

# A Comparative Study of MRI Data using Various Machine Learning and Pattern Recognition Algorithms to Detect Brain Abnormalities

<sup>1</sup>Lavneet Singh, <sup>2</sup>Girija Chetty

<sup>1,2</sup>Faculty of Information Sciences and Engineering  
University of Canberra, Australia  
Lavneet.singh@canberra.edu.au  
Girija.chetty@canberra.edu.au

## Abstract

In this study, we present the investigations being pursued in our research laboratory on magnetic resonance images (MRI) of various states of brain by extracting the most significant features, and to classify them into normal and abnormal brain images. We propose a novel method based on deep and extreme machine learning on wavelet transform to initially decompose the images, and then use various features selection and search algorithms to extract the most significant features of brain from the MRI images. By using a comparative study with different classifiers to detect the abnormality of brain images from publicly available neuro-imaging dataset, we found that a principled approach involving wavelet based feature extraction, followed by selection of most significant features using PCA technique, and the classification using deep and extreme machine learning based classifiers results in a significant improvement in accuracy and faster training and testing time as compared to previously reported studies.

**Keywords:** -Deep Machine Learning, Extreme Machine Learning, MRI, PCA

## 1. INTRODUCTION

Magnetic Resonance Images (MRI) is an advance technique used for medical imaging and clinical medicine and an effective tool to study the various states of human brain. MRI images provide the rich information of various states of brain which can be used to study, diagnose and carry out unparalleled clinical

analysis of brain to find out if the brain is normal or abnormal. However, the data extracted from the images is very large and it is hard to make a conclusive diagnosis based on such raw data. In such cases, we need to use various image analysis tools to analyze the MRI images and to extract conclusive information to classify into normal or abnormalities of brain. The level of detail in MRI images is increasing rapidly with availability of 2-D and 3-D images of various organs inside the body.

Magnetic resonance imaging (MRI) is often the medical imaging method of choice when soft tissue delineation is necessary. This is especially true for any attempt to classify brain tissues (Fletcher-Heath, L. M., Hall, L. O., Goldgof, D. B. and Murtagh, F.R. 2001). The most important advantage of MR imaging is that it is non-invasive technique (Chaplot, S., Patnaik, L.M. and Jagannathan N.R. 2006). The use of computer technology in medical decision support is now widespread and pervasive across a wide range of medical area, such as cancer research, gastroenterology, heart diseases, brain tumors etc. (Gorunescu, F. 2007, Kara, S. and Dirgenali, F. 2007). Fully automatic normal and diseased human brain classification from magnetic resonance images (MRI) is of great importance for research and clinical studies. Recent work (Chaplot, S., Patnaik, L.M. and Jagannathan N.R. 2006, Maitra, M. and Chatterjee A. 2007) has shown that classification of human brain in magnetic resonance (MR) images is possible via machine learning and classification techniques such as artificial neural networks and support vector machine (SVM) (Chaplot, S., Patnaik, L.M. and Jagannathan N.R. 2006, Mishra, Anurag, Singh, Lavneet and Chetty, Girija 2012), and unsupervised techniques such as self-organization maps (SOM) (Chaplot, S., Patnaik, L.M. and Jagannathan N.R. 2006, Singh, Lavneet and Chetty, Girija. 2012) and fuzzy c-means combined with appropriate feature extraction techniques (Maitra, M. and Chatterjee A. 2007). Other supervised classification techniques, such as k-nearest neighbors (k-NN), which group pixels

---

<sup>1</sup> Copyright (c) 2012, Australian Computer Society, Inc. This paper appeared at the 10th Australasian Data Mining Conference (AusDM 2012), Sydney, Australia, December 2012. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 134, Yanchang Zhao, Jiuyong Li, Paul Kennedy, and Peter Christen, Ed. Reproduction for academic, not-for profit purposes permitted provided this text is included.

based on their similarities in each feature image (Fletcher-Heath, L. M., Hall, L. O., Goldgof, D. B. and Murtagh, F.R. 2001, Abdolmaleki, P., Mihara, F., Masuda, K. and DansoBuadu, Lawrence. (1997), Rosenbaum, T., Engelbrecht, V., Krolls, W. and Lenard, H. 1999, Cocosco, C., Zijdenbos, Alex P. and Evans, Alan C. 2003) can be used to classify the normal/pathological T2-weighted MRI images.

Out of several debilitating ageing related health conditions, white matter lesions (WMLs) are commonly detected in elders and in patients with multiple brain abnormalities like Alzheimer's disease, Huntington's disease and other neurological disorders. According to previous studies, it is believed that total volume of the lesions (lesion load) and their progression relate to the aging process as well as disease process. Therefore, segmentation and quantification of white matter lesions via texture analysis is very important in understanding the impact of aging and diagnosis of various brain abnormalities. Manual segmentation of WM lesions, which is still used in clinical practices, shows the limitation to differentiate brain abnormalities using human visual abilities. Such methods can produce a high risk of misinterpretation and can also contribute to variation in correct classification. Automated texture analysis algorithms have been developed to detect brain abnormalities using image segmentation techniques and machine learning algorithms. The signal of homogeneity and heterogeneity of abnormal areas in Region of Interest (ROI) in white matter lesions of brain in T2-MRI images can be quantified by texture analysis algorithms [reference]. The ability to measure small differences in MRI images is essential and important to reduce the diagnosis errors of brain abnormalities. The supervised feature classification from T2 MRI images, however, suffers from two problems. First, because of the large variability in image appearance between different datasets, the classifiers need to be retrained from each data source to achieve good performances. Second, these types of algorithms rely on manually labeled training datasets to compute the multi-spectral intensity distribution of the white matter lesions making the classification unreliable. Inspired by new segmentation algorithms in computer vision and machine learning, we propose an efficient semi-automatic and deep learning algorithm for white matter (WM) lesion segmentation around ROI based on extreme and deep machine learning. Further, we compare this novel approach with some of the other supervised machine learning techniques reported previously.

