Original Research Article

Multiple Sclerosis Lesion Segmentation Using Ensemble Machine Learning

Randa ElSebely[*], Bassem Abdullah, Ashraf A Salem, Ahmed Hassan Yousef

Department of Computer & Systems Engineering, Ain Shams University, Cairo, Egypt

*Corresponding author: Randa ElSebely

Abstract

A lesion is an area of tissue that has been damaged through injury or disease. So a brain lesion is an area of injury or disease within the brain. While the definition sounds simple, understanding brain lesions can be complicated. That's because there are many types of brain lesions. They can range from small to large, from few to many, from relatively harmless to life-threatening. magnetic resonance imaging (MRI) is increasingly used nowadays. Manual delineation of Multiple Sclerosis (MS) lesions in MR images by human experts is time-consuming, subjective and prone to inter-expert variability. Therefore, automatic segmentation is needed as an alternative to manual segmentation. In this paper, the 2D discrete wavelet transform (DWT) is used to extract local information from analyzing MR images. Ensemble decision trees (EDT) and Ensemble support vector machines (ESVM) are used to segment MS lesions and automatically differentiate between blocks in regions of MS lesions and the blocks in non-MS. lesions. We evaluate our approach on real MRI data sets. We can detect MS lesions with accuracy more than 98 %. Technique evaluated using real MRI datasets. The results compared with ground truth. The main contribution of the proposed techniques described in this paper is the use of DWT with Ensemble Machine Learning and solving the problem of imbalanced classification data without changing or losing trained data.

Keywords: MRI, Multiple Sclerosis, Automatic segmentation, Ensemble support vector machine, Ensemble Decision Tree, Lesions.

INTRODUCTION

MS is an autoimmune disease of the central nervous system. It may result in a variety of symptoms from blurred vision to severe muscle weakness and degradation, depending on the affected regions in the brain. As a result, MS is categorized as an auto-immune disease [1]. Nerve fibers covered by myelin protect the nerves and help them to conduct the electrical pulses. In MS patients, the myelin disappeared which is called Demyelination [2, 3]. Recent medical researches suggest that genetic and environmental factors mainly cause MS. The fact that a monozygotic twin with co-twin suffering from MS has a chance of 25% of developing the disease while the same chance for a non-twin sibling stands at only 3%, is an indicator of the role of genetic influence in developing MS [4]. Also, MS mostly affects adults in developed countries. Moreover, roughly 2.5 million people in the world suffer MS. Iran is considered as a country with high MS prevalence (51.52 per 100000) in the Middle East. Although there is no definite treatment of MS, early

detection is important to decrease the progression of disease and managing symptoms. There are several ways used to diagnose MS such as blood testing, evoked potential test, lumbar puncture, etc [5]. To understand the Effect of The disease, MRI is used. MRI is known as a useful method for monitoring and diagnosing MS disease. MRI is based on the magnetic characteristics of the imaged tissue [6-8]. This method is a painless, safe, fast, noninvasive imaging technique that yields images of the body structures, specifically capturing images within brain layers, and provides valuable information for medical diagnosis. It is quite a tedious and challenging experience for an expert to analyze MR images because there are complications and difficulties in terms of understanding and interpreting anatomical borders that are invisible almost in all images. Manual detection of MS lesions in MR images by a human is time-consuming Therefore, automatic segmentation is an alternative to manual segmentation [9, 10]. A Lot of Methods have been proposed for the automatic segmentation. DWT is used to extract local information from analyzing MR images

and Textural features have been used to differentiate between lesion white matter, normal white matter and normal-appearing white matter. In this paper, we use a technique that calculates textural features to describe each Block of the 3D MRI Brain Image and DWT to extract features from MR Images [16]. Machine Learning Techniques is used To Classify between the blocks and detect the blocks that have MS lesions. With advanced technology, diagnosis by using machine learning techniques is easier. In this study, Ensemble support vector machine and Ensemble Decision Tree approaches are implemented to obtain more accurate results for MS lesion detection and to provide a comparative result. The aim of this study in the long term is to create a decision support system for physicians by using image processing and Machine Learning techniques. Future studies will focus on a completely automatic version of the decision support system by using machine learning techniques. The literature studies discuss many different statistical and machine learning techniques that have been applied to develop MS prediction models, such as logistic regression, linear discriminate analysis, naïve Bayes, decision trees, artificial neural networks, k-nearest neighbor, and support vector machine methods [2-10]. More specifically, studies comparing some of the above-mentioned techniques have shown that SVM performs better than many of the other related techniques. It is necessary when constructing the SVM classifier to determine a specific kernel function, such as a polynomial or radial basis function (RBF), which is an important learning parameter. However, there have been very few studies focused on assessing the prediction performance of SVM classifiers constructed using different kernel functions. Besides, it is known that combining multiple classifiers or classifier ensembles, another active research area of pattern classification, often gives better performance than single classifiers [11-20]. However, except who show that EDT constructed by bagging perform better than the single decision tree model. Another complicating factor is that the collected dataset for MS segmentation is usually class imbalanced, with the minority class containing a small number of patients with MS and the majority class containing a large number of patients with Non-MS. This means that using only prediction accuracy or classification accuracy to evaluate the prediction models is insufficient [9, 1]. Therefore, our research objective is to compare between MS segmentation using EDT and ESVM .Their performance will be assessed by different evaluation metrics, including classification accuracy, Dice Similarity, Sensitivity, Detected Lesion Load (DLL), True positive rate (TPR) and Positive Predictive Value (PPV). Consequently, the findings of this paper should allow future researchers to easily choose the most effective baseline technique that can provide the optimal prediction performance for future comparison.

