

# AN EFFICIENT WAVELET BASED FEATURE REDUCTION AND CLASSIFICATION TECHNIQUE FOR THE DIAGNOSIS OF DEMENTIA

T.R. Sivapriya

Department of Computer Science, Lady Doak College, Madurai  
spriya.tr@gmail.com

## ABSTRACT

*This research paper proposes an improved feature reduction and classification technique to identify mild and severe dementia from brain MRI data. The manual interpretation of changes in brain volume based on visual examination by radiologist or a physician may lead to missing diagnosis when a large number of MRIs are analyzed. To avoid the human error, an automated intelligent classification system is proposed which caters the need for classification of brain MRI after identifying abnormal MRI volume, for the diagnosis of dementia. In this research work, advanced classification techniques using Support Vector Machines based on Particle Swarm Optimisation and Genetic algorithm are compared. Feature reduction by wavelets and PCA are analysed. From this analysis, it is observed that the proposed classification of SVM based PSO is found to be efficient than SVM trained with GA and wavelet based feature reduction technique yields better results than PCA.*

## KEYWORDS

*Classification, MRI, SVM, PSO, BPN, Wavelet, PCA.*

## 1. INTRODUCTION

Automated classification methods are commonly used for the analysis of neuroimaging studies. Several multiresolution approaches have been proposed to detect significant changes in the brain volume using neighbourhood information. Various computer-aided techniques have been proposed in the past and include the study of texture changes in signal intensity [1], grey matter (GM) concentrations differences [2], atrophy of subcortical limbic structures [3]–[5], and general cortical atrophy [6]–[8].

Magnetic Resonance Images are examined by radiologists based on visual interpretation of the films to identify the presence of tumour abnormal tissue. The shortage of radiologists and the large volume of MRI to be analysed make such readings labour intensive, cost expensive and often inaccurate. The sensitivity of the human eye in interpreting large numbers of images decreases with increasing number of cases, particularly when only a small number of slices are affected. Hence there is a need for automated systems for analysis and classification of such medical images. The MRI may contain both normal slices and defective slices. The defective or

abnormal slices are identified and separated from the normal slices and then these defective slices are further investigated for the detection of tumour tissues.

Brain image analyses have widely relied on univariate voxel-wise analyses, such as voxel-based morphometry (VBM) for structural MRI [9]. In such analyses, brain images are first spatially registered to a common stereotaxic space, and then mass univariate statistical tests are performed in each voxel to detect significant group differences. However, the sensitivity of these approaches is limited when the differences are spatially complex and involve a combination of different voxels or brain structures [10]. Recently, there has been a growing interest in support vector machines (SVM) methods [11, 12] to overcome the limits of these univariate analyses. These approaches allow capturing complex multivariate relationships in the data and have been successfully applied to the individual classification of a variety of neurological conditions [13-16].

Alzheimer's dementia (AD) is more prevalent today. The brain volume is significantly changed in Alzheimer's Dementia patients compared to healthy subjects of the same age group. Visual assessment of ventricle volume or shape change has shown to be quite reliable [17][18] in detecting AD. Zhu [19] applied Fourier analysis for image features, Ferrarini [20], Chaplot [21] applied wavelets, Selvaraj et al [22] applied Haralick components to extract features for brain MRI analysis.

Matthew C. Clarke et al. [23] developed a method for abnormal MRI volume identification with slice segmentation using Fuzzy C-means (FCM) algorithm. Luiza Antonie [24] proposed a method for Automated Segmentation and Classification of Brain MRI in which an SVM classifier was used for normal and abnormal slices classification with statistical features.

## **2. IMAGE FEATURE ANALYSIS**

### **2.1 TEXTURE ANALYSIS**

Texture is an image feature that provides important characteristics for surface and object identification from image [25-27]. Texture analysis is a major component of image processing and is fundamental to many applications such as remote sensing, quality inspection, medical imaging, etc. It has been studied widely for over four decades. Recently, multiscale filtering methods have shown significant potential for texture description, where advantage is taken of the spatial-frequency concept to maximize the simultaneous localization of energy in both spatial and frequency domains [28]. The use of wavelet transform as a multiscale analysis for texture description was first suggested by Mallat [29]. Recent developments in the wavelet transform [30, 31] provide good multiresolution analytical tool for texture analysis and can achieve a high accuracy rate.

### **2.2. FEATURE REDUCTION**

A standard method for reducing the dimensionality of medical images is principal components analysis (PCA), a data adaptive orthonormal transform whose projections have the property that for all values of  $N$ , the first  $N$  projections have the most variance possible for an  $N$ -dimensional subspace. PCA ignores spatial information. It treats the set of spectral images as an unordered set of high dimensional pixels. Wavelets are an efficient and practical way to represent edges and image information at multiple spatial scales. Image features at a given scale, such as region of interest can be directly enhanced by filtering the wavelet coefficients. For many tasks, wavelets may be a more useful image representation than pixels. The wavelet transform will take place

spatially over each image band, while the PCA transform will take place spectrally over the set of images. Thus, the two transforms operate over different domains.