Rest of the paper is organized as follows. Next Section gives a brief background of materials and methods used in Section 2. The details of the feature extraction, and feature selection, and other classifiers techniques used is described in same Section 2, 3 and Section 4 presents some of the experimental work carried. The paper concludes with in section 5 with some outcomes of the experimental work using proposed approach, and outlines plans for future work.

## 2. Materials and Methods

### 2.1 Datasets

The input dataset consists of axial, T2-weighted, 256 X 256 pixel MR brain images (Fig. 1). These images were downloaded from the (Harvard Medical School website ([http:// med.harvard.edu/AANLIB/](http://med.harvard.edu/AANLIB/), Harvard Medical School 1999). Only those sections of the brain in which lateral ventricles are clearly seen are considered in our study. The number of MR brain images in the input dataset is 60 of which 6 are of normal brain and 54 are of abnormal brain. The abnormal brain image set consists of images of brain affected by Alzheimer's and other diseases. The remarkable feature of a normal human brain is the symmetry that it exhibits in the axial and coronal images. Asymmetry in an axial MR brain image strongly indicates abnormality. Hence symmetry in axial MRI images is an important feature that needs to be considered in deciding whether the MR image at hand is of a normal or an abnormal brain. A normal and an abnormal T2-weighted MRI brain image are shown in Fig. 1(a) and 1(b), respectively. Indeed, for multilayer learning models like deep and extreme machine learning algorithms needed big datasets for training, however due to lack of availability of proper datasets in MRI imaging, we used this dataset for examining the performance of proposed approaches for this paper, but acquiring other suitable datasets for future studies.

### 2.2 Coarse Image Segmentation

Color image segmentation is useful in many applications. From the segmentation results, it is possible to identify regions of interest and objects in the scene, which is very beneficial to the subsequent image analysis or annotation. However, due to the difficult nature of the problem, there are few automatic algorithms that can work well on a large variety of data. The problem of segmentation is difficult because of image texture. If an image contains only homogeneous color regions, clustering methods in color space are sufficient to handle the problem. In reality, natural scenes are rich in color and texture. It is difficult to identify image regions containing color-texture patterns. The approach taken in this work assumes the following:

- Each region in the image contains a uniformly distributed color-texture pattern.
- The color information in each image region can be represented by a few quantized colors, which is true for most color images of natural scenes.
- The colors between two neighboring regions are distinguishable - a basic assumption of any color image segmentation algorithm.

#### **K-Means clustering based Coarse Image Segmentation**

K-Means clustering algorithm is a well-known unsupervised clustering technique to classify any given input dataset. This algorithm classifies a given dataset into discrete k-clusters using which k-centroids are defined, one for each cluster. The next step is to take each

point in the given input data set and associate it to the possible nearest centroid. This process is repeated for all the input data points, based on which next level of clustering and the respective centroids are obtained. This procedure is iterated until it converges. This algorithm minimizes the following objective function.

$$J = \sum_{j=1}^k \sum_{i=1}^k \|x_i^j - c^j\|^2$$

Where  $\|x_i^j - c^j\|^2$  is a chosen distance measure between a data point  $(x_i)^j$  and the cluster centre,  $c_j$  is an indicator of the distance of the  $k$  data points from their respective cluster centers. The proposed unsupervised segmentation algorithm uses the principle of K-means clustering.

The proposed technique segments the region of interest (ROI) of an input image (input\_img) by an interactive user defined shape of square or rectangle to obtain select\_img. Then, the number of bins for coarse data computation (bin size), the size of overlapping kernel to partition (w-size) and the maximum number of clusters for segmentation (max\_class) are fed as input data for the computation of coarse data. The coarse data identified by each kernel is aggregated to form the final\_coarse\_data which is further clustered using the principle of K-means clustering in order to produce the segment\_img. The algorithmic description of the proposed technique is given herein under:

**Algorithm**

1. Read a grayscale image as input\_img  
/\* Define the area to be segmented as a runtime interactive input. The shape of the selection can either be a square or a rectangle \*/
2. Let select\_img is the selected subimage of input\_img
3. Assign:
  - a. binsize=5  
/\* number of bins for coarse data computation \*/
  - b. wsize= 7  
/\* wsize is the size of overlapping kernel to partition the select\_img \*/
  - c. max\_class= 3  
/\* maximum number of clusters for segmentation \*/
4. Repeat step 5 and 6 in algorithm until the select\_img is read
5. Read select\_img in the order of (wsize\*wsize) as window\_img
6. Compute coarse\_img for window\_img as coarse\_win\_data
7. Aggregate coarse\_win\_data for select\_img as final\_coarse\_data
8. Cluster final\_coarse\_data using K-means clustering technique using max\_class in order to obtain segment\_img
9. Stop

