

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/323907323>

# Alzheimer's Disease Detection Using Extreme Learning Machine, Complex Dual Tree Wavelet Principal Coefficients and Linear Discriminant Analysis

**Preprint** · June 2018

DOI: 10.1166/j.mhi.2018.2381

CITATIONS

0

READS

195

**5 authors**, including:



**Debesh Jha**

Chosun University

**15** PUBLICATIONS **21** CITATIONS

[SEE PROFILE](#)



**Saruar Alam**

Chosun University

**10** PUBLICATIONS **7** CITATIONS

[SEE PROFILE](#)



**Jae-Young Pyun**

Chosun University

**123** PUBLICATIONS **616** CITATIONS

[SEE PROFILE](#)



**Goo-Rak Kwon**

Chosun University

**58** PUBLICATIONS **135** CITATIONS

[SEE PROFILE](#)

**Some of the authors of this publication are also working on these related projects:**



Alzheimer's Disease [View project](#)



# Alzheimer's Disease Detection Using Extreme Learning Machine, Complex Dual Tree Wavelet Principal Coefficients and Linear Discriminant Analysis

Debesh Jha<sup>1</sup>, Saruar Alam<sup>1</sup>, Jae-Young Pyun<sup>1</sup>, Kun Ho Lee<sup>2</sup>, and Goo-Rak Kwon<sup>1,\*</sup>

<sup>1</sup>Dept. of Information and Communication Engineering, Chosun University, 375 Seosuk-Dong, Dong-Gu, Gwangju 501-759, Republic of Korea

<sup>2</sup>National Research Center for Dementia and Department of Biomedical Science, Chosun University, Gwangju, 375 Seosuk-Dong, Dong-Gu, Gwangju 501-759, Republic of Korea

The early detection and classification of Alzheimer's disease (AD) are important clinical support tasks for medical practitioners in customizing patient treatment programs to have better manage the development and progression of these diseases. Efforts are being made to diagnose these neurodegenerative disorders in the early stages. Efficient early categorization of the AD and mild Cognitive Impairment (MCI) from HC is necessary as prompt preventive care could assist to mitigate risk factors. For analysis and prognosis of disease, Magnetic resonance imaging (MRI). In this paper, we proposed a novel computer-aided diagnosis (CAD) cascade model to discriminate patients with the AD from healthy controls using dual-tree complex wavelet transforms (DTCWT), principal component analysis, linear discriminant analysis, and extreme learning machine (ELM). The proposed method obtained accuracy of  $90.26 \pm 1.17$ , a specificity of  $90.20 \pm 1.56$  and sensitivity of  $90.27 \pm 1.29$  on the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset and accuracy of  $95.72 \pm 1.54$ , a sensitivity of  $96.59 \pm 2.34$  and specificity of  $93.03 \pm 1.67$  on the Open Access Series of Imaging Studies (OASIS) dataset. The proposed method is effective and superior to the existing models.

**Keywords:** Alzheimer's Disease, Computer-Aided Diagnosis, Dual-Tree Complex Wavelet Transform, Principal Component Analysis, Linear Discriminant Analysis, Extreme Learning Machine, Alzheimer's Disease Neuroimaging Initiative, Open Access Series of Imaging Studies.

## 1. INTRODUCTION

The most widely recognized reason for dementia is Alzheimer's disease (AD), where it is proved by the fact that about 50% to 80% of all dementia victims are sufferers of this disease. The illness influences people's memorization, cognizance, and actions. Various kinds of degeneration transpire in the hippocampus and other different parts of the brain since the AD is a neurodegenerative illness. Even though it is not a usual disease, but still holds the 6th position amongst the prominent reasons for death in the USA. As of now, there is no cure for this disease; nevertheless, few precautionary steps can be taken to reduce the risk factors and decelerate the retrogressive growth. According to a survey, approximately \$605 billion universally and \$220 billion in the USA is spent every year on diagnosing AD. Many individuals experience the ill effects of AD around the world and requests

on analysts are developing promptly. MRI is a compelling therapeutic image development method, as it has the demonstrated potential to see architectural changes in the human brain, interior organs, and different tissues.