### 2.2.1.PCA

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application [32-34], it is also named the discrete Karhunen–Loeve transform (KLT), the Hotelling transform or proper orthogonal decomposition (POD).

However, PCA over a *subset* of wavelet coefficients can be used to find eigenspectra that maximize the energy of that subset of wavelet coefficients. For example, PCA on only the vertical wavelet subbands will result in eigenspectra that maximize vertical wavelet energy. More generally, we use the term *Wavelet PCA* to refer to computing principal components for a masked or modified set of wavelet coefficients to find Wavelet PCA eigenspectra, and then projecting the original image onto the Wavelet PCA eigenspectra basis. In this way, features at a particular scale are indirectly emphasized by the computed projection basis, enhancing the reduced dimensionality images without filtering artifacts.

### 2.2.2. WAVELET APPLICATION FOR MULTIREOLUTION ANALYSIS AND DIMENSION REDUCTION

The wavelet transform (WT) has gained widespread acceptance in signal processing and image compression. Because of their inherent multi-resolution nature, wavelet-coding schemes are especially suitable for applications where *scalability* and *tolerable degradation* are important. Wavelet transform decomposes a signal into a set of basis functions. These basis functions are called *wavelets*. Wavelets are obtained from a single prototype wavelet  $y(t)$  called mother *wavelet* by *dilations* and *shifting*:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

where  $a$  is the scaling parameter and  $b$  is the shifting parameter

The wavelet transform is computed separately for different segments of the time-domain signal at different frequencies. Multi-resolution analysis analyzes the signal at different frequencies giving different resolutions. MRA is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. It is good for signal having high frequency components for short durations and low frequency components for long duration. e.g. images and video frames.

Since wavelet transform [35-43] has been successful in signal processing application a continuity relationship connecting (i-1), i and (i+1) feature sets is preferred, as it exists in case of signals. The localization property of wavelet transform has the capability of extracting the finer details from the spatial signals. These finer detail features are processed and approximated thus

extracting the knowledge at all levels of wavelet decomposition of the spatial signal. The DWT has been used for texture classification [44] and image compression [45] due to multiresolution decomposition property. The wavelet decomposition technique was also used to extract the intrinsic features for face recognition [46].

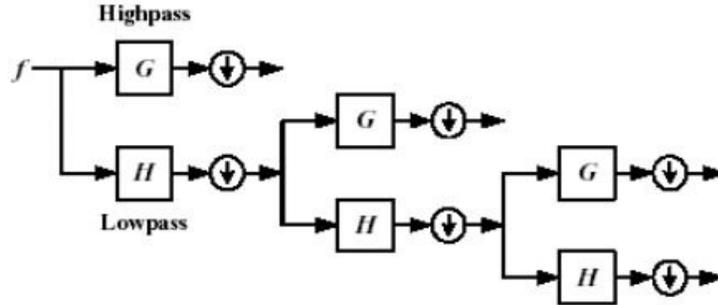


Figure 1. 2D DWT

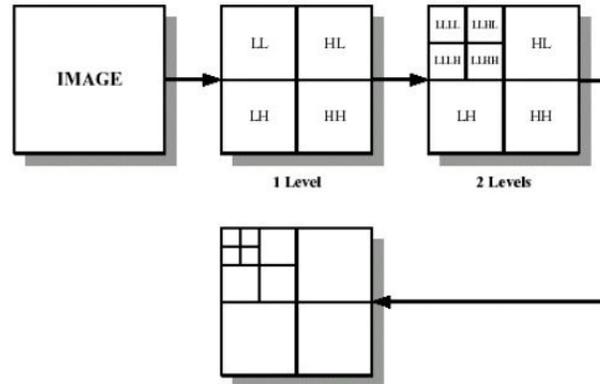


Figure 2. 2D DWT for Image

In wavelet transform the approximation from N dimensions to the lower dimension generates the approximation coefficients and the error vectors generate the corresponding detail coefficients. Thus the entire process of multi resolution knowledge mining involves analyzing the error vectors at different vector spaces  $V_k$  where  $k < N$  for being capable of representing a stable knowledge. In the process of dimensionality reduction we search for the appropriate error vector  $e_k$  in the lowest K dimensional space where  $K < N$  that would hold the optimum or enough detail for classifier and also extract the knowledge expressed by the error vectors at different dimensions.

### 3. TRAINING OF SVM

#### 3.1. SUPPORT VECTOR MACHINES

Support Vector Machines are new learning techniques that were introduced in 1995 by Vapnik [47]. In terms of theory the SVMs are well founded and proved to be very efficient in classification tasks. The advantages of such classifiers are that they are independent of the dimensionality of the feature space and that the results obtained are very accurate, although the training time is very high. Support Vector Machines (SVMs) are feedforward networks with a single layer of nonlinear units.

Their design has like the GRBF networks good generalization performance as an objective and follows for that reason the principle of *structural risk minimization* that is rooted in VC dimension theory. The solution for a typical two-dimensional case can have the form shown in Figure 1.