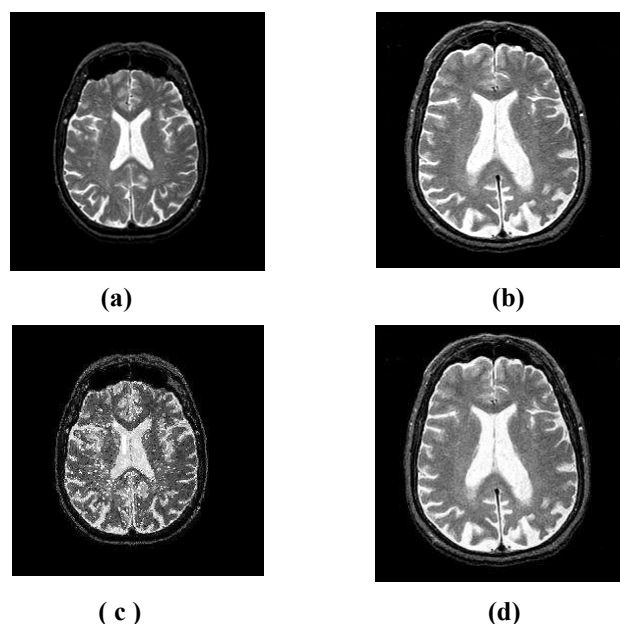
This algorithm can segment an object either fully or partially based on user's choice. If the image has a background and object(s) then it partitions the object from the background and displays its coarse image. If the image has no background, then the segmented image reveals the inner details of the object.



**Figure 1. (a) Coarse Segmented MRI Image based on above algorithm (b) ROI segmented image of White Lesions**

**3. Decomposition of images Using Wavelets**

Wavelets are mathematical functions that decompose data into different frequency components and then study each component with a resolution matched to its scale. Wavelets have emerged as powerful new mathematical tools for analysis of complex datasets. The Fourier transform provides representation of an image based only on its frequency content. Hence this representation is not spatially localized while wavelet functions are localized in space.



**Fig.2. (a) T2, weighted an axial MRI Brain Image; (b) T2, weighted an axial MR brain image as abnormal brain; (c) and (d) T2, weighted an axial MR brain image as normal and abnormal brain after Wavelets Decomposition and denoising**

**Discrete wavelets transform (DWT)**

The DWT is an implementation of the wavelet transform using a discrete set of the wavelet scales and translation obeying some defined rules. For practical computations, it is necessary to discretize the wavelet transform. The scale parameters are discretized on a logarithmic grid. The translation parameter ( $\tau$ ) is then discretized with respect to the scale parameter, i.e. sampling is done on the dyadic (as the base of the

logarithm is usually chosen as two) sampling grid. The discretized scale and translation parameters are given by,  $s = 2^{-m}$  and  $t = n2^{-m}$ , where  $m, n \in \mathbb{Z}$ , the set of all integers. Thus, the family of wavelet functions is represented in Eq. (1) and (2),

$$\psi_{m,n}(t) = 2^{\frac{m}{2}} \psi(2^m t - n) \quad (1)$$

$$W\psi(a, b) = \int_{-\infty}^{\infty} f(x) * \psi_{a,b}(t) dx \quad (2)$$

In case of images, the DWT is applied to each dimension separately. This result in an image  $Y$  is decomposed into a first level approximation component  $Y_a^1$  and detailed components  $Y_h^1$ ,  $Y_v^1$  and  $Y_d^1$  corresponding to horizontal, vertical and diagonal details. Fig.1 depicts the process of an image being decomposed into approximate and detailed components.

The approximation component ( $Y_a$ ) contains low frequency components of the image while the detailed components ( $Y_h$ ,  $Y_v$  and  $Y_d$ ) contain high frequency components. Thus,

$$Y = Y_a^1 + \{ Y_h^1 + Y_v^1 + Y_d^1 \} \quad (3)$$

At each decomposition level, the length of the decomposed signals is half the length of the signal in the previous stage. Hence the size of the approximation component obtained from the first level decomposition of an  $N \times N$  image is  $N/2 \times N/2$ , second level is  $N/4 \times N/4$  and so on. As the level of decomposition is increased, compact but coarser approximation of the image is obtained. Thus, wavelets provide a simple hierarchical framework for interpreting the image information.

#### 4. Deep Belief Nets

DBNs (Hinton, G.E. and Salakhutdinov, R.R. 2006) are multilayer, stochastic generative models that are created by learning a stack of Restricted Boltzmann Machines (RBMs), each of which is trained by using the hidden activities of the previous RBM as its training data. Each time a new RBM is added to the stack, the new DBN has a better variation lower bound on the log probability of the data than the previous DBN, provided the new RBM is learned in the appropriate way (Hinton, G.E. and Osindero, S. 2006).

