance depends upon the initialization process.

Image segmentation can be regarded as a classification problem [6] in which image is divided into similar regions and a best label is given to each pixel. The SVMs is a kernel-based machine learning technique [7] in which classification is performed by calculating the best possible linear and nonlinear hyper plane with maximum margin between the separating classes from the input. SVMs are an inductive method of the

Corresponding author: Zia Ur Rahman, Department of Computer Science, Bacha Khan University, Charsadda, KPK, Pakistan. Email: zia.cs@bkuc.edu.pk

Structured Risk Minimization which reduces the likelihood of the generalization errors. The results show [7] that the image segmentation done by SVMs method gives better results than the previously used methods.The Support Vector Machines (SVM) is important to use for the arrangement of human brain tissues because of the effective results exposed in various pattern recognition system and which is use of Kernel functions and give results in high generalization capability [8].

SVMs offer a number of advantages over other classification techniques but it has some shortcomings because it makes a liberal use of the kernel function which requires more computational time, estimation of error/trade off parameter is a across validation process which is wastage of both data and computation and the kernel function must satisfy the Mercer's condition which means it will be a positive integral operator of continuous symmetry [9]. While RVM has no such disadvantage and offers probabilistic Bayesian learning framework that can give accurate prediction models with the use of less Kernel functions. The advantages offered by RVM makes it a better candidate for the classification of the MR images of the brain tissues. The MR image of brain consists of three types of tissues [WM, GM, CSF] and the background [10]. The two dimensional MR images are provided to RVM classifiers which are trained by supervised learning methodology to verify whether a pixel belongs to one of three tissues or background. The results will be compared with the SVM, FCM and BP-MLP. The criteria for comparison are classification error rate and total time taken in training and testing the data is taken from the McConnell Brain Image Center, Montreal Neurological Institute, McGill University which is available at http://www.bic.mni.mcgill.ca/brainweb/. The size of image volume and thickness is 181x 217x60 pixels and 3mm respectively. The data taken has 3% noise level and 40% intensity non-uniformity. The data is given labels for training purpose as follows 0: background, 1: WM, 2: GM and 3: CSF.

2. MATERIALS AND METHODS

The aim of the support vector classification is to device a computationally efficient way of learning 'good' separating hyper planes is a high dimensional feature space, where by 'good' hypothesis we will understand ones optimizing the generalization bounds and by 'computationally efficient' we will mean algorithms able to deal with sample sizes of the order of 100000 instances [11]. The generalization theory gives a clear guidance about how to control capacity and hence prevent over fitting by controlling the hyper plane margin measures, while optimization theory provides the mathematical techniques necessary to find hyper planes optimizing these measures [11]. Different generalization bound exist motivating different algorithms e.g. optimizing the maximal margin, the margin distribution, the number of support vectors etc.

Most common and well-established approaches which reduce the problem to minimizing the norm of the weight vector are discussed.

Support Vector Machines (SVM): Support Vector Machines (SVM) are used for the classification of brain tissues because of the efficient results shown in many pattern recognition system and use of Kernel functions which results in high generalization ability [10]. The aim of the support vector classification is to device a computationally efficient way of learning 'good' separating hyper planes is a high dimensional feature space, where by 'good' hypothesis we will understand ones optimizing the generalization bounds and by 'computationally efficient' we will mean algorithms able to deal with sample sizes of the order of 100000 instances [11]. The generalization theory gives a clear guidance about how to control capacity and hence prevent over fitting by controlling the hyper plane margin measures, while optimization theory provides the mathematical techniques necessary to find hyper planes optimizing these measures [11]. Different generalization bound exist motivating different algorithms e.g. optimizing the maximal margin, the margin distribution, the number of support vectors etc. Most common and well-established approaches which reduce the problem to minimizing the norm of the weight vector are discussed. These are the simplest model of SVM and also first ones. They work on data which is linearly separable in the feature space and hence cannot be used in many real-world situations. It is easiest algorithm and main building block for more complex SVMs [12].The algorithm which bounds the generalization error of linear machines in terms of the margin ms (f) of the hypothesis f with respect to the training set S. The maximal margin classifier optimizes this bound by separating the data with the maximal margin hyper plane and given that the bound does not depend on the dimensionality of the space, this separation can be sought in any kernel-induced feature space.