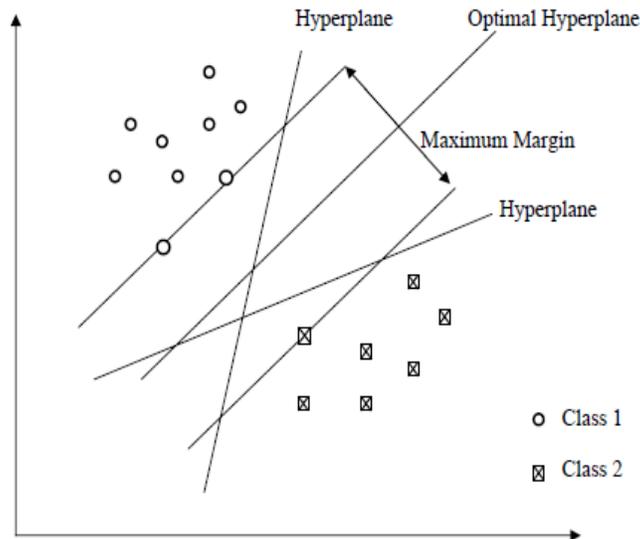


Figure 3. SVM

Those training points for which the equality in of the separating plan is satisfied (i.e.)  $y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i$ , those which wind up lying on one of the hyperplanes  $H_1, H_2$ , and whose removal would change the solution found, are called *Support Vectors (SVs)*.

This algorithm is firmly grounded in the framework of statistical learning theory – Vapnik Chervonenkis (VC) theory, which improves the generalization ability of learning machines to unseen data [48,49]. In the last few years Support Vector Machines have shown excellent performance in many real-world medical diagnosis applications [50] including object recognition, and diagnosis in medical [51, 52].

SVM guarantees the existence of unique, optimal and global solution since the training of SVM is equivalent to solving a linearly constrained QP. On the other hand, because the gradient descent algorithm optimizes the weights of BPN in a way that the sum of square error is minimized along the steepest slope of the error surface, the result from training may be massively multimodal, leading to non-unique solutions, and be in the danger of getting stuck in a local minima.

### 3.2. PARTICLE SWARM OPTIMISATION

Particle Swarm Optimisers (PSO) are a new trend in evolutionary algorithms, being inspired in group dynamics and its synergy. PSO had their origins in computer simulations of the coordinated motion in flocks of birds or schools of fish. As these animals wander through a three dimensional space, searching for food or evading predators, these algorithms make use of particles moving in an n-dimensional space to search for solutions for a variable function optimization problem.

In PSO individuals are called particles and the population is called a swarm used to train the classifiers with best values [53]. PSO are inspired in the intelligent behaviour of beings as part of an experience sharing community as opposed to an isolated individual reactive response to the environment.

### 3.3.SVM-PSO

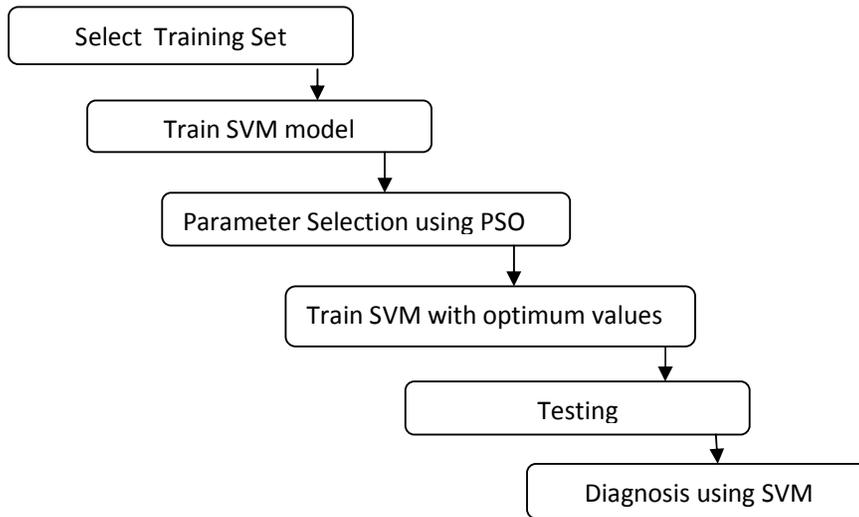


Figure 4. Training SVM with PSO

### 3.4. GENETIC ALGORITHM

The genetic algorithm (GA) is a popular optimization method that attempts to incorporate ideas of natural evolution. Its procedure improves the search results by constantly trying various possible solutions with some kinds of genetic operations. In general, the process of GA proceeds as follows.