A Restricted Boltzmann Machine (RBMs) is a complete bipartite undirected probabilistic graphical model. The nodes in the two partitions are referred as hidden and visible units. An RBM is defined as

$$p(v, h) = \frac{e^{-E(v, h)}}{\sum_u \sum_g e^{-E(u, g)}} \quad (4)$$

Where  $v \in V$  are the visible nodes and  $h \in H$  are the latent random variables. The energy function  $E(v, h, W)$  is described as

$$E = - \sum_{i=1}^D \sum_{j=1}^K v_i W_{ij} h_j \quad (5)$$

Where  $W \in \mathbb{R}^{D \times K}$  are the weights on the connections, and where we assume that the visible and hidden units both contain a node with value of 1 that acts to introduce bias. The conditional distribution for the binary visible and hidden units are defined as

$$p(v_i = 1/h, W) = \sigma(\sum_{j=1}^K W_{ij} h_j) \quad (6)$$

$$p(h_j = 1/v, W) = \sigma(\sum_{i=1}^D W_{ij} v_i) \quad (7)$$

Where  $\sigma$  is the sigmoid function. Using above equations, it easy to go back and forth between the layers of RBM. While training, it consists of some input to the RBM on the visible layer, and updating the weights and the biases such that  $p(v)$  is high. In generalized way, in as set of  $C$  training cases  $\{v^c \in \{1, \dots, C\}\}$ , the objective is to maximize the average log probability defined as

$$\sum_{c=1}^C \text{logp}(v^c) = \sum_{c=1}^C \log \frac{\sum_g e^{-E(v^c, g)}}{\sum_u \sum_g e^{-E(u, g)}} \quad (8)$$

The whole training process involves updating the weights with several numbers of epochs and the data is split in 20 batches which we take it randomly and the weights are update at the end of every batch. We use the binary representation of hidden units' activation pattern for classification and visualization. The activation of each hidden unit is defined as

$$f(x) = g(Wx + b) \quad (9)$$

Where  $g(z) = 1/(1 + \exp(-z))$  is the logistic sigmoid function, applied component-wise to the vector  $z$ ,  $W$  is a weight vector between visible nodes and hidden nodes and  $b$  is a bias. The autoencoder with  $N_h$  hidden nodes is trained and fine-tuned using back-propagation to minimize squared reconstruction error, with a term encouraging low average activation of the units.

#### 5. Extreme Machine Learning

The Extreme Learning Machine (Lin, M.B., Huang, G.B., Saratchandran P. and Sudararajan N. 2005, Huang, G.B., Zhu, Q.Y. and Siew, C.K. 2006, Huang, G.B., Zhu, Q.Y. and Siew, C.K. 2002, Mishra, Anurag, Singh, Lavneet and Chetty, Girija 2012, Singh, Lavneet and Chetty, Girija 2012 is a Single hidden Layer Feed forward Neural Network (SLFN) architecture. Unlike traditional approaches such as Back Propagation (BP) algorithms which may face difficulties in manual tuning control parameters and local minima, the results obtained after ELM computation are extremely fast, have good accuracy and has a solution of a system of linear equations. For a given network architecture, ELM does not have any control parameters like stopping criteria, learning rate, learning epochs etc., and thus, the implementation of this network is very simple. The main concept behind this algorithm is that the input weights (linking the input layer to the hidden layer) and the hidden layer biases are randomly chosen based on

some continuous probability distribution function such as uniform probability distribution in our simulation model and the output weights (linking the hidden layer to the output layer) are then analytically calculated using a simple generalized inverse method known as Moore – Penrose generalized pseudo inverse (Serre, D. 2002).

Given a series of training samples  $(x_i, y_i)_{i=1, 2 \dots N}$  and  $\hat{N}$  the number of hidden neurons where  $x_i = (x_{i1}, \dots, x_{in}) \in \mathbb{R}^n$  and  $y_i = (y_{i1}, \dots, y_{im}) \in \mathbb{R}^m$ , the actual outputs of the single-hidden-layer feed forward neural network (SLFN) with activation function  $g(x)$  for these  $N$  training data is mathematically modeled as

$$\sum_{k=1}^{\hat{N}} \beta_k g((w_k, x_i) + b_k) = 0_i, \forall i = 1, \dots, N \quad (10)$$

Where  $w_k = (w_{k1}, \dots, w_{kn})$  is a weight vector connecting the  $k^{\text{th}}$  hidden neuron,  $\beta_k = (\beta_{k1}, \dots, \beta_{km})$  is the hidden neuron. The weight vectors  $w_k$  are randomly chosen. The term  $(w_k, x_i)$  denotes the inner product of the vectors  $w_k$  and  $x_i$  and  $g$  is the activation function. The above  $N$  equations can be written as  $H\beta = O$  and in practical applications  $\hat{N}$  is usually much less than the number  $N$  of training samples and  $H\beta \neq Y$ , where

$$H = \begin{bmatrix} g((w_1, x_1) + b_1) & \cdots & g((w_{\hat{N}}, x_1) + b_{\hat{N}}) \\ \vdots & \ddots & \vdots \\ g((w_1, x_{1N}) + b_1) & \cdots & g((w_{\hat{N}}, x_N) + b_{\hat{N}}) \end{bmatrix}_{N \times \hat{N}}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{\hat{N}} \end{bmatrix}_{\hat{N} \times m} \quad O = \begin{bmatrix} O_1 \\ \vdots \\ O_N \end{bmatrix}_{N \times m} \quad Y = \begin{bmatrix} Y O_1 \\ \vdots \\ Y O_N \end{bmatrix}_{N \times m} \quad (11)$$

The matrix  $H$  is called the hidden layer output matrix. For fixed input weights  $w_k = (w_{k1}, \dots, w_{kn})$  and hidden layer biases  $b_k$ , we get the least-squares solution  $\hat{\beta}$  of the linear system of equation  $H\beta = Y$  with minimum norm of output weights  $\beta$ , which gives a good generalization performance. The resulting  $\hat{\beta}$  is given by  $\hat{\beta} = H^+ Y$  where matrix  $H^+$  is the Moore-Penrose generalized inverse of matrix  $H$  (Serre, D. 2002).