The maximal margin classifier forms the strategy of the first SVM, namely to find the maximal margin hyper plane in an appropriately chosen kernel-induced feature space. This strategy implemented by reducing it to a convex optimization problem minimizing a quadratic function under linear inequality constraints. Linear classifier has inherent degree of freedom because the function associated with the hyper plane (w, b) does not change if the rescaling of the hyper plane to $(\lambda w, \lambda b)$ is performed [13]. There will be a change in the margin as measured by the function output as opposed to the geometric margin. The margin of the function output is referred as the functional margin.

Hence, an equally well optimize the geometric margin by fixing the functional margin to be equal to 1 (hypothesis with functional margin 1 are sometimes known as canonical hyper planes) and minimizing the norm of the weight vector.

The weight vector realizing a functional margin of 1 on the positive point x^+ and the negative point x^-. The geometric margin is computed as follows.A functional margin of 1 implies.

$$\left\langle w.x^+ \right\rangle + b = +1, \left\langle w.x^- \right\rangle + b = -1 \quad \textbf{(1)}$$

To compute geometric margin w must normalize. The geometric margin γ is then the functional margin of the resulting classifier.

$$\gamma = \frac{1}{2}\left[\left\langle \frac{w}{\|w\|}.x^+ \right\rangle - \left\langle \frac{w}{\|w\|}.x^- \right\rangle\right] = \frac{1}{\|w\|_2} \quad \textbf{(2)}$$

The resulting geometric margin will be equal to

$$\frac{1}{\|w\|_2} \quad \textbf{(3)}$$

Proposition 1

Linear separable training sample

$$s = ((x_1, y_1), \cdots\cdots\cdots (x_{1,} y_1)) \quad \textbf{(4)}$$

The hyperplane (w, b) that solve the optimization problem

$$\text{Minimize}_{w,b} \left\langle w.w \right\rangle, \quad \textbf{(5)}$$

$$\text{Subject to} \quad \begin{array}{l} y_i(\left\langle w.x_i \right\rangle + b) \geq 1 \\ i = 1, \cdots\cdots, l \end{array} \quad \textbf{(6)}$$

It shows that the maximum margin hyperplane

$$\gamma = \frac{1}{\|w\|} \quad \textbf{(7)}$$

Now the optimization problem is transform into its corresponding dual form.

The primal langrangian is

$$L(w,b,\alpha) = \frac{1}{2}\left\langle w.w \right\rangle - \sum_{i=1}^{l} \alpha_i \left[y_i(\left\langle w.x_i \right\rangle + b) - 1 \right] \quad \textbf{(8)}$$

Where $\alpha_i \geq 0$ are the langrangian multipliers

The corresponding dual form is formed by differentiating with respect to (w, b), we get

$$w = \sum_{i=1}^{l} y_i \alpha_i x_i, \quad \textbf{(9)}$$

$$0 = \sum_{i=1}^{l} y_i \alpha i, \quad \textbf{(10)}$$

Putting above vales into the primal form we get the following equation.

$$L(w,b,\alpha) = \sum_{i=1}^{l} \alpha_i - \tfrac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \left\langle x_i, x_j \right\rangle \quad \textbf{(11)}$$

It shows the substitution relation that the hypothesis can be described as the linear combination of the training points. The dual representation is also required for kernels.

Now minimum of <W. W> is found which equivalent to find maximum is of $w(\alpha)$.

Maximize $$w(\alpha) = \sum_{i=1}^{l} \alpha_i - \tfrac{1}{2}\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \left\langle x_i . x_j \right\rangle, \quad \textbf{(12)}$$

Subject to $$\sum_{i=1}^{l} y_i \alpha_i = 0, \quad \alpha_i \geq 0, i = 1, \cdots, l. \quad \textbf{(13)}$$

$$w^* = \sum_{i=1}^{l} y_i \alpha_i^* x_i \quad \textbf{(14)}$$

It realizes that the maximum margin hyperplane with geometric margin.