MRI creates finest basic images, giving unique tissue data, which upgrades both the exactness of brain pathology analysis and nature of treatment. A key preferred standpoint of this procedure is its non-invasiveness. Many types of research have been directed utilizing multivariate investigation calculations and architectural/functional MRI to group neurological sicknesses.<sup>1-3</sup> An essential concentration of these examinations was the substantial dimensionality of extracted characteristics and the distinguishing proof of disease signs among them where most the discriminative data of said illnesses exists. Results demonstrated huge cerebral structural changes in a few cerebrum ROIs, especially in the hippocampus and entorhinal cortex.<sup>4</sup> Cosmic and inner intensity-based characteristics,<sup>3,5</sup> and additionally

\* Author to whom correspondence should be addressed.

geometric and surface-based highlights,<sup>6,7</sup> have utilized as a part of prior investigations for categorizing disease. The authors showed an electroencephalogram (EEG) intelligibility analysis of Alzheimer's disease by utilizing a probabilistic neural network (PNN) and demonstrated critical precision in recognizing genuine AD from the control classes.<sup>8</sup> Sandeep et al.<sup>9</sup> stratified AD utilizing discrete wavelet coefficients as an element for preparing and testing Support Vector Machines (SVM) and neural network classifiers. Getting fundamental biased features from MRI images is basic for a capable study of disease analysis. The favored element extraction techniques, among those most commonly utilized, are independent component analysis,<sup>10</sup> wavelet transform,<sup>11</sup> and Fourier transform.<sup>12</sup> This examination has been led utilizing discrete wavelet features and the k-nearest neighbor technique (k-NN)<sup>12</sup> on an Artificial Neural Network (ANN).<sup>12</sup> Zhang and Wang<sup>14</sup> ran AD prediction models using displacement field estimation between AD and healthy controls using some categorizers such as-SVM, Generalized Eigenvalue Proximal SVM (GEP-SVM) and Twin Support Vector Machine (TWSVM). Tomar et al.<sup>15</sup> looked into a few sorts of twin SVM techniques, their advancement issues, and their applications. Siyuan Lu et al.<sup>34</sup> have examined pathological brain recognition applying Extreme Learning Machine (ELM) by bat technique. Jha et al.<sup>13</sup> utilized Wiener-filtering, 2D-Discrete Wavelet Transform, Probabilistic PCA, and Random Subspace Ensemble Classifier for classification of pathological brain images on a Harvard dataset. Lama<sup>37</sup> et al. has used structural MRI Images using a regularized extreme learning machine and PCA features for diagnosis of Alzheimer's disease. Jha et al.<sup>38</sup> designed an efficient cascade model for Pathological Brain Image Detection by Magnetic Resonance Imaging for the pathological brain detection.

The biomarkers utilized as a part of our proposed strategy are MRI images from the ADNI and OASIS datasets. Our essential purpose behind utilizing DTCWT over DWT is its viable portrayal of singularities (curves and lines), despite the fact that DWT has the benefit of representing the functions in multi-scale and packed structures. In DTCWT, changes in magnitude variance can be accomplished to a higher degree.<sup>16</sup> In our recommended strategy, DTCWT coefficient-based AD grouping has been proposed utilizing primary segment examination and direct discriminant investigation of separated coefficients; an ELM was used as a regulating process. Classification efficiency is reported with respect to precision, affectability, and specificity, after implementing 10-fold cross validation and execute the process for 10 to 20 times. Our technique delivered prevalent outcomes when contrasted with few other traditional AD classification techniques.

## 2. MATERIALS AND METHODS

Total 172 cases from the dataset of ADNI were used; 86 AD and 86 HC. Moreover, we utilized 95 subjects from the OASIS dataset; 98 HC and 28 subjects experiencing AD.

### 2.1. Overview of Experimental Data

Information utilized as a part of the readiness of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

The ADNI was propelled in 2003 as a public-private organization drove by Principal Investigator Michael W. Weiner, MD.