MATERIALS & METHODS
Figure-1 shows the workflow of the proposed method discussed below in detail.
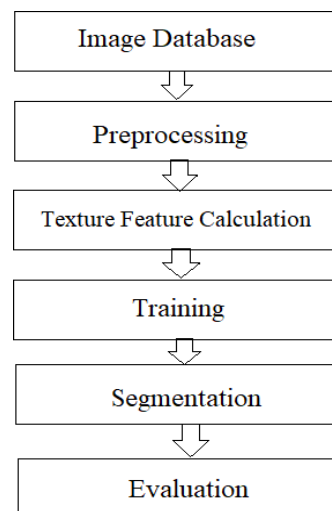

Fig-1: The workflow

Image Database
The sources of these datasets are the workshop of MS lesion Segmentation Challenge 2008 [18, 19] and real MRI studies for MS subjects acquired at the University of Miami. Datasets used for evaluation include 62 cases the datasets are divided into labeled cases used for training (20 cases) and non-labeled cases used for testing (31 cases). All Brain MRI are 3D Raw Type (512x512x512) with resolution 0.5mm x 0.5 mm x 0.5mm . The two approaches, EDT and ESVM, both implemented in MATLAB environment on real brain MS data.

Preprocessing
Raw images are not suitable for analysis due to having many artifacts such as intensity inhomogeneity, extra cranial tissues, and noises. These unwanted data reduce the accuracy and efficiency of segmentation. Therefore, suitable pre-processing techniques must be used to improve the quality of the image for further steps. A literature study shows several pre-processing and feature extraction methods for MR brain image analysis. In this part, pre-processing and segmentation methods will be discussed in detail. An intensity & Histogram correction Applied to 3D MRI Image by applying a contrast- brightness correction to maximize the intersection between the histogram of the training and segmentation datasets followed by using a 3D RF filter [20, 21]. Figure 2 & 3 show the effect of the preprocessing technique on improving MRI Brain slice.
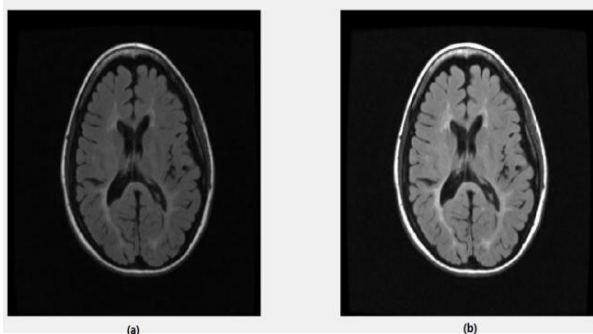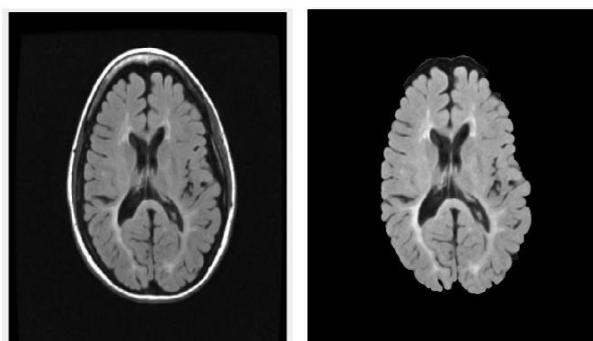
Fig-2(a): A slice from the subject CHB_train_Case 01 before preprocessing. (b) A slice from subject CHB_train_Case01 after preprocessing
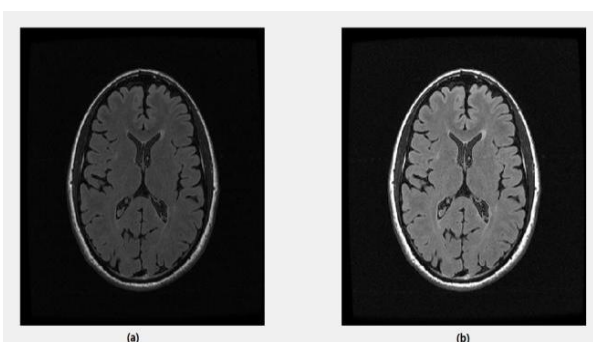
Fig-3(a): A slice from the subject CHB_train_Case 07 before preprocessing. (b) A slice from subject CHB_train_Case07 after preprocessing

There are some non-brain tissues in MR brain images, and these tissues have no valuable data for analysis, so it is necessary to remove these tissues for further processing. The Graph Cut is one of the well-known brain splitting tools which removes non-brain tissues in MR brain images. This method is simple and one of the robust brain extraction tools. Graph Cut tool is publicly available in MATLAB Image Processing Toolbox. Figure-4 shows the effect of the Applying Graph Cut technique on improving FLAIR slice [22-26].