The value of b does not appear is dual problem so b* must be found by making use of the primal constraints.

$$b^* = -\frac{\max_{y_i=-1}(\left\langle w^* . x_i \right\rangle) + \min_{y_i=1}(\left\langle w^* . x_i \right\rangle)}{2} \quad \textbf{(15)}$$

The optimal hyper plane can be expressed in the dual representation in terms of this subset of the parameters.

$$f\left(x, a^*, b^*\right) = \sum_{i \in SV}^{l} y_i a_i^* \left\langle x_{i.} . x \right\rangle + b \quad \textbf{(16)}$$

The langrage multipliers associated with each point become the dual variable giving them intuitive interpretation quantifying how important a given training point is in forming the final solution. Non-support vector points have no influence on solution.

3. RESULTS AND DISCUSSIONS

SVMs are used for classification of magnetic resonance images [MRI] because many conventional classification techniques [10] for example neural networks suffer from the generalization. The brain image consists of three types of brain tissues and background. The SVMs are used [10] because of the effectiveness of the technique in pattern recognition, usage of the Kernel-functions which maps the input into high dimensional feature space where simple linear classifiers are used for classification and the usage of the Structured Risk Minimization method which provides better classification as compared to the already used classification techniques. The result obtained are compared with two well-known technique known as BP-MLP [14] a supervised learning technique and fuzzy c-mean method known as FCM [15-17], an unsupervised learning technique in terms of the average computational time and classification error rate.

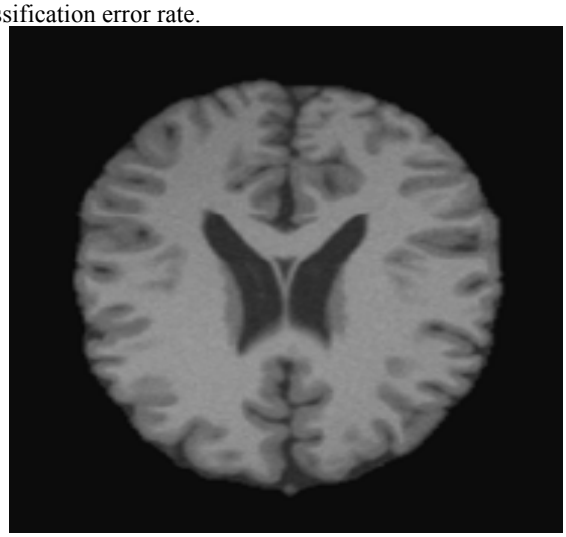

Fig 1: MR image of the brain[10]

The MRI of the brain is divided into four parts the three types of brain tissues known as WM, GM, CSF and background. The results [10] show a significant amount of improvement in classification error rate of 9.89 percent for SVM as compared to 11.65 percent and 12.22 percent for FCM and BP-MLP respectively. Similarly the total time taken by SVM [10] is 6.18 seconds which is also a significant improvement over 173.51 second and 244.63 seconds for the FCM and BP-MLP respectively.

Method	Training Time (s)	Testing Time (s)	Total Time (s)	Classification Error Rate (%)
FCM		173.51	173.51	11.65
BP-MLP	242.91	1.72	244.63	12.22
SVM	1.12	5.06	6.18	9.89

Table 1: Performance comparison of SVM, BP-MLP and FCM[10]

Image Data Extraction through EMMA (Extensible Matlab Medical Image Analysis): EMMA is used for extracting the data and to be used in MATLAB. The following programming codes were implemented for the extraction of the image data, scaled and then use for the classification of the data by randomly splitting the data into train data, test data and prediction.