First of all, GA generates a set of solutions randomly that is called an initial population. Each solution is called a chromosome and it is usually in the form of a binary string. After the generation of the initial population, a new population is formed that consists of the fittest chromosomes as well as offspring of these chromosomes based on the notion of survival of the fittest. The value of the fitness for each chromosome is calculated from a user-defined function. Typically, classification accuracy (performance) is used as a fitness function for classification problems. In general, offspring are generated by applying genetic operators. Among various genetic operators, selection, crossover and mutation are the most fundamental and popular operators. The selection operator determines which chromosome will survive. In crossover, substrings from pairs of chromosomes are exchanged to form new pairs of chromosomes. In mutation, with a very small mutation rate, arbitrarily selected bits in a chromosome are inverted. These steps of evolution continue until the stopping conditions are satisfied [54, 55].

### 3.5 OPTIMIZATION OF SVM USING GA

This study analyses training of SVM with Genetic Algorithm in which the kernel parameter settings of the SVM model are globally optimized, in order to improve prediction accuracy of typical SVM.

Step 1: Generate initial population to find optimum factors, kernel parameters. The chromosomes are initiated with random values. The values for feature selection and instance selection are set to '0' or '1' indicating rejection and selection respectively. The Gaussian Radial Basic Function is used as the Kernel function of SVM. The upper bound  $C$  and kernel parameter  $\delta$  are key variables that affect the performance of the SVM.

Step 2: Train the SVM process using the assigned value of the factors in the chromosomes, and calculate the performance of each chromosome. The performance of each chromosome can be calculated through the fitness function for GA. In this study, the main goal is to find the optimal or near optimal parameters that produce the most accurate prediction solution. The fitness function for the test data is set to the prediction accuracy of the test dataset.

Step 3: A new generation of the population is produced by applying genetic operators such as selection, crossover, and mutation. According to the fitness values for each chromosome, the chromosomes whose values are high are selected and used for the basis of crossover. The mutation operator is also applied to the population with a very small mutation rate. After the production of a new generation, Step 2 is performed again. From this point, Step 2 and Step 3 are iterated again and again until the stopping conditions are satisfied. When the stopping conditions are satisfied, the genetic search finishes and the chromosome that shows the best performance in the last population is selected as the final result.

Sometimes the optimized parameters determined by GA fit quite well with the test data, but they don't fit well with the unknown data. The phenomenon occurs when the parameters fit too well with the given test data set. Hence, in the last stage, the system applies the finally selected parameters – the optimal selections of features and instances, and the optimal kernel parameters – to the unknown data set in order to check the generalizability of the determined factors.

#### 4. METHODOLOGY

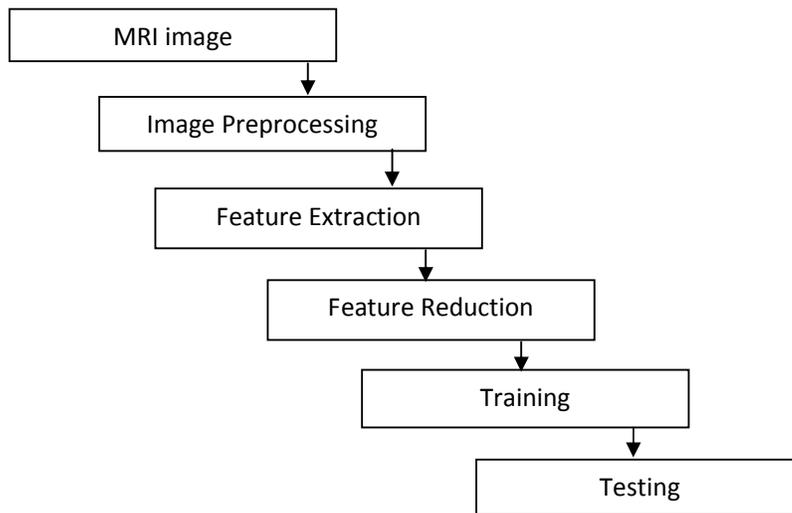


Figure 5. Methodology of the study

The MRI images are preprocessed and enhanced before feature extraction using wavelet based Haralick features. Image features are reduced using wavelets, PCA. Data is divided in to Training and Testing data according to Figure 5.

#### 4.1.FEATURE EXTRACTION PHASE AND CLASSIFICATION

For feature extraction, the following steps are done:

- The voxel-wise texture features of image I (x, y, z) are extracted at each slice of 3D ROI by convoluting with 2D Gabor filters [56,57]and averaging inside the ROI. The 2D Gabor filters are mathematically described at location (x, y) as  $\lambda = 1/f$  is the wavelength,  $\theta$  the orientation,  $\gamma$  the spatial aspect ratio which determines the eccentricity of the convolution kernel.
- A 32X32 gray level co-occurrence matrix with 32 gray levels are generated for the first level approximation images of each training image considered for feature extraction.
- Haralick features listed in Table-1 is calculated for each training image.
- For each texture feature compute the average of different angles.
- Compute the overall mean of each feature of the entire set of images considered for feature extraction.
- Feature reduction using wavelets, PCA
- Feed features as input vectors to SVM.
- Test the sensitivity, specificity and accuracy of the classification based on the texture features.