## 6. Trained Classifiers

In this study, apart from deep learning based on Restricted Boltzmann machines and extreme machine learning based on Single hidden Layer Feed forward Neural Network (SLFN) architecture as classifiers, several other classifiers are also examined in terms of accuracy and performance, including K-nearest neighbor (Wang, Jun. and Zucker, Daniel J. 2000), SVM (Vapnik, V. 1995), Naive Bayes George, H. and John, Pat Langley. 1995), MultiboostAB (Webb, Geoffrey. 2000), RotationForest (Rodriguez, Juan J., Kuncheva, Ludmila I. and Alonso, Carlos J. 2006), VFI (Quinlan, Ross. 1993), J48 (Breiman, Leo. 2001) and Random Forest (Hall, M. A. 1998).

J48 (Kohavi, Ron and John, George H. 1997) is an implementation of C4.5 algorithm that produces

decision trees from a set of labeled training data using the concept of information entropy. It examines the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data into smaller subsets. To make the decision, the attribute with the highest normalized information gain is used. The KNN algorithm (Wang, Jun. and Zucker, Daniel J. 2000) compares the test sample with the available training samples and finds the ones that are more similar (“nearest”) to it. When the  $k$ -nearest training samples are found, the class label in majority is assigned to the new sample. Learning in the VFI algorithm (Quinlan, Ross. 1993) is achieved by constructing feature intervals around each class for each attribute (basically discretization) on each feature dimension. Class counts are recorded for each interval on each attribute and classification is performed by a voting scheme.

The Naïve Bayesian Classifier (George, H. and John, Pat Langley. 1995) assumes that features are independent. Given the observed feature values for an instance and the prior probabilities of classes, the a posteriori probability that an instance belongs to a class is estimated. The class prediction is the class with the highest estimated probability. The SVMs (Vapnik, V. 1995) first map the attribute vectors into a feature space (possibly with higher dimensions), either linearly or nonlinearly, according to the selected kernel function. Then, within this feature space, an optimized linear division is sought; i.e., a hyper plane is constructed which separates two classes (this can be extended to multiple classes). MultiBoosting (Webb, Geoffrey. 2000) is an extension to the highly successful AdaBoost technique for forming decision committees. MultiBoosting can be viewed as combining AdaBoost with wagging. It is able to harness both AdaBoost's high bias and variance reduction with wagging's superior variance reduction. Using C4.5 as the base learning algorithm, Multi-boosting is demonstrated to produce decision committees with lower error than either AdaBoost or wagging significantly more often than the reverse over a large representative cross-section of data set. It offers the further advantage over AdaBoost of suiting parallel execution.

## 7. Feature Selection

In machine learning, during the training of the classifiers, if the numbers of image features are large, it can lead to ill-posing and over fitting, and reduce the generalization of the classifier. One way to overcome this problem is to reduce the dimensionality of features. To reduce the dimensionality of the large set of features of dataset, in our study, we propose the use of three optimal attribute selection algorithms: correlation based feature selection (CFS) method (Kohavi, Ron and John, George H. 1997), which evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them, secondly an approach based on wrappers (Hughes, N.P. and Tarassenko, L. 2003) which evaluates attribute sets by using a learning scheme. Also in this study, three search methods are

also examined: the Best First, Greedy Stepwise and Scatter Search algorithms. These search algorithms are used with attribute selector’s evaluators to process the greedy forward, backward and evolutionary search among attributes of significant and diverse subsets. In total, these feature selection algorithms were tested to select nearly 10 optimal and significant features out of 1024 features.

When we do PCA, we need to do an eigen-decomposition of the covariance matrix. The procedure of PCA is as follows:

1. Compute the mean:

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m \tilde{x}_i$$

2. Generate the zero-mean data matrix:

$$x_i = \tilde{x}_i - \bar{x}$$

$$A = (x_1, x_2, \dots, x_m)$$

3. Construct the covariance matrix:

$$C = AA^T$$

The covariance matrix C is symmetric and positive definite. So the eigenvalues of C is real and non-negative.

4. Eigen-decomposition:

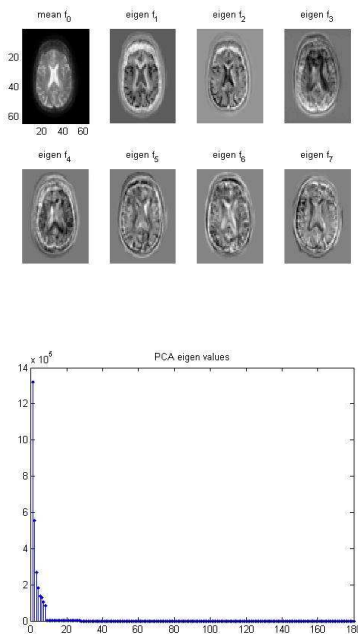
The eigenvalues  $\lambda_i$  and the eigenvectors  $v_i$  of C satisfy

$$Cv_i = \lambda_i v_i$$

5. So we have the eigen-decomposition of the covariance matrix:

$$C = V\Lambda V^{-1} = V\Lambda V^T$$

**Figure 3. (a) Eigen values of segmented MRI images (b) Eigen vectors after PCA**



## 8. Experiments and Results

### 8.1 Level of wavelet decomposition

We obtained wavelet coefficients of 60 brain MR images, each of whose size is 256 X 256. Level-1 HAR wavelet decomposition of a brain MR image produces 16384 wavelet approximation coefficients; while level-2

and level-3 produce 4096 and 1024 coefficients, respectively. The third level of wavelet decomposition greatly reduces the input vector size but results in lower classification percentage. With the first level decomposition, the vector size (16384) is too large to be given as an input to a classifier. The preliminary experimental analysis of the wavelet coefficients through simulation in Matlab 7.10., we showed that level-2 features are the best suitable for different classifiers, whereas level-1 and level-3 features results in lower classification accuracy. The second level of wavelet decomposition not only gives virtually perfect results in the testing phase, but also has reasonably manageable number of features (4096) that can be handled without much hassle by the classifier. We also use the DAUB-4 (Daubachies) as mother wavelets to get decomposition coefficients of MRI images at Level 2 for comparative evaluation of two wavelets decomposition methods in terms of classification accuracy.