Fig-4(a): A slice from the subject CHB_train_Case01 before Applying Graph Cut. (b) A slice from subject CHB_train_Case01 after Applying Graph Cut

Feature Extraction and Calculation

After Pre Processing of each 3D MRI Image slice is processed through a trained detector engine to obtain initial MS lesions regions. The detector engine in our method is implemented using a machine learning algorithm. Training the detector engine is done by processing the training dataset and dividing its slices into square blocks and assigning a binary class for each block. If the block contains at least one pixel manually labeled as MS, it is classified as an MS block (class 1). Otherwise, the block is classified as a non-MS block (class 0) if all of its pixels are labeled as non-MS pixels. Each block is described by a feature vector which mainly represents textural features of the block. During segmentation, the slice to be segmented is divided into square non- overlapping blocks and each block is classified by the trained engine as MS block or non-MS block [8]. The Size of each Block is wxw pixel [2].

Fourier's representation functions as a superposition of sines and cosines and represents the frequency domain of the original data. The main flaw of the Fourier transform is that it fails to represent where the signal has several discontinuities and sharp spikes [43, 44]. For solving this problem, Wavelets have been introduced for time- frequency analysis. Grossmann and Morlet first introduced wavelets. Wavelets are mathematical functions that divide data or signals into different frequency components. DWT is the most famous transformation technique adopted for image compression. In DWT, a digital signal is analyzed in terms of time and frequency content which profit from the filtering methods [45, 46]. The DWT is applied to MR images and extracts an approximated and detailed version of images hierarchically [47]. As decomposition level increases more stable approximation coefficients are obtained. In our method, we use 2 level DWT, and the design simulated in MATLAB.
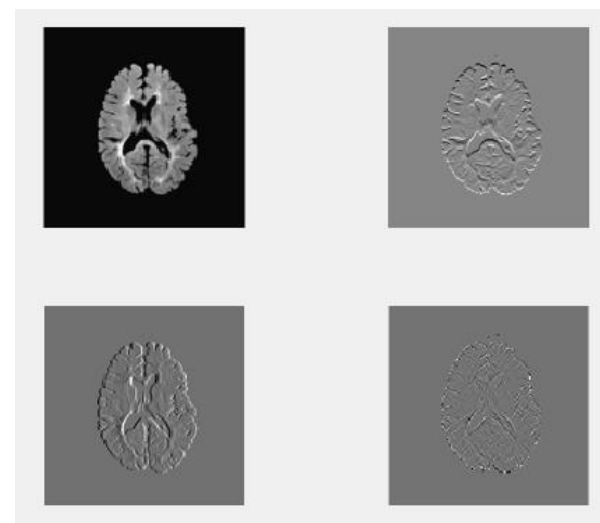
Fig-5: Feature Extracted from MRI using 2D-DWT

Textural features can be categorized according to the matrix or vector used to calculate the feature. We evaluate histogram, gradient, run-length matrix and co-occurrence based features. For all feature calculations, the image is represented by a function $f(x,y)$ of two space variables x and y, x=0,1,…N-1, and y=0,1,…, M-1. The function $f(x,y)$ can take any value i=0,1,…,G-1

where G is a total number of intensity levels in the image. Histogram based features The intensity level is a function h(i) providing, for each intensity level i, the number of pixels in the whole image having this intensity.

$$H(i) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \delta(f(x,y), i); \delta(i,j) = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases} \dots\dots\dots\dots (1)$$

We Calculate (Mean, Variance) From the Histogram based feature. Gradient-based features g(x,y) is defined for each pixel in the image based on the neighborhood size. For a 3x3 pixel's neighborhood, g is defined as follows [27-31].

$$\Delta x = f(x-1, y) - f(x+1, y) \dots\dots\dots\dots (2)$$
$$\Delta y = f(x, y-1) - f(x, y+1)$$
$$g(x,y) = \sqrt{\Delta x^2 + \Delta y^2}$$

We Calculate (gradient mean and gradient Variance) From gradient-based feature. Run-length matrix based features are defined in a specific direction. Usually, a matrix is calculated for the horizontal, vertical, 45° and 135° directions. Co-occurrence matrix based features are a form of a second-order histogram that is defined for certain angle θ and certain distance d. Usually, the co-occurrence matrix is calculated for d = 1 and 2 with angles θ = 0°, 45°, 90°, and 135°. We Calculate (contrast, entropy, and absolute value) From the Co-occurrence matrix. Position Features are the location of Each Slice of MRI Brain image with respect to the bottom slice and the radial Euclidean distance between the block's top-left pixel and the center of the slice normalized by dividing it by the longest diameter of the slice. The neighboring block's features are the difference between the mean intensity of the current block and the mean intensity of each of the eight neighboring blocks as shown in if Figure-6.
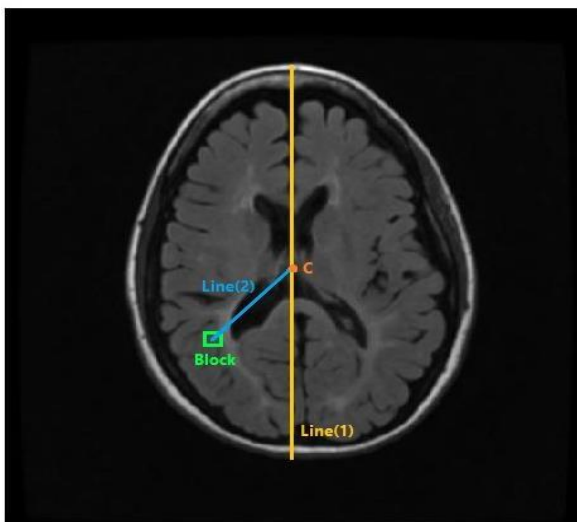