**Table I. Summary of subject's demographics status.**

	AD	Normal
No. of subjects	86	86
	Male-43, Female-43	Male-46, Female-40
Average age	77.30	76.05
Average education points	14.65	15.93
MMSE	23.48	29.08

To quantify the movement of MCI and early-beginning Alzheimer's Disease (AD), the main objective of the ADNI dataset is to verify in case Positron Emission Tomography (PET), serial MRI, other biological attributes, and clinical and neuropsychological evaluation can be consolidated.

For up-to-date information, visit [www.adni-info.org](http://www.adni-info.org). (For upgraded data, visit [www.adni-info.org](http://www.adni-info.org)). Table I is representing the demographical analysis of data which are utilized from the ADNI dataset.

(Moreover, we used downloaded MRI images of OASIS dataset). (OASIS is a database intended to accumulate MRI datasets and make them uninhibitedly available to mainstream researchers). (OASIS accumulates two sorts of information: cross-sectional MRI information and longitudinal MRI information). (Our study used cross-sectional MRI information, as our main points are to build up a programmed framework for recognizing the AD, for which longitudinal MRI information is not ideal).

(The OASIS dataset comprises of 416 Subjects aged in the vicinity of 18 and 96 years. Our examination contained 51 AD patients (35 with CDR = 0.5 and 16 with CDR = 1) among 100 subjects having dementia, and 44 HC among 98 typical subjects). (Table II demonstrates the statistic points of interest of the subjects utilized as a part of our research. Both men and women are incorporated and all subjects are right-handed. The size of the CDR is recorded in Table III).

### 2.2. Proposed Approach

The proposed method comprises of four stages: preprocessing and slice extraction, characteristic extraction, the projection of features into bringing down measurement, and productive characterization of the disease. Figure 4 demonstrates all stages in detail.

#### 2.2.1. Preprocessing and Slice Extraction

(All MRI images utilized for training and testing the ELM of our proposed method are seen utilizing the ONIS toolbox and traded as 2D MRI image cuts. All pictures are in PNG format, and the measurements of OASIS picture cuts are  $176 \times 208$ ; the measurements of the ADNI picture cuts are  $256 \times 166$ . The scope of determination of those slices was performed physically from

**Table II. Statistical OASIS data details used in our learning.**

Factors	Normal	AD
No. of patients	98	28
Age	84.40 (76–96)	82.11 (76–96)
Education	3.34 (1–5)	3.13 (1–5)
Socioeconomic status	2.31 (1–5)	2.82 (1–5)
CDR (0/0.5/1)	98/0/0	0/0/28
MMSE	28.72 (25–30)	24.82 (18–30)

Table III. Clinical dementia scale.

CDR	Rank
0.5	Very mild dementia
1	Mild
2	Moderate
3	Severe

the tissue center for data clarity. The images are resized to  $256 \times 256$  for further processing. An example of a brain image slice is depicted in Figure 1. We have utilized LibSVM toolbox for kernel SVM simulation in the MATLAB.

### 2.2.2. Dual Tree Complex Wavelet Transform

The wavelet transform is a standout amongst the most every now and again utilized feature extraction procedures for MRI images. For boosting directional selectivity and impaired by DWT, DTCWT is suggested. The main motivation beyond utilizing DTCWT is it gives the solution for “shift-invariant problems” and directional selectivity in two directions or more dimensions,” that is the drawback of a traditional DWT.

For our proposed method, the DTCWT<sup>16</sup> coefficients were extracted from the input MRI images. The highlights of the fifth resolution scale were utilized as they created higher classification execution when contrasted and other determination levels. DTCWT has a multi-resolution portrayal, as with CWT). For proficient disease characterization, it is desirable to utilize a couple of intermediate scales of the extricated coefficients as input to a categorizer, as the least resolution scales lose fine points of interest and high-resolution scales often noisy. Accordingly, we like to pick a couple of moderate sizes of DTCWT coefficients. These coefficients were used as input data for principal component analysis (PCA). CWT can be presented as complex-valued scaling capacities and complex-valued wavelets. DTCWT draws in two genuine DWTs, which individually provides the real and imaginary component of the wavelet transform. Moreover, two-channel bank types are set: analysis channel banks and synthesis channel banks. These filter banks are utilized for executing DTCWT to guarantee that general transformation turns out to be roughly analytic as shown in Figure 2.