### 8.2 Attribute Selection and Classification

Table 1 shows the accuracy of classification (percentage of correctly classified samples), True Positive Rate (TP), False Positive Rate (FP) and Average Classification Accuracy (ACC) over all pair-wise combination with different feature evaluators and search algorithms with respect to multi-class classification.

Table 1 shows the performance of several learning classifiers, including K-nearest neighbor, SVM, Naive Bayes, MultiboostAB, Rotation Forest, VFI, J48 and Random Forest. Among the pair-wise classification, the lowest accuracy is observed for the classification VFI classifiers of 74.16% and the highest accuracy for the classification by Rotational forest of 97.06%. Moreover, the combination of CFS feature evaluator with the of Best First search algorithm gives the highest classification accuracy.

While Table 1 shows the performance of individual classifiers, Table 2 defines the comparative results of various combined search techniques and feature evaluators using above prescribed classifiers. Table 3 compares the proposed method against a popular dimensionality reduction method, known as Principal Component Analysis (PCA). PCA applies an orthogonal linear transformation that transforms data to a new coordinate system of uncorrelated variables called principal components. We have applied PCA to reduce the number of attributes or feature to 18 attributes and plotted the ROC curves using several above mentioned learning classifiers in terms of True Positive and False Positive Rate, as seen in figure 4. As can be seen in figure 4, ROC curves for all the trained learning classifiers examined in this study, the curves lie above the diagonal line describing the better classification rather than any other random classifiers. The optimal points of various trained classifiers are indicated by bold solid circles as False Positive rate (FP) and True Positive rate (TP). These optimal points in ROC curves

show the maximum optimal value (FP, TP) of all trained classifiers.

**Table 1. Various Classifiers comparison with respect Average Classification Accuracy(%) and other parameters**

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ACC (%)
KNN	0.935	0.917	0.826	0.853	0.839	91.04
SVM	0.912	0.912	0.831	0.912	0.87	91.17
Naive Bayes	0.868	0.916	0.828	0.868	0.847	86.76
MultiboostAB	0.91	0.91	0.829	0.91	0.868	91.04
Rotation Forest	0.971	0.285	0.971	0.971	0.968	97.06
VFI	0.742	0.049	0.932	0.742	0.796	74.16
J48	0.96	0.314	0.958	0.96	0.957	95.98
Random Forest	0.97	0.271	0.97	0.97	0.968	97.01

**Table 2. Comparison of pair wise combination of various Attribute Selectors and classifiers with respect to ACC (%)**

Evaluator	Search Algorithm	Classifier	N	ACC (%)
CFS	Best First	K-NN	6	91.04
CFS	Greedy Stepwise	K-NN	2	89.70
CFS	Scatter Search	K-NN	4	88.23
Wrapper	Best First	K-NN	5	89.32
Wrapper	Greedy Stepwise	K-NN	4	87.56
Wrapper	Scatter Search	K-NN	4	88.20
CFS	Best First	SVM	6	91.17
CFS	Greedy Stepwise	SVM	6	89.23
CFS	Scatter Search	SVM	4	91.04
Wrapper	Best First	SVM	2	90.65
Wrapper	Greedy Stepwise	SVM	2	90.65
Wrapper	Scatter Search	SVM	5	89.56
CFS	Best First	Naive Bayes	8	86.76
CFS	Greedy Stepwise	Naive Bayes	8	82.78
CFS	Scatter Search	Naive Bayes	7	82.12
Wrapper	Best First	Naive Bayes	4	85.44
Wrapper	Greedy Stepwise	Naive Bayes	2	85.44
Wrapper	Scatter Search	Naive Bayes	2	80.12
CFS	Best First	MultiboostAB	5	91.04
CFS	Greedy Stepwise	MultiboostAB	5	91.04
CFS	Scatter Search	MultiboostAB	4	86.54
Wrapper	Best First	MultiboostAB	5	89.39
Wrapper	Greedy Stepwise	MultiboostAB	5	90.45
Wrapper	Scatter	MultiboostAB	4	88.76