Fig-6: Position Features Extraction on a Sample Slice. Calculation of the normalized radial distance between Block and slice center C (length of line 1 / length of line 2)

Neighboring features is the Calculation of the eight neighboring blocks features (the difference between mean the gray scale of the centered red block and mean gray scale of each of the eight neighboring green blocks in the grid circled by the yellow circle) as shown in Figure-7.
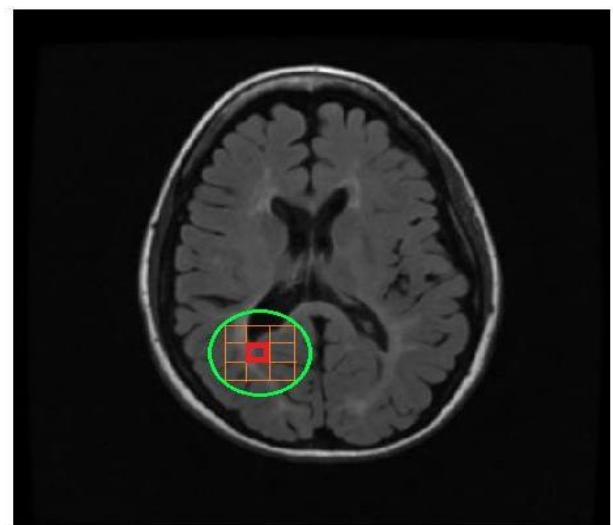
Fig-7: Pixel and Neighboring Pixels

Training

In the training phase, after, removing unwanted data and noise reduction from MR images are needed. This step is an essential part because if we do not remove the skull correctly from the raw images; all the next processes will be affected. In this step, the graph cut was applied to have the most valuable area of the brain. The segmentation of MS lesions amounts to a binary classification problem, i.e., y_i is either 0 or 1. The training entry is said to be a positive entry if y_i is 1 and negative in the other case. Training the classifier with more than one subject dataset allows the classifier to learn features of real MS-blocks, Training Data Applied by MATLAB2018. Cross-Validation [32] is a common method that is applied for running comparisons amongst learning algorithms. The data are separated into two sections; one section is used for learning or training the model, and the other one issued for validating the model. The most famous form of cross-validation is the k-fold cross-validation method. There are other forms of cross-validation such as the Hold-Out Cross-Validation method. The following table

shows the differences between these two popular cross-validation methods [23, 33].

Imbalanced data distribution is an important challenge of machine learning workflow. An imbalanced dataset means instances of one of the two classes is higher than the other, in another way, the number of observations is not the same for all the classes in a classification dataset. This problem is faced not only in the binary class data but also in the multi-class data. To solve this problem important techniques described below will help you to deal with our imbalanced data [37-39].

Oversampling

This technique is used to modify the unequal data classes to create balanced datasets. When the quantity of data is insufficient, the oversampling method tries to balance by incrementing the size of rare samples. A primary technique used in oversampling is SMOTE (Synthetic Minority Over-sampling Technique) [39]. In this technique, the minority class is over-sampled by producing synthetic examples rather than by over- sampling with replacement and for each minority class observation, it calculates the k nearest neighbors (k- NN). The oversampling technique solves the imbalanced data problem but the drawback of oversampling is of course higher speed required for the training (higher complexity and cost), speed, and require more time so an overall slower speed [49].
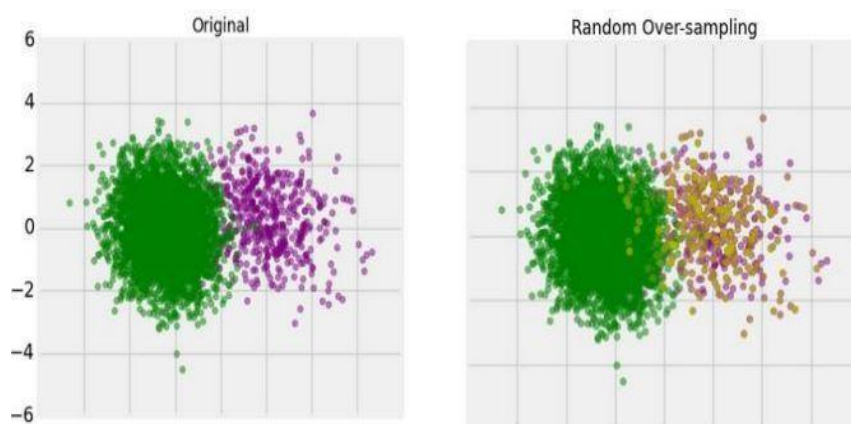