Table 1. Haralick features

Property	Formula
Angular second moment	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [p(i, j)]^2$
Correlation	$\frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} ij p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Inertia	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2 p(i, j)$
Entropy	$-\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i, j) \log_2 [p(i, j)]$
Absolute value	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1}  i-j  p(i, j)$
Inverse difference	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{p(i, j)}{1+(i-j)^2}$

#### 4.2. CROSS VALIDATION OF THE CLASSIFIERS

Cross Validation is a statistical analysis method used to verify the performance of classifiers. The basic idea is that the original dataset is divided into training datasets which are used for training

classifiers, and validation datasets for testing the trained models to obtain the classification accuracy as the evaluation performance of classifiers. This paper uses Leave-One-Out Cross Validation.

### **4.3. MRI DATASET**

OASIS provides brain imaging data that are freely available for distribution and data analysis. This data set consists of a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease (AD). For each subject, 3 or 4 individual T1-weighted MRI scans obtained within a single imaging session are included. The subjects are all right-handed and include both men and women. 100 of the included subjects over the age of 60 have been diagnosed with very mild to mild AD. Additionally, for 20 of the non-demented subjects, images from a subsequent scan session after a short delay (less than 90 days) are also included as a means of assessing acquisition reliability in the proposed study.

For each subject, a number of images are taken for analysis, including: 1) images corresponding to multiple repetitions of the same structural protocol within a single session to increase signal-to-noise, 2) an average image that is a motion-corrected co-registered average of all available data, 3) a gain-field corrected atlas-registered image to the 1988 atlas space of Talairich and Tournoux (Buckner et al., 2004), 4) a masked version of the atlas-registered image in which all non-brain voxels have been assigned an intensity value of 0, and 5) a gray/white/CSF segmented image (Zhang et al., 2001).

The SVM can be applied to different dimensional data by introducing a Kernel function to find a maximal margin hyper-plane in a high feature space that is well suited to the different classification problem structures. At the same time, it reduces the amount of training and testing, thereby increasing the classification accuracy for classification problems. The proposed method obtained the highest classification accuracy for the brain MRI classification problems. However, the number of features selected is less in the proposed method. This means that not all features are needed to achieve total classification accuracy. These results indicate that for different classification problems, the proposed method (binary particle swarm optimization) can serve as a pre-processing tool and help optimize the training process, which leads to an increase in classification accuracy. A good feature selection process reduces feature dimensions and improves accuracy.

## **5. RESULTS**

For SVMs, correct parameter adjustment is crucial, since many parameters are involved. This can have a profound influence on the results. For different classification problems, different parameters have to be set for SVMs. The two factors  $r$  and  $C$  are especially important. A suitable adjustment of these parameters results in a better classification hyper-plane found by the SVM, and thereby enhances the classification accuracy. Bad parameter settings affect the classification accuracy negatively.

The computation time used in PSO is less than in GAs. The parameters used in PSO are also fewer. However, suitable parameter adjustment enables particle swarm optimization to increase the efficiency of feature selection. Figure 6 indicates that Wavelet based feature reduction provides comparatively higher classification accuracy. Figures 7 and 8 indicate that SVM trained with PSO has higher accuracy and sensitivity compared to GA based SVM.