	Search			
CFS	Best First	Rotation Forest	9	97.06
CFS	Greedy Stepwise	Rotation Forest	9	96.21
CFS	Scatter Search	Rotation Forest	8	91.66
Wrapper	Best First	Rotation Forest	5	93.78
Wrapper	Greedy Stepwise	Rotation Forest	6	93.78
Wrapper	Scatter Search	Rotation Forest	6	89.54
CFS	Best First	VFI	3	74.16
CFS	Greedy Stepwise	VFI	2	71.01
CFS	Scatter Search	VFI	4	71.01
Wrapper	Best First	VFI	3	72.22
Wrapper	Greedy Stepwise	VFI	2	72.85
Wrapper	Scatter Search	VFI	4	72.85
CFS	Best First	J48	7	95.98
CFS	Greedy Stepwise	J48	7	95.98
CFS	Scatter Search	J48	6	91.41
Wrapper	Best First	J48	7	95.98
Wrapper	Greedy Stepwise	J48	7	95.98
Wrapper	Scatter Search	J48	6	91.41
CFS	Best First	Random Forest	8	97.01
CFS	Greedy Stepwise	Random Forest	8	95.47
CFS	Scatter Search	Random Forest	8	95.47
Wrapper	Best First	Random Forest	5	96.25
Wrapper	Greedy Stepwise	Random Forest	6	96.25
Wrapper	Scatter Search	Random Forest	5	90.01

**Table 3. Classification Comparison using PCA and other feature attribute evaluators in terms of ACC (%)**

Classifier	PCA	CFS-Best First	Wrapper-Best First
KNN	91.38	91.04	89.32
SVM	96.24	91.17	90.65
Naive Bayes	85.63	86.76	85.44
MultiboostAB	94.52	91.04	89.39
Rotation Forest	97.06	97.06	93.78
VFI	77.12	74.16	72.22
J48	95.34	95.98	95.98
Random Forest	97.34	97.01	96.25

**Fig.4. Shows the ROC curve of the above mentioned trained classifiers.**

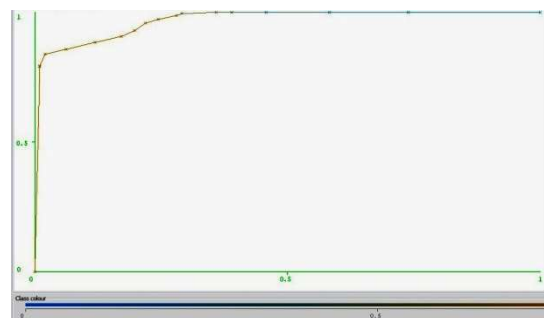


Table 4 describes the classification results using Extreme Machine Learning and Deep Machine Learning. In table 4, we compared the training time,



testing time and classification error using extreme and deep machine Learning. As we can see in the table both learning algorithms are processed to many hidden layers and their evaluations is done in terms of various factors. As depicted in Table 4, it clearly shows that deep machine learning plays a major role in reducing the classification error. As Deep and extreme machine learning are designed to work on large datasets for it is difficult to compare the performance. However, they result in acceptable accuracy levels, and we are currently examining several other publicly available large MRI datasets for enhancing the performance of these two novel approaches (Deep learning and Extreme machine learning approaches).

**Table 4. describes the classification results using Extreme Machine Learning and Deep Machine Learning.**

Hidden Layers	Training Time(s)			Classification Error		
	10	15	20	10	15	20
Deep Learning	0.56	0.47	0.72	0.083	0.065	0.071
Extreme Learning	0.31	0.31	0.61	0.042	0.042	0.061

The good factor in using deep and extreme learning classification algorithms is that the model is skipped by using dimension reduction evaluators and can be used on unlabeled datasets of MRI brain images where the ROI are classified as unlabeled and can be labeled and classified using these algorithms. But, still on small datasets of our current study is positive and encouraging in terms of low classification error and computation time for training and testing of data.

## 9. Conclusions

In this study, we have presented a principled approach for investigating brain abnormalities based on wavelet based feature extraction, PCA based feature selection and deep and extreme machine learning based classification comparative to various others classifiers. Experiments on a publicly available brain image dataset show that the proposed principled approach performs significantly better than other competing methods reported in the literature and in the experiments conducted in the study. The classification accuracy of more than 93% in case of deep machine learning and 94% in case of extreme machine learning demonstrates the utility of the proposed method. In this paper, we have applied this method only to axial T2-weighted images at a particular depth inside the brain. The same method can be employed for T1-weighted, proton density and other types of MR images. With the help of above approaches, one can develop software for a diagnostic system for the detection of brain disorders like Alzheimer's, Huntington's, Parkinson's diseases etc. Further, the proposed approach uses reduced data by incorporating feature selection algorithms in the processing loop and still provides an improved recognition and accuracy. The training and testing time for the whole study used by deep and extreme machine learning is much less as compared to SVM and other

traditional classifiers reported in the literature. Further work will be pursued to classify different type of abnormalities, and to extract new features from the MRI brain images on various parameters as age, emotional states and their feedback.