Fig-8: Oversampling Technique

Undersampling

Unlike oversampling, this technique balances the imbalance dataset by reducing the size of the class which is in abundance. While the under- sampling is fast and solves the imbalanced data problem the disadvantage is that it discards potentially useful data. Which affects the classification performance [49].
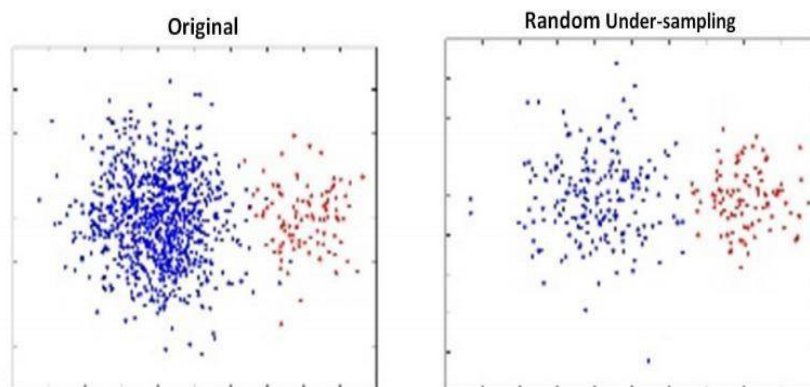
Fig-9: Undersampling Technique

Cost-Sensitive Learning Technique

Cost-Sensitive Learning (CSL) takes the misclassification costs into consideration by minimizing the total cost. The goal of this technique is mainly to pursue a high accuracy of classifying examples into a set of known classes. It is playing as one of the important roles in machine learning algorithms including real-world data mining applications. In this technique, the costs of false- positive (FP), false negative (FN), true positive (TP), and true negative (TN) can be represented in a cost matrix as shown below where $C(i,j)$ represents the misclassification cost of classifying an instance and also "i" the predicted class and "j" is the actual class. Here is an example of the cost matrix for binary classification [49].

	Actual negative	Actual positive
Predict negative	C(0,0), or TN	C(0,1), or FN
Predict positive	C(1,0), or FP	C(1,1), or TP

Ensemble Learning Techniques

The ensemble-based method is another technique that is used to deal with imbalanced data sets, and the ensemble technique is combined with the result or performance of several classifiers to improve the performance of single classifier. This method modifies the generalization ability of individual classifiers by assembling various classifiers. It mainly combines the outputs of multiple base learners. There are various approaches in ensemble learning such as Bagging, Boosting, etc. Bagging or Bootstrap Aggregating tries to implement similar learners on a smaller dataset and then takes a mean of all the predictions. The Boosting is an iterative technique that rectifies the weight of an observation depending on the last classification. This method decreases the bias error and builds strong predictive models. This paper proposed method uses CSL and Ensemble Learning Techniques due to reliability, speed and without losing any data [49].

SVM is a supervised machine learning technique for classification. In this technique, the model tries to figure out the best hyper plane separating the data into varying classes [11, 12]. A wide margin has crucial importance for accuracy. There are different application routines of SVM depending on the dataset such as linear, quadratic, radial basis function (RBF), etc. In this study, SVM is chosen because it was sufficient to classify the labels, where labeling the data was done for -1 and 1 in 2D space. The approach finds a hyper-surface dividing the negative and positive samples (pixels, labels) with the largest possible margin on all sides of the hyper-plane. Here, the classification problem results as either -1 or 1 as in equation 1. For each slice of the input dataset, each group of connected pixels is labeled and analyzed [13, 14]. Training the algorithm and SVM structure was created using training data was implemented to increase the training capability of SVM as well as increasing the accuracy of the test data. Therefore, the training dataset was divided into several groups; each new dataset was used to train an SVM classifier, and the Ensemble SVM was achieved when these classifiers were combined. Randomly, different SVM classifier structures were obtained by dividing the training data into several parts and using them as inputs of the SVM classification. The block diagram of this method is shown in Figure-10.
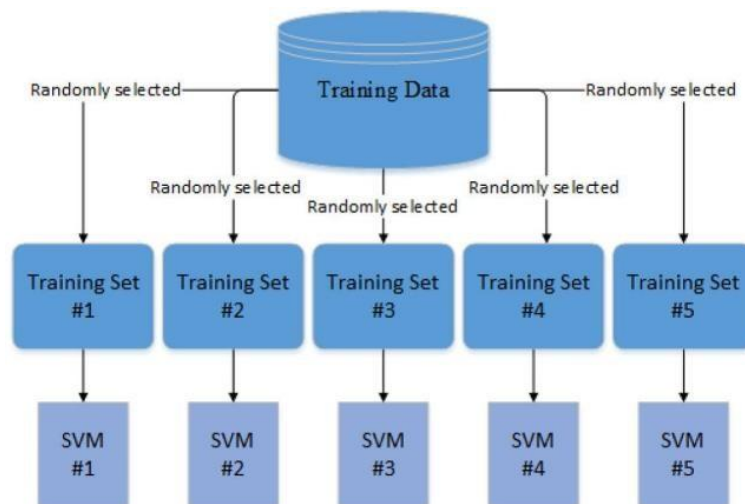