The DTCWT can be designated in matrix form as:

$$D = [D_h \ D_g] \quad (1)$$

Where,  $D_h, D_g$  are rectangular matrices.

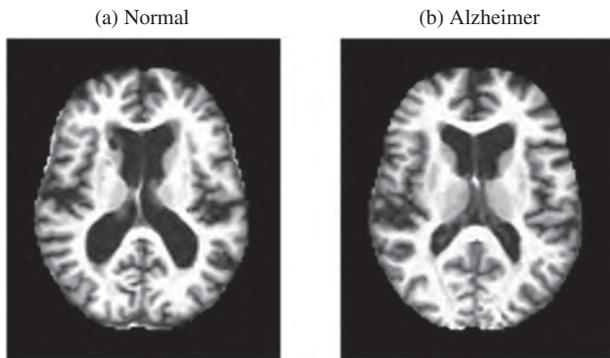


Fig. 1. MR Image slice sample (Axial slice view after preprocessing). (a) Normal, (b) Alzheimer.

For the input data image  $x$ , complex wavelet, coefficients can be presented as:

$$T_h + jT_g \quad (2)$$

where  $T_h = D_h * x$  is the real part, and  $T_g = D_g * x$  is the imaginary part.

The DTCWT coefficients of input images are constant while shifting; even if an image is switched in time or space, the coefficients do not change. Moreover, DTCWT utilizes isolation of 6 distinct headings ( $\pm 15, \pm 30, \pm 45$ ) for 2D pictures and 28 unique bearings for 3D pictures, while regular DWT takes into account confinement of horizontal and vertical directions). We have drawn out 5 levels of DTCWT coefficients from one ratio for each of the 2-dimensional piece image subject.

### 2.2.3. Principal Component Analysis

Primary Component Analysis (PCA)<sup>17</sup> is a dimensionality lessening strategy that is used to map features onto lower dimensional space. The transformation of data can be linear or nonlinear. One of the most commonly utilized linear-transformation is PCA that is an orthogonal transformation utilized to change over conceivably connected samples to directly uncorrelated factors. The number of principal components is not exactly or equivalent to the quantity of unique factors). The PCA transformation process is demonstrated in Figure 3.

The PCA is compiled as:

- (i) Getting the average of the initial dataset and zero mean matrix.
- (ii) Constructing the covariance matrix of the resulting matrixes form step (i).
- (iii) Calculating the eigenvalue and the eigenvector.
- (iv) Projecting the data matrix accomplished by eigenvectors and from the highest to lowest eigenvalues.

### 2.2.4. Linear Discriminant Analysis

A generalizing Fisher linear discriminant<sup>18</sup> is utilized for the linear projection of features to detach two or more classes. For making influential and discriminating projected features, PCA coefficients can be predicted on to a recent LDA projection axis.

It is required to choose between-class and within-class variability.

The between class variable dataset matrix can be designated by sample variance as:

$$S_B = \frac{1}{c} \sum_{j=1}^c (m_j - m)(m_j - m)^T \quad (3)$$

Within class variance matrix can be entitled as:

$$S_w = \sum_{j=1}^c \sum_{z_k \in w_j} (z_k - m_i)(z_k - m_i)^T \quad (4)$$

where,  $z_k$  is  $k$ -th sample variable belongs to a class.

The generalized Rayleigh Coefficient is:

$$J(w) = \frac{W^T S_B W}{W^T S_w W} \quad (5)$$

where,  $W$  is the matrix for LDA coefficients. It can be classified using the generalized eigenvalue issue as:

$$S_B W = \lambda S_w W \quad (6)$$

where,  $\lambda$  is the eigenvalue.