- Handle = openimage ('c:\matlab701\work\normal.mnc'); //the code creates an handle for the image data

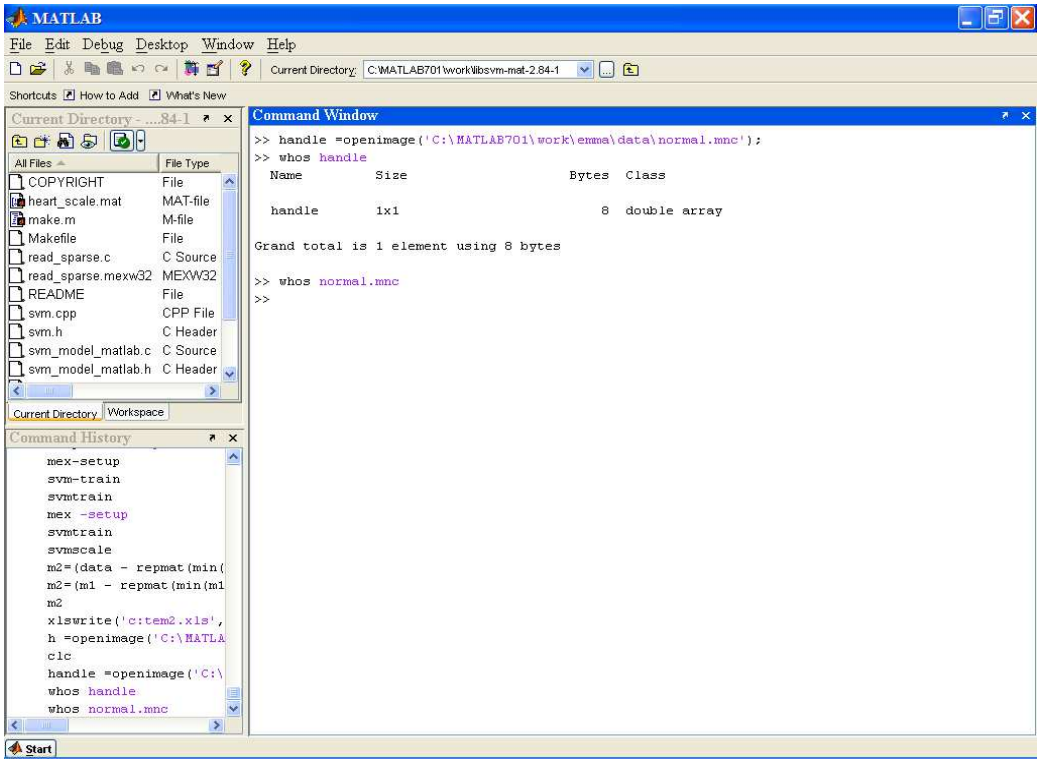

Fig 2 MATLAB snapshot (OPENIMAGE function)

OPENIMAGE function is used in MATLAB for reading MINC files. It takes the file name as a parameter while the user defined variable stores the value, in this case it is handle. This function creates all variables which are required by get/put functions such as getimages and putimages.
The fig 2 demonstrates the implementation of OPENIMAGE command on the image data normal.mnc.
Then whos function is used which lists variables and their sizes. In this case the variable name is handle whose size is 1x1, byte is zero and its class is double array.

- Datainfo_ns = getimageinfo (handle, NumSlices); //this line extract the num of slices contained in the image data

- Datainfo_ns = getimageinfo (handle, NumFrames); //this line extract the num of frames contained in the image data