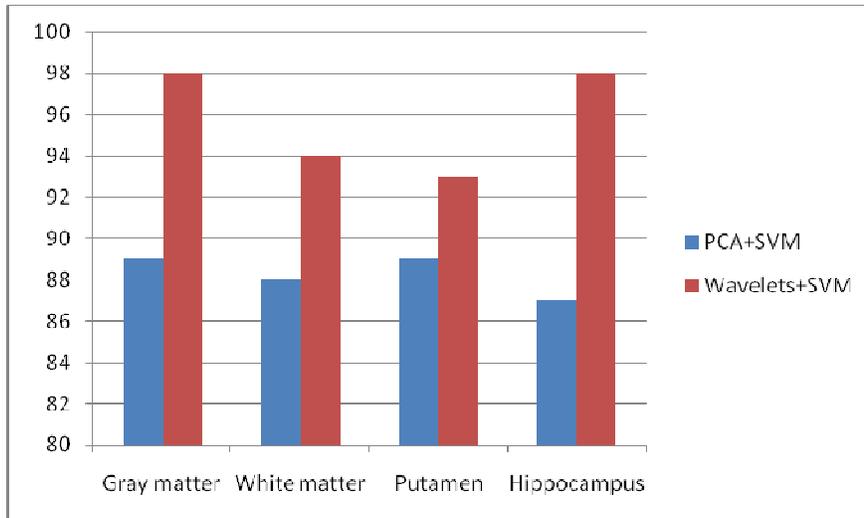


Figure 6. Efficiency of Wavelet and PCA feature reduction techniques

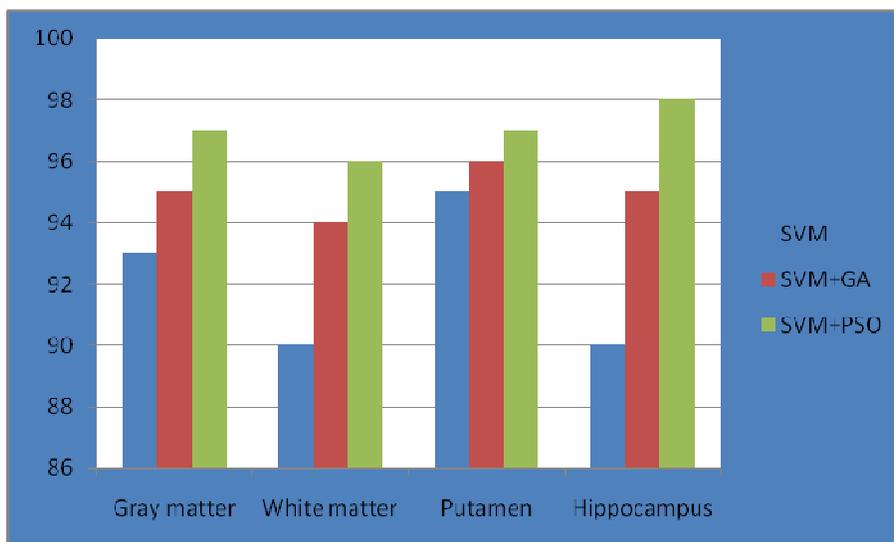


Figure 7. Efficiency of Classifiers for the Longitudinal OASIS database

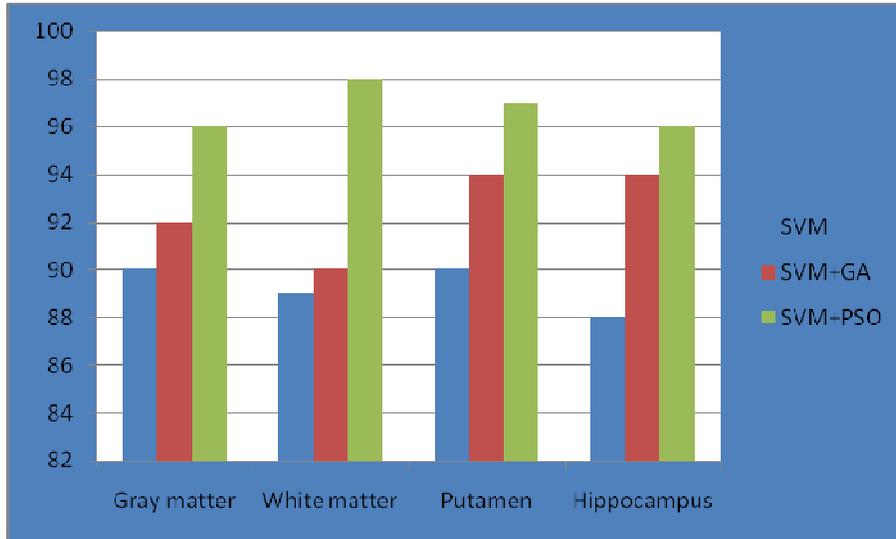


Figure 8. Efficiency of Classifiers for the Cross-sectional OASIS database

## 6. CONCLUSION

Building an efficient classification model for classification problems with different dimensionality and different sample size is important. The main tasks are the selection of the features and the selection of the classification method. In this paper, feature reduction using wavelets as well as PCA is analysed. Wavelet based feature reduction preserves the important details as well as reduces the features effectively than PCA. Combining wavelet with SVM provides better classification than with PCA.

The performance of SVM trained with PSO is compared with SVM trained by Genetic Algorithm. Experimental results show PSO based SVM method reduced the total number of parameters needed effectively, thereby obtaining a higher classification accuracy compared GA based SVM. The proposed method can serve as an ideal pre-processing tool to help optimize the feature selection process, since it increases the classification accuracy and, at the same time, keeps computational resources needed to a minimum.

## REFERENCES

- [1] P. A. Freeborough and N. C. Fox (1998), "MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 475–479.
- [2] G. B. Frisoni, C. Testa, A. Zorzan, F. Sabbatoli, A. Beltramello, H. Soininen, and M. P. Laakso (2002), "Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry," *J. Neurol. Neurosurg. Psychiatry*, vol. 73, pp. 657–664.
- [3] P. M. Thompson, K. M. Hayashi, G. I. De Zubicaray, A. L. Janke, S. E. Rose, J. Semple, M. S. Hong, D. H. Herman, D. Gravano, D. M. Doddrell, and A. W. Toga (2004) , "Mapping hippocampal and ventricular change in alzheimer disease," *NeuroImage*, vol. 22, pp. 1754–1766.
- [4] G. B. Frisoni, F. Sabbatoli, A. D. Lee, R. A. Dutton, A. W. Toga, and P. M. Thompson (2006), "In vivo neuropathology of the hippocampal formation in AD: A radial mapping MR-based study," in *NeuroImage*.