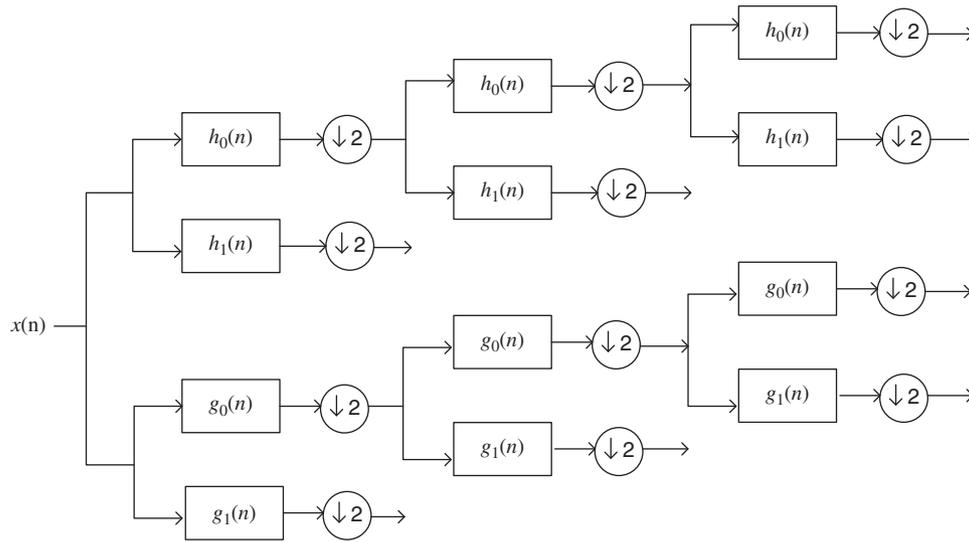


Fig. 2. Block diagram for a 3-level DTCWT.

If  $S_w$  is singular matrix, the above Eq. (6) can be simplified as:

$$S_w^{-1} S_B W = \lambda W \quad (7)$$

where, the eigenvectors of  $S_w^{-1} S_B$  will be  $W$ . The eigenvector matrix will be  $W_{lda}$ ,

$$W_{lda} = [W_1 W_2 W_3 \dots W_k], \quad k \in Z \quad (8)$$

The PCA coefficients values can be projected onto  $l$  lower dimensional LDA projection entitled by eigenvectors correlating with non-zero higher efficient energy eigenvalues,

$$W'_{lda} = [W_1 W_2 W_3 \dots W_l], \quad l \in Z \quad (9)$$

where,  $l \leq k$ .

$F$ , the output feature model is estimated as:

$$F = (W'_{lda})^T \cdot \psi(x)_{pc} \quad (10)$$

### 2.2.5. Extreme Learning Machine Algorithm

Extreme learning machine (ELM) has been lengthened to various research fields and gained great breakthrough. ELM is exceptionally fast training, good generalization, and has universal approximation capability. It has been successful in various applications, for example, ship detection, image quality enhancement, and face recognition. We applied Extreme Machine Learning (ELM) rather than using deep learning based classification because the deep-based model needs a large number of samples for optimum performance. For smaller data, the ELM performs better and as studied by Huang et al.<sup>19</sup> A brief overview of ELM is depicted below.

Consider the training set  $S = \{(x_i, y_i)\}_{i=1}^n$ , where  $x_i \in R^d$  is the training sample, and  $y_i \in \{-1, +1\}$  is its corresponding class level, for the single hidden layer feed-forward neural network (SLFN) with  $L$  hidden neurons. The output function of the problem of can be solved as a

$$f(x_i) = \sum_{k=1}^L \beta_k h_k(a_k, b_k, x_i) = h(x_i) \beta, \quad i = 1, \dots, N$$

Where,  $\beta_k$  is the weight matrix of the output,  $h(x_i) = h_k(a_k, b_k, x_i)$ ,  $k = 1, \dots, L$  is the output of the network with

respect to its training sample  $x_i$ ,  $h(\cdot)$  is a nonlinear continuous function, and  $a_k, b_k$  are the corresponding parameters of  $k$ -th hidden neuron. The suitable parameters are obtained by minimizing the error function  $\|H\beta - Y\|$ . The explanation of a SLFN is shown in Figure 4. ELM utilizes random hidden node parameters and the tuning free training approach for feedforward neural networks, unlike iterative weight updating process as done in conventional gradient descent algorithm. The optimization problem can be solved by least squares (LS) algorithm easily.<sup>21</sup> The network output weights can also be solved by dual optimization problem.<sup>22</sup>