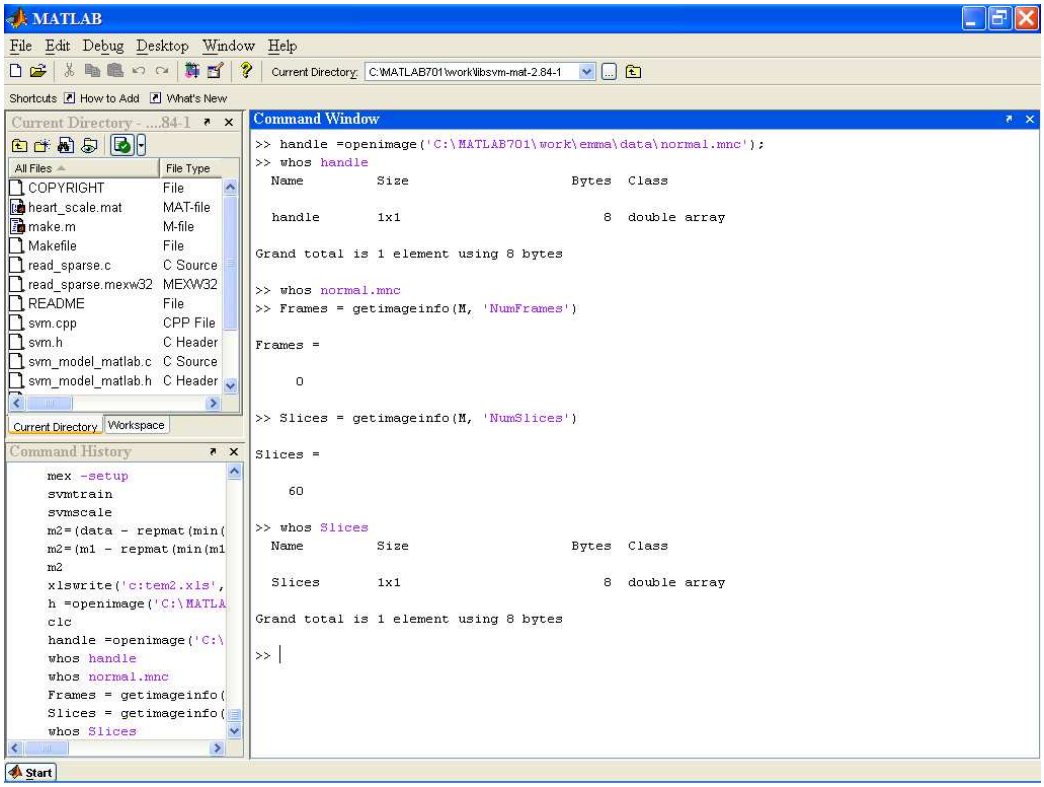

Fig 3 MATLAB Snapshot (GETIMAGEINFO command)

GETIMAGEINFO has important details of an open image. It has two parameters, first is handle of the image and second is the type of information required from that image which is then stored in another variable. Different types of information can be obtained from an image for example time, Filename, NumFrames, NumSlices, ImageHeight, ImageWidth, ImageSize, DimSizes, FrameLengths, FrameTimes, MidFrameTimes, MinMax, AllMin, AllMax, Steps, Start, DirCosines, Permutation.

The fig 3 demonstrates the implementation of GETIMAGEINFO command on the image data for NumFrames and NumSlices.The current experiment has used NumFrames and NumSlices. NumFrames is the number of frames in the study; if it is zero then it means it is a non-dynamic study (equivalent to "time"). In this case the study is non-dynamic that's why it has returned zero.

NumSlices is the number of slices in the study (0 if no slice dimension). In this case the number of slices is 60 as shown in the fig 3. Then again whos function is used to get information about slices variable which has recorded information about the NumSlices.

- Images = getimages (handle, 60, 1:0); // this line creates the image for the original image data
- View image (images (7, 1));

VIEWIMAGE function is used for showing image from a data matrix.

4. FUTURE WORK

Many limitations have been faced ranging from SVM tools which require specifically data format which are not developed for strings, and data downloading and decrypting. Beside, these problem the most challenging problem was EMMA as the links given was not accessible. Many emails were sent to the researchers in the McConnell Brain Image Centre, Montreal Neurological Institute, McGill University. In the end the EMMA was found and the data downloaded is decrypted. The scaling was a big issue but after many attempts it is performed through LIBSVM. The actual classification is performed through SVMLITE. The data should be given a general format if possible so that it can be used easily by researchers. So, for future research appropriate scaling mechanisms should be included in the new versions of LIBSVM which can scale MINC format. The database for brain data is under development due to which there is a possibility to have reduced error rate. Thus the RVM

can be a major tool for MR image classification of brain tissues which can be extremely helpful in the diagnosis of the brain diseases.