## 10. References

- Fletcher-Heath, L. M., Hall, L. O., Goldgof, D. B. and Murtagh, F.R. (2001): Automatic segmentation of non-enhancing brain tumors in magnetic resonance images; *Artificial Intelligence in Medicine* 21, pp. 43-63.
- Chaplot, S., Patnaik, L.M. and Jagannathan N.R. (2006): Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network; *Biomedical Signal Processing and Control* 1, pp. 86-92.
- Gorunescu, F. (2007): *Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection*; PWASET Volume 25 November, ISSN 1307-6884, PP. 427-430.
- Kara, S. and Dirgenali, F. (2007): A system to diagnose atherosclerosis via wavelet transforms, principal component analysis and artificial neural networks; *Expert Systems with Applications* 32, pp. 632-640.
- Maitra, M. and Chatterjee A. (2007): Hybrid multi-resolution Slantlet transform and fuzzy c-means clustering approach for normal-pathological brain MR image segregation, *MedEng Phys*, doi:10.1016/j.medengphy.2007.06.009.
- Abdolmaleki, P., Mihara, F., Masuda, K. and DansoBuadu, Lawrence. (1997): Neural networks analysis of astrocytic gliomas from MRI appearances *Cancer Letters* 118, pp. 69-78.
- Rosenbaum, T., Engelbrecht, V., Krolls, W. and Lenard, H. (1999): MRI abnormalities in neuro-bromatosis type 1 (NF1): a study of men and mice; *Brain & Development* 21, pp. 268-273.
- Cocosco, C., Zijdenbos, Alex P. and Evans, Alan C. (2003): A fully automatic and robust brain MRI tissue classification method; *Medical Image Analysis* 7, pp. 513-527.
- Harvard Medical School (1999): *The whole Brain Atlas*. [ONLINE] Available at: <http://www.med.harvard.edu/aanlib/home.html>.
- Hinton, G.E. and Salakhutdinov, R.R. (2006): Reducing the dimensionality of data with neural networks. *Science*, Vol. 313. No. 5786, pp. 504-507.
- Hinton, G.E. and Osindero, S. (2006): A fast learning algorithm for deep belief nets, *Neural Computation* 18, pp 1527-1554.



- Lin, M.B., Huang, G.B., Saratchandran P. and Sudararajan N. (2005): Fully complex extreme learning machine, *Neurocomputing*, vol (68), pp 306 – 314.
- Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2006): *Extreme Learning Machine: Theory and Applications*, *Neurocomputing*, vol (70), pp 489-501.
- Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2002): Real-Time Learning Capability of Neural Networks, *IEEE Transactions on Neural Networks*, vol 17(4), pp 863-878.
- Serre, D. (2002): *Matrices: Theory and Applications*, Springer Verlag, New York Inc.
- Wang, Jun. and Zucker, Daniel J. (2000): Solving Multiple-Instance Problem: A Lazy Learning Approach. In: 17th International Conference on Machine Learning, 1119-1125.
- Vapnik, V. (1995): *The Nature of Statistical Learning Theory*, Springer, New York.
- George, H. and John, Pat Langley. (1995): Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345.
- Webb, Geoffrey. (2000): *MultiBoosting: A Technique for Combining Boosting and Wagging*. Machine Learning. Vol.40 (No.2).
- Rodriguez, Juan J., Kuncheva, Ludmila I. and Alonso, Carlos J. (2006): Rotation Forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 28(10):1619-1630G.
- Quinlan, Ross. (1993): *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.
- Breiman, Leo. (2001): Random Forests. *Machine Learning*. 45(1):5-32.
- Hall, M. A. (1998): *Correlation-based Feature Subset Selection for Machine Learning*. Hamilton, New Zealand.
- Kohavi, Ron and John, George H. (1997): Wrappers for feature subset selection. *Artificial Intelligence*. 97(1-2):273-324.
- Hughes, N.P. and Tarassenko, L. (2003): Novel signal shape descriptors through wavelet transforms and dimensionality reduction. In *Wavelet Applications in Signal and Image Processing X*, pages 763–773.
- Mishra, Anurag, Singh, Lavneet and Chetty, Girija (2012): A Novel Image Water Marking Scheme Using Extreme Learning Machine, *Proceedings of IEEE World Congress on Computational Intelligence (WCCI 2012)*, Brisbane, Australia, IEEE Explore.
- Singh, Lavneet and Chetty, Girija (2012): Hybrid Approach in Protein Folding Recognition using Support Vector Machines, *Proceedings of International Conference on Machine Learning and Data Mining (MLDM 2012)*, Berlin, Germany, LNCS, Springer.
- Singh, Lavneet and Chetty, Girija (2012): Review of Classification of Brain Abnormalities in Magnetic Resonance Images Using Pattern Recognition and Machine Learning, *Proceedings of International Conference of Neuro Computing and Evolving Intelligence, NCEI 2012*, Auckland, New-Zealand, LNCS Bioinformatics, Springer.
- Singh, Lavneet and Chetty, Girija (2012): A Novel Approach for protein Structure prediction Using Pattern Recognition and Extreme Machine Learning, *Proceedings of International Conference of Neuro Computing and Evolving Intelligence, NCEI 2012*, Auckland, New-Zealand, LNCS Bioinformatics, Springer.
- Singh, Lavneet and Chetty, Girija (2012): Using Hybrid Neural Networks for Identifying the Brain Abnormalities from MRI Structural Images, *Proceedings of International Conference of Neuro and Image Processing (ICONIP 2012)*, LNCS, Springer.
- Singh, Lavneet and Chetty, Girija (2012): A Novel Approach to Protein Structure Prediction Using PCA or LDA Based Extreme Learning Machines, *Proceedings of International Conference of Neuro and Image Processing (ICONIP 2012)*, LNCS, Springer.
- Singh, Lavneet and Chetty, Girija (2012): Investigating Brain Abnormalities from MRI Images Using Pattern Recognition and Machine Learning Techniques, *Proceedings of International Conference on Pattern Recognition in Bioinformatics (PRIB 2012)*, LNCS, Springer.
- Singh, Lavneet and Chetty, Girija (2012): A Novel Approach to Protein Structure Prediction Using PCA Based Extreme Learning Machines and Multiple Kernels, *Proceedings of The 12th International Conference on Algorithm and Architectures for Parallel Processing (ICA3PP 2012)*, LNCS, Springer.