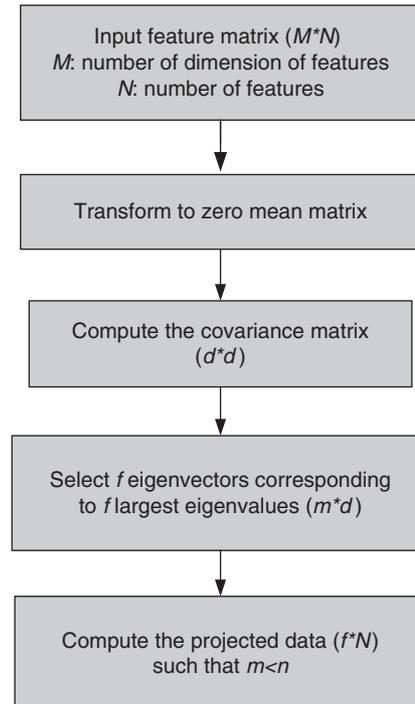


Fig. 3. PCA implementation for feature reduction.

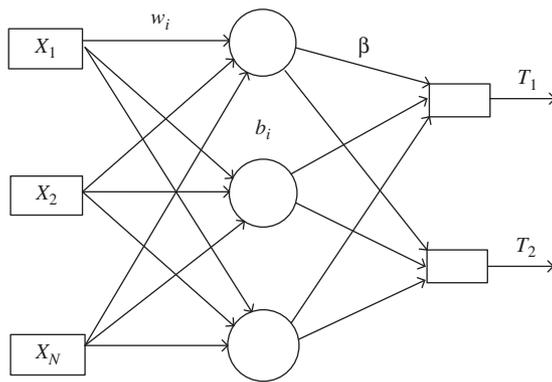


Fig. 4. Sketch of extreme learning machine.

### 3. RESULTS AND DISCUSSION

#### 3.1.

##### 3.1.1. Background

In this article, our proposed method is presented by utilizing Fisher linear discriminant analysis of DTCWT principal components mechanism. The detail description of our proposed method is shown in Figure 5. The benefit of WT above FT is its multiple scaled representations and frequency components with spatial domain information. Fourier coefficients just produce image frequency information, when in fact wavelets contain effective observations of the spatial and frequency domain in a multi-scaled layout. Additionally, wavelets description is spatially localized; Fourier functions are not spatially localized, as they comprise only of image frequency components. MRI images can be illustrated and can be treated at different resolutions, and as a result, they are used as an incisive framework for handling multi-resolution images. Eventually, DWT coefficients can be extracted by utilizing arrays of low and high pass filter banks.

Nevertheless, there are several disadvantages to traditional wavelet transform. These involve drift in wavelet coefficient oscillation in the direction of positive and negative around singularities. Shift variance of the signal which might cause oscillation of wavelet coefficient examples around singularities, substantial aliasing of sufficiently spaced wavelet coefficient patterns, and absence of directional selectivity disturb to the process as well as the model of the geometric image properties (such as edges and ridges). In these circumstances, flaws on the subject of conventional DWT are not proficient by Fourier transform. Influenced by Fourier transform, our enhanced DTCWT is used to overwhelm these drawbacks. Earlier research shows that DTCWT feature-based AD disease detection which performs superior to typical DWT-based feature extraction.<sup>20</sup> Additionally, DTCWT

gives superior singularities of the line and curve representation. Therefore, the discriminative feature can be extracted moderately, which is vital for any pattern classification problem.

Mis-classification rates and higher dimensionality features present the problems concerning pattern classification. For smooth classification, dimensionality reduction methods are utilized to transform data from higher to lower dimensional spaces. PCA is the most commonly used linear transformation and addresses these concerns when it is needed. Extracted features are examined using PCA for feature reduction. For each MRI images from the OASIS as well as ADNI datasets, there are 49152 ( $1536 \times 32$ ) features. After applying PCA method, these results are reduced to  $128 \times 127$  for OASIS dataset and  $172 \times 171$  for ADNI dataset.

After PCA, the classification may still not be enough, as PCA does not define for the variability of features within a class or between classes. To confirm the PCs more separable, it is required to transform data into another space incorporating directions that will find axes, which will inflate the gap between different classes. Thus, LDA is enforced to project PCs onto new projection axes more effective disease classification.