- [5] J. G. Csernansky, L. Wang, S. Joshi, J. P. Miller, M. Gado, D. Kido, D. McKeel, J. C. Morris, and M. I. Miller (2000), "Early DAT is distinguished from aging by high-dimensional mapping of the hippocampus. Dementia of the alzheimer type," *Neurology*, vol. 55, pp. 1636–1643.
- [6] P. M. Thompson, K. M. Hayashi, G. de Zubicaray, A. L. Janke, S. E. Rose, J. Semple, D. Herman, M. S. Hong, S. S. Dittmer, D. M. Doddrell, and A. W. Toga (2003), "Dynamics of gray matter loss in Alzheimer's disease," *J. Neurosci.*, vol. 23, pp. 994–1005.
- [7] D. Chan, J. C. Janssen, J. L. Whitwell, H. C. Watt, R. Jenkins, C. Frost, M. N. Rossor, and N. C. Fox (2003), "Change in rates of cerebral atrophy over time in early-onset alzheimer's disease: Longitudinal MRI study," *Lancet.*, vol. 362, pp. 1121–1122.
- [8] J. P. Lerch, J. C. Pruessner, A. Zijdenbos, H. Hampel, S. J. Teipel, and A. C. Evans (2005), "Focal decline of cortical thickness in alzheimer's disease identified by computational neuroanatomy," *Cereb Cortex*, vol. 15, pp. 995–1001.
- [9] J. Ashburner and K.J. Friston (2000), "Voxel-based morphometry—the methods. *NeuroImage*", 11(6):805–21.
- [10] C. Davatzikos (2004). Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, 23(1):17–20.
- [11] V.N. Vapnik (1995), *The Nature of Statistical Learning Theory*. Springer-Verlag.
- [12] B. Scholkopf and A.J. Smola. *Learning with Kernels* (2001), MIT Press.
- [13] Z. Lao et al (2004). Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, 21(1):46–57.
- [14] Y. Fan et al. (2007) COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE TMI*, 26(1):93–105.
- [15] S. Klöppel et al. (2008) Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(3):681–9.
- [16] P. Vemuri et al. (2008) Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *NeuroImage*, 39(3):1186–97.
- [17] Mahmoud-Ghoneim, D., Toussaint, G., Constans JM., et al. (2003) "Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas". *Magn Reson Imaging* ;21:983–87.
- [18] Snyder, AZ., Girton, LE., Morris, JC., Buckner, RL. (2009), Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, 64: 1032-1039.
- [19] Zhu, H., Goodyear, B. G., Lauzon, M. L., Brown, R. A., et al (2003). A new local multiscale Fourier analysis for medical imaging. *Med. Phys.*, 30: 1134-1141.L.
- [20] Ferrarini, WM. Palm, H. Olofsen, MA. Van Buchem, JH. Reiber, F. Admiraal-Behloul (2006), Shape Differences of the brain ventricles in Alzheimer's Disease, *Neuroimage*.
- [21] Chaplot, S., Patnaik, L.M., 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*.
- [22] Selvaraj, H., Thamarai Selvi, S., Selvathi, D., Gewali, L. (2007), Brain MRI Slices Classification Using Least Squares Support Vector Machine. *IC-MED*, Vol. 1, No. 1, Issue 1, Page 21 of 33.

- [23] M. C. Clark, L. O. Hall, D. B. Goldgof, L. P. Clarke, R. P. Velthuizen, and M. S. Silbiger (1994), "MRI Segmentation using Fuzzy Clustering Techniques", IEEE Engineering in Medicine and Biology, pp. 730-742.
- [24] Automated Segmentation and Classification of Brain Magnetic Resonance Imaging – by Luiza Antonie <http://www.cs.ualberta.ca/~luiza/c615/proj.pdf>.
- [25] Smith, G., and Burns, I. (1997), Measuring texture classification algorithms, Pattern Recognition Letters, Vol. 18, pp. 1495-1501, <http://www.cssip.elec.uq.edu.au/guy/meastex/meastex.html>.
- [26] Arivazhagan, S., Ganesan, L., and Priyal, S.P. (2006), "Texture classification using Gabor wavelets based rotation invariant features," Pattern recognition letters, 27(16): 1976-1982.
- [27] Vassili, A. Kovalev, Frithjof Kruggel, Hermann-Josef Gertz, D. Yves von Cramon, D. (2001): IEEE Three-Dimensional Texture Analysis of MRI Brain Datasets, Transactions on Medical Imaging, Vol. 20, No. 5.
- [28] Valkealahti, K., Oja, E. (1998): Reduced multidimensional cooccurrence histograms in texture classification, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, , pp. 90-94.
- [29] Mallat, S.G. (1989), "Multifrequency channel decompositions of images and wavelet models, IEEE Trans. on Acoustics Speech And Signal Processing, Vol. 11, No. 7, pp. 674-693,.
- [30] Gonzalez, R.C., Woods, R.E. (2002), Digital Image Processing (2nd Edition). Prentice Hall.
- [31] T.R. Sivapriya, V. Saravanan, P. Ranjit Jeba Thangaiah Texture analysis of brain MRI and classification with BPN for the diagnosis of dementia, CCSEIT 2011, ISBN 978-3-642-24042-3, pp. 555-565.
- [32] Wen Ge, Xu Hongzhe, Zheng Weibin et al., (2009), Multi-kernel PCA based high-dimensional images feature reduction, International Conference on Electric Information and Control Engineering (ICEICE), 2011, ISBN: 978-1-4244-8036-4, doi:10.1109/ICEICE.2011.5778352, pp: 5966 – 5969.
- [33] Yihui Luo, Shuchu Xiong, Sichun Wang, (2008) A PCA Based Unsupervised Feature Selection Algorithm, International Conference on Genetic and Evolutionary Computing, WGEC '08. ISBN: 978-0-7695-3334-6, doi:10.1109/WGEC.2008.109, pp: 299-302.
- [34] G. C. Feng, P. C. Yuen, and D. Q. Dai, (2000), Human face recognition using PCA on wavelet subband, Journal of Electronic Imaging, vol. 9, no. 2, pp. 226–233.
- [35] F. Abramovich, T. Bailey, and T. Sapatinas (2000), "Wavelet analysis and its statistical applications" JRSSD, (48):1–30, 2000.
- [36] A. Antoniadis and G. Oppenheim, editors (1995), Wavelets and Statistics, Lecture Notes in Statistics. Springer-Verlag.
- [37] Y. Meyer (1993) Wavelets—Algorithms and Application. SIAM.
- [38] R. Young (1993) "Wavelet Theory and its Application" Kluwer Academic Publishers, Boston.
- [39] D. Keim and M. Heczko (2001). Wavelets and their applications in databases. Tutorial Notes of ICDE 2001.
- [40] D. Hand, H. Mannila, and P. Smyth (2001). Principles of Data Mining. The MIT Press.
- [41] S. Mallat (1989), "A theory for multiresolution signal decomposition: the wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7):674–693.