5. CONCLUSIONS

Asalready disscussed that there are several dangrous diseases in human brain, just like, alzheimer movements disorder diseases, post traumatic syndrome, white matter metabolic and parkinson related diseases etc. The Support Vector Machine is successfully used for the classification of the Magnetic Resonance Images and the results produced are better than the previously used supervised and unsupervised learning techniques. The Support Vector Machine is more powerful and has similar functional form which is a motivation for using the Relevance Vector Machines (RVM) for the Magnetic Resonance images.

REFERENCES

[1] Bellon, O. R. P., Silva, L.: New improvements to range image segmentation by edge detection, IEEESignal Processing Lett., 9(2), 2002, 43-45.

[2] Cheng, H. D., Jiang, X. H., Wang, J. L.: Color image segmentation based on homogram thresholding and region merging, *Pattern Recognition*, **35**(2), 2002 373-393.

[3] Kuntimad, G., Ranganath, H. S.: Perfect image segmentation using pulse coupled neural networks, IEEE Trans. Neural Networks, 10(3), 1999, 591-598.

[4] Shen, X., Spann, M., Nacken, P.: Segmentation of 2D and 3D images through a hierarchical clustering based on region modelling, *Pattern Recognition*, **31**(9), 1998, 1295-1309.

[5] Wong, K. P., Feng, D., Meikle, S. R., Fulham, M. J.: Segmentation of dynamic PET images using cluster analysis, *IEEE Trans. Nucl. Sci.*, **49**(1), 2002, 200-207.

[6] Cheng, H. D., Jiang, X. H., Sun, Y. & Wang, J. (2001): Color image segmentation: advances and prospects; Available at: http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V14-43W07HY1&_user=4581748&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000046823&_version=1&_urlVersion=0&_userid=4581748&md5=5de983d9542b3ba98da2881bf5567d09

[7] Yu, Y., Chang, C: Scenery Image Segmentation Using Support Vector Machines, *Fundamental Informatics 61 (2004) 379–388*

[8] Z. U. Rahman et al., "Magnetic Resonance Images Classification through Relevance Vector Machine," Journal of Applied Environmental and Biological Sciences, vol. 7, no. 1, pp. 213–217, Jan. 2017.

[9] Tipping, M.P. & Smola, A. (2001): Sparse Bayesian learning and the Relevance Vector Machine: Journal of Machine Learning Research. 1 211-244

[10] Zhang, X., Xiao, X.L., Tian, J.W., Liu, J., & Xu, G.Y.: Application of Support Vector Machines in classification of magnetic resonance images, International Journal of Computers and Applications, 28(2), 2006 122-128

[11] Cristianini, N. & Taylor, J.S. (2006) Support vector machines and other kernel -based learning methods. Cambridge: Cambridge University press

[12] Hearst, M. A. (1998). Trends and controversies: Support vector machines. IEEE intelligent systems 18-21

[13] Cortes, C. & Vapnik, V. (1995). Support vector networks. Machine learning, 20: 273 – 297.

[14] W.E. Reddick, J.O. Glass, E.N. Cook, & T.D. Elkin, Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks, IEEE Trans. on Medical Imaging, 16 (6), 1997, 911–918.

[15] D.L. Pham & J.L. Prince, (1999) An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities, Pattern Recognition Letters, 20(1), 57–68.

[16] F. Dad et al., "Optimal Path Selection Using Dijkstra's Algorithm inCluster-based LEACH Protocol," Journal of Applied Environmental and Biological Sciences, vol. 7, no. 2, pp. 194–198, Feb. 2017.

[17] Z. U. Rahman et al., "Investıgatıng the Pakıstan's Offshore Software Industry Infrastructure," Journal of Applied Environmental and Biological Sciences, vol. 7, no. 3, pp. 237–243, Mar. 2017