Fig-10: The training section block diagram in the proposed ESVM algorithm

In order to tackle the problem of extremely imbalanced data in our dataset, the EDT classifier is used to categorize each MS or non-MS brain tissue and to improve the accuracy of the minority class (e.g. MS). EDT is one of the predictive modeling approaches used in statistics, data mining and machine learning. It uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). Tree models where the target variable can take a discrete set of values are called classification trees in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels [17]. In order to tackle the problem of extremely imbalanced data in our dataset, the EDT classifier is used to categorize each MS or non-MS brain tissue and to improve the accuracy of the minority class (e.g. MS). EDT is an ensemble technique that uses multiple decision trees. EDT combines several decision trees to produce better predictive performance than utilizing a single decision tree. The main principle behind the ensemble model is that a group of weak learners come together to form a strong learner.
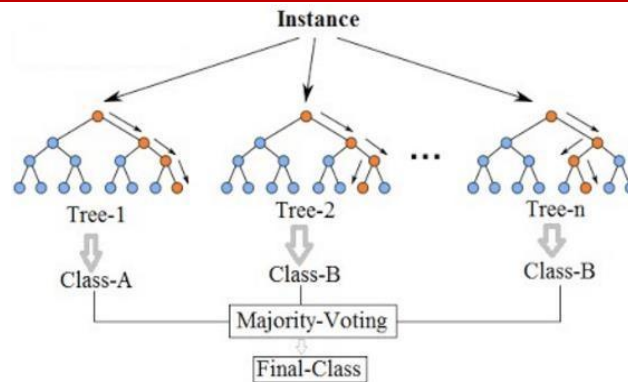
Fig-11: Example of EDT Algorithm used for Classification

Segmentation

Each of the slices in MRI Image is divided into blocks of size wxw pixels. The features vector for each block is calculated. Then we use Trained ESVM and EDT to predict the class label of each Block. For any block classified as MS block, assuming true positive classification, this does not mean that all pixels of the block should be classified as MS pixels because the SVM engine is trained to detect the blocks that contain MS pixels completely or partially.

Evaluation of the Proposed Method

To evaluate the performance of the automatic segmentation method, comparisons are performed with state of the art methods. Evaluation of segmentation with respect to the ground truth. The voxels that are marked as MS in both automatic segmentation and manual segmentation are the true positives (TP), voxels appear only in automatic segmentation are false positives (FP), voxels appear only in manual segmentation are false negatives (FN) and voxels that are marked as non-MS in both sets are true negatives (TN) [48].

Dice Similarity (DS)

The dice similarity (DS) is a measure of the similarity between the manual segmentation(X) and the automatic segmentation (Y). The equation for the calculation can be written as:

$$\text{Dice} = 2 \, X \cap Y / (X + Y) \dots \dots \dots \dots (3)$$

Sensitivity (Sen)

A measure of how many lesions are detected. It can be calculated as the percentage of true positive voxels to the total number of MS voxels in the ground truth. For the two sets manual segmentation (X) and

automatic segmentation (Y), the sensitivity is calculated as*:*

$$\text{Sensitivity} = /X \cap Y / / /X/ \dots \dots \dots \dots (4)$$

Detected Lesion Load (DLL)

We introduce the Detected Lesion Load (DLL) metric as a percentage of detected lesion volume with reference to the original lesion volume [11, 48]. The detected lesion volume takes into account all the positive lesions whether true or false.

$$\text{DLL} = |Y| / |X| \dots \dots \dots \dots \dots \dots \dots \dots . (5)$$

True Positive Rate (TPR) and Positive Predictive Value (PPV)

For the two sets manual segmentation (X) and automatic segmentation (Y), true positives (TP), false positives (FP) and false negatives (FN) can be calculated as:

$$TP = |X \cap Y| \dots \dots \dots \dots \dots \dots \dots (6)$$
$$FP = |Y - X| \dots \dots \dots \dots \dots \dots \dots (7)$$
$$FN = |X - Y| \dots \dots \dots \dots \dots \dots \dots (8)$$

The True positive rate (TPR) and Positive Predictive Value (PPV) are defined as:

$$TPR = TP / (TP + FN) \dots \dots \dots \dots \dots (9)$$
$$PPV = TP / (TP + FP) \dots \dots \dots \dots \dots (10)$$
$$\text{Accuracy (ACC)} = (TP + TN) / (TP + TN + FP + FN) * 100 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (11)$$

EXPERIMENTAL RESULTS

MATLAB 2018 Functions are used to train data. Table-1 obtains the comparison of classification result between ESVM with RBF kernel and EDT. Cross-Validation is also applied in the training phase. In the shown table we obtain Accuracy, Sensitivity, DLL m TPR, and PPV Parameters.