- [42] R.Polikar. The wavelet tutorial. Internet Resources:  
<http://engineering.rowan.edu/polikar/WAVELETS/WTtutorial.html>.
- [43] I. Daubechies (1992), Ten Lectures on Wavelets. Capital City Press, Montpelier, Vermont.
- [44] T. Chang and C.-C.J. Kuo (1993), "Texture Analysis and Classification with Tree-Structured Wavelet Transform" IEEE Trans. Image Processing, vol. 2, no. 4,pp. 429-441.
- [45] A. Averbuch, D. Lazar, and M. Israeli (1996), "Image Compression Using Wavelet Transform and Multiresolution Decomposition" IEEE Trans. Image Processing, vol. 5, no. 1.
- [46] R. Foltyniewicz (1996), "Automatic Face Recognition via Wavelets and Mathematical Morphology," Proc. Int'l Conf. Pattern Recognition, pp. 13-17.
- [47] V. N. Vapnik (1995), "The Nature of Statistical Learning Theory" Springer-Verlag, New York .
- [48] N. Tsiaparas et al (2011), Comparison of Multiresolution Features for Texture Classification of Carotid Atherosclerosis From B- Mode Ultrasound IEEE Transactions on Information Technology in Biomedicine, Volume:15, Issue:1, Page(s): 130 – 137.
- [49] Ulacl Bagcl, Li Bai (2007), A comparison of daubechies and gabor wavelets for classification of mr images,IEEE International Conference on Signal Processing and Communications (ICSPC 2007), 24-27.
- [50] Vemuri P, Wiste HJ, Weigand SD, Knopman DS, Trojanowski JQ, Shaw LM, Bernstein MA, Aisen PS, Weiner M, Petersen RC, Jack CR Jr (2010), Alzheimer's Disease Neuroimaging Initiative: Serial MRI and CSF Biomarkers in Normal Aging, MCI and AD. Neurology 2010, 75:143-151
- [51] Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, et al. (2008) Alzheimer's disease diagnosis in individual subjects using structural mr images: Validation studies. NeuroImage 39: 1186–1197.
- [52] S. Vibha, S.Vyas and Priti Rege (2006), "Automated Texture Analysis with Gabor filter", GVIP Journal, Volume 6, Issue 1.
- [53] S. Sivakumari, R. Praveena Priyadarsini, P. Amudha (2009). Performance evaluation of svm kernels using hybrid PSO-SVM, ICGST-AIML Journal, ISSN: 1687-4846, Volume 9, Issue I, pp:19-25.
- [54] Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A., & Jain, A. K. (2000). Dimensionality reduction using genetic algorithms. IEEE Transactions on Evolutionary Computation, 4(2), pp:164–171.
- [55] Oh, I.-S., Lee, J.-S., and Moon (2004), B.-R. "Hybrid Genetic Algorithms for Feature Selection". IEEE Trans. Pattern Analysis and MachineIntelligence, vol. 26, no.11.
- [56] Mozer, M., Jordan, M., & Petsche, T., (Eds.) (1997), Advances in Neural Information Processing Systems (pp. 475–481). Cambridge, MA: MIT Press.
- [57] Randen, T., Husoy, J.H. (1999), "Filtering for Texture Classification: A Comparative Study", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 21, no. 4, pp. 291-1.