Table-I: Esvm Classification Results

Case	Performance Evaluation					
	ACC	DS	Sen	DLL	TPR	PPV
C01	99.4%	0.31	0.23	0.38	0.5	0.23
C02	99.53%	0.28	0.107	0.38	0.62	0.12
C03	99.44%	0.45	0.38	0.42	0.89	0.33
C04	99.4%	0.36	0.28	0.36	0.79	0.25
C05	98.72%	0.40	0.36	0.38	0.87	0.266
C06	99.58%	0.52	0.50	0.40	0.69	0.313

Table-II: Edt Classification Results

Case	Performance Evaluation					
	ACC	*DS*	*Sen*	*DLL*	*TPR*	*PPV*
C01	*98.7%*	0.547	0.58	0.834	*0.55*	*0.46*
C02	98.08%	0.575	0.90	1.63	1.01	0.90
C03	99.41%	0.533	0.5	0.529	0.71	0.45
C04	98.8%	0.583	0.52	0.667	0.70	0.79
C05	99.4%	0.48	0.42	0.52	0.48	0.45
C06	99%	0.514	0.6	0.55	0.59	0.46

CONCLUSION

We have developed an automated method for detection of MS lesions in brain MR images by applying a textural- based ESVM with RBF Kernel and EDT to one channel MRI data. The methods have been tested using real MRI datasets. The aim of this study is to determine MS lesion with solving the problem of imbalanced data without losing data by using ESVM and EDT approach. Real data results are subject to this work, where the results are compared to the given and known input data. From the experimental results, it is observed that ESVM and EDT show classification accuracy of 98.96% and 99.4% respectively and demonstrated that the suggested methods perform well and achieved very good recognition results for MS classification.

REFERENCES

1. WebMd: Brain Lesions: Causes, Symptoms, Treatments [online] https://www.webmd.com/brain/brain-lesions-causes-symptoms-treatments#1

2. Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., ... & Lublin, F. D. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of neurology*, *69*(2), 292-302.

3. Ghribi, O., Sellami, L., Slima, M. B., Mhiri, C., Dammak, M., & Hamida, A. B. (2018). Multiple sclerosis exploration based on automatic MRI modalities segmentation approach with advanced volumetric evaluations for essential feature extraction. *Biomedical Signal Processing and Control*, *40*, 473-487.

4. Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D., Ebers, G. C., & Canadian Collaborative Study Group. (2003). Twin concordance and sibling recurrence rates in multiple sclerosis. *Proceedings of the National Academy of Sciences*, *100*(22), 12877-12882.

5. Traboulsee, A. L., & Li, D. K. (2006). The role of MRI in the diagnosis of multiple. *Adv Neurol*, *98*, 125-46.

6. Miller, D. H., Grossman, R. I., Reingold, S. C., & McFarland, H. F. (1998). The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain: a journal of neurology*, *121*(1), 3-24.

7. McDonald, W. I. (1989). Diagnosis of multiple sclerosis, *British Medical Journal (BMJ)*, 299(6700):635-637.

8. Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., ... & Fujihara, K. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, *17*(2), 162-173.

9. Confavreux, C., Vukusic, S., Grimaud, J., & Moreau, T. (1999). Clinical progression and decision making process in multiple sclerosis. *Multiple Sclerosis Journal*, *5*(4), 212-215.

10. Van Leemput, K. (2001). Quantitative analysis of signal abnormalities in MR imaging for multiple sclerosis and Creutzfeldt-Jakob disease.

11. Raghtate, G. S., & Salankar, S. S. (2015, December). Automatic brain MRI classification using modified ant colony system and neural network classifier. In *2015 International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 1241-1246). IEEE.

12. Tang, X., Zeng, W., Shi, Y., & Zhao, L. (2018). Brain activation detection by modified neighborhood one-class SVM on fMRI data. *Biomedical Signal Processing and Control*, *39*, 448-458.

13. Abdullah, N., Ngah, U. K., & Aziz, S. A. (2011, May). Image classification of brain MRI using support vector machine. In *2011 IEEE International Conference on Imaging Systems and Techniques* (pp. 242-247). IEEE.

14. Trigui, R., Mitéran, J., Walker, P. M., Sellami, L., & Hamida, A. B. (2017). Automatic classification and localization of prostate cancer using multi-parametric MRI/MRS. *Biomedical Signal Processing and Control*, *31*, 189-198.

15. Havaei, M., Jodoin, P. M., & Larochelle, H. (2014, August). Efficient interactive brain tumor segmentation as within-brain kNN classification. In *2014 22nd International Conference on Pattern Recognition* (pp. 556-561). IEEE.

16. Anbeek, P., Vincken, K. L., & Viergever, M. A. (2008). Automated MS-lesion segmentation by k-nearest neighbor classification. *MIDAS Journal*.

17. Geremia, E., Menze, B. H., Clatz, O., Konukoglu, E., Criminisi, A., & Ayache, N. (2010, September). Spatial decision forests for MS lesion segmentation in multi-channel MR images. In *International Conference on Medical Image*

Computing and Computer-Assisted Intervention (pp. 111-118). Springer, Berlin, Heidelberg.

18. Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., ... & Warfield, S. (2008). 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. *Midas Journal*, *2008*, 1-6.

19. Lesion, M. S. (2008). Segmentation Challenge. [Online] http://www.ia.unc.edu/MSseg/

20. Prima, S., Ourselin, S., & Ayache, N. (2002). Computation of the mid-sagittal plane in 3-D brain images. *IEEE transactions on medical imaging*, *21*(2), 122-138.

21. Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., & Mazziotta, J. C. (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of computer assisted tomography*, *22*(1), 139-152.

22. Kalavathi, P. (2014). Improved Brian Extraction tool using marker-controlled watershed segmentation. *IOSR Journal of computer Engineering*, *16*(5), 106-111.

23. Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, *17*(3), 143-155.

24. Denitto, M., Farinelli, A., Figueiredo, M. A., & Bicego, M. (2017). A biclustering approach based on factor graphs and the max-sum algorithm. *Pattern Recognition*, *62*, 114-124.

25. Somasundaram, K., & Kalavathi, P. (2012, December). A novel skull stripping technique for T1-weighted MRI human head Scans. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing* (pp. 1-8).

26. Alansary, A., Ismail, M., Soliman, A., Khalifa, F., Nitzken, M., Elnakib, A., ... & Zurada, J. M. (2015). Infant brain extraction in T1-weighted MR images using BET and refinement using LCDG and MGRF models. *IEEE journal of biomedical and health informatics*, *20*(3), 925-935.

27. Nixon, M. S., & Aguado, A. S. (2014). Feature extraction & image processing for computer vision, Academic Press.

28. Xu, D., & Wang, Y. (2007). An automated feature extraction and emboli detection system based on the PCA and fuzzy sets. *Computers in Biology and Medicine*, *37*(6), 861-871.

29. Mayer, R., Simone II, C. B., Skinner, W., Turkbey, B., & Choykey, P. (2018). Pilot study for supervised target detection applied to spatially registered multiparametric MRI in order to non-invasively score prostate cancer. *Computers in biology and medicine*, *94*, 65-73.

30. Abdullah, B. A. (2011). Segmentation of multiple sclerosis lesions in brain MRI.

31. Zhang, J., Tong, L., Wang, L., & Li, N. (2008). Texture analysis of multiple sclerosis: a comparative study. *Magnetic resonance imaging*, *26*(8), 1160-1166.

32. Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, *10*(7), 1895-1923.

33. Kwan, R. K. S., Evans, A. C., & Pike, G. B. (1996, September). An extensible MRI simulator for post-processing evaluation. In *International Conference on Visualization in Biomedical Computing* (pp. 135-140). Springer, Berlin, Heidelberg.

34. Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., & Evans, A. C. (1998). Design and construction of a realistic digital brain phantom. *IEEE transactions on medical imaging*, *17*(3), 463-468.

35. Van Leemput, K. (2001). Quantitative analysis of signal abnormalities in MR imaging for multiple sclerosis and Creutzfeldt-Jakob disease.

36. Rosner, L. J., & Ross, S. (1992). *Multiple Sclerosis*. New York: Simon and Schuster.

37. Wu, G., & Chang, E. Y. (2003). Adaptive feature-space conformal transformation for imbalanced-data learning. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)* (pp. 816-823).

38. Akbani, R., Kwek, S., & Japkowicz, N. (2004, September). Applying support vector machines to imbalanced datasets. In *European conference on machine learning* (pp. 39-50). Springer, Berlin, Heidelberg.

39. Batuwita, R., & Palade, V. (2010, July). Efficient resampling methods for training support vector machines with imbalanced datasets. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

40. Mathew, J., Luo, M., Pang, C. K., & Chan, H. L. (2015, November). Kernel-based SMOTE for SVM classification of imbalanced datasets. In *IECON 2015-41st Annual Conference of the IEEE Industrial Electronics Society* (pp. 001127-001132). IEEE.

41. Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., ... & Ye, X. (2017). Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. *International journal of computer assisted radiology and surgery*, *12*(2), 183-203.

42. Sun, Z., & Chi, M. (2015, June). Superpixel-based active learning for the classification of hyperspectral images. In *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (pp. 1-4). IEEE.

43. Graps, A. (1995). An Introduction to Wavelets, IEEE Computational Science and Engineering. In *IEEE Computer Society,* 2(2):50-61.

44. Lewis, A. S., & Knowles, G. (1992). Image compression using the 2-D wavelet transform. *IEEE Transactions on image Processing*, *1*(2), 244-250.

45. Gupta, D., & Choubey, S. (2015). Discrete wavelet transform for image processing. *International Journal of Emerging Technology and Advanced Engineering*, *4*(3), 598-602.

46. Wang, S., Zhang, Y., Dong, Z., Du, S., Ji, G., Yan, J., ... & Phillips, P. (2015). Feed- forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *International Journal of Imaging Systems and Technology*, *25*(2), 153-164.

47. Heidari Gheshlaghi, S., Ranjbar, A., Abolfazl Suratgar, A., Bagher Menhaj, M., & Faraji, F. (2019). A Superpixel Segmentation Based Technique for Multiple Sclerosis Lesion Detection. *arXiv preprint arXiv:1907.03109*.

48. Abdullah, B. A. (2012). Segmentation of Multiple Sclerosis Lesions in Brain MRI.

49. https://analyticsindiamag.com/5-important-techniques-to-process-imbalanced-data-in-machine-learning/