ELM is an emerging effective pattern classification method<sup>23,35</sup> that can be used in many fields such as image processing, video applications, medical applications, etc. The ELM has very quick learning speed, better generalization performance compared to a gradient-based method such as back-propagation methods, less prone to problems like local minima, inappropriate learning rate, and over-fitting, etc. ELM is flexible with hidden activation functions. It has the benefit of quick learning process, relatively superior performance over SVM and its variants.

All programs are executed in MATLAB 2015b which was installed in Intel (R) Core (TM) i3-4160 CPU system. Due to the time complication of the extraction of DTCWT and DWT coefficients from a 2D MRI image slices, are 0.5148 and 0.5109 respectively. There is no meaningful difference in CPU elapsed time when comparing transform approaches. As a dimensionality reduction method, we utilized PCA to neglect higher dimensional input features.

In addition, it is not possible to train and test a classifier with higher dimensional features due to elapsed time. CPU elapsed time to accomplishing ELM classification performance was approximately 70.40 seconds without lowering dimensions. The time needed for our proposed method is approximately roughly 15 seconds; faster than the approach that does not use Fisher discriminant analysis.

##### 3.1.2. Performance Evaluation

The act of a binary classifier can be evoked by utilizing confusion matrix, as shown in Table IV. The number of examples

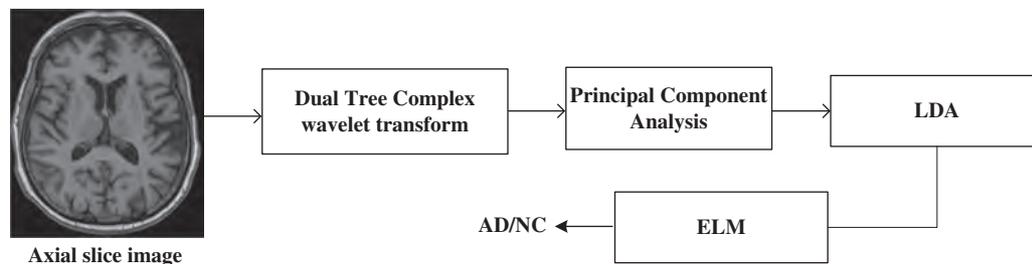


Fig. 5. Flowchart of DTCWT based classification performance of AD from HC.

**Table IV. Confusion matrix for a binary classifier to differentiate between two classes ( $S_1$  and  $S_2$ ).**

True class	Predicted class	
	$S_1$ (Patients)	$S_2$ (Controls)
$S_1$ (Patients)	TP	FN
$S_2$ (Controls)	FP	TN

correctly forecasted by the classifier, which stands at the diagonal. These may be categorized into true positives TP, representing correct identified patients, and true negatives TN, and which also represent correct identified controls. There are several models wrongly which are inaccurately stratified by the classifier which may be allocated into false positives FP, showing controls inaccurately distinguished as patients, and false negatives FN, showing patients inaccurately classified as controls.

Accuracy is determined by calculating the ratio of examples that are accurately distinguished by a classifier:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (11)$$

This result may not be an absolute performance metric in the class distribution of the dataset is unstable. For example, if class  $C_1$  is much greater than  $C_2$ , a high accuracy value could be obtained by a classifier that marks all examples as belonging to class  $C_1$ . Sensitivity is the rate of true positives (TP), and specificity the rate true negatives (TN). Sensitivity and specificity are defined as:

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{and} \quad \text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (12)$$

Sensitivity deals with the proportion of correctly identified patients, and specificity deals with the proportion of correctly identified controls. Furthermore, some other commonly used statistical performance evaluation measures such as precision, recall,  $f\_measure$ , and  $g\_mean$  are also calculated.

These measures are defined as:

$$\begin{aligned} \text{recall} &= \text{sensitivity}, \\ \text{precision} &= \text{TP}/(\text{TP} + \text{FP}), \\ f\_measure &= 2 * ((\text{precision} * \text{recall})/(\text{precision} + \text{recall})), \\ g\_mean &= \text{sqrt}(\text{TP rate} * \text{TN rate}) \end{aligned} \quad (13)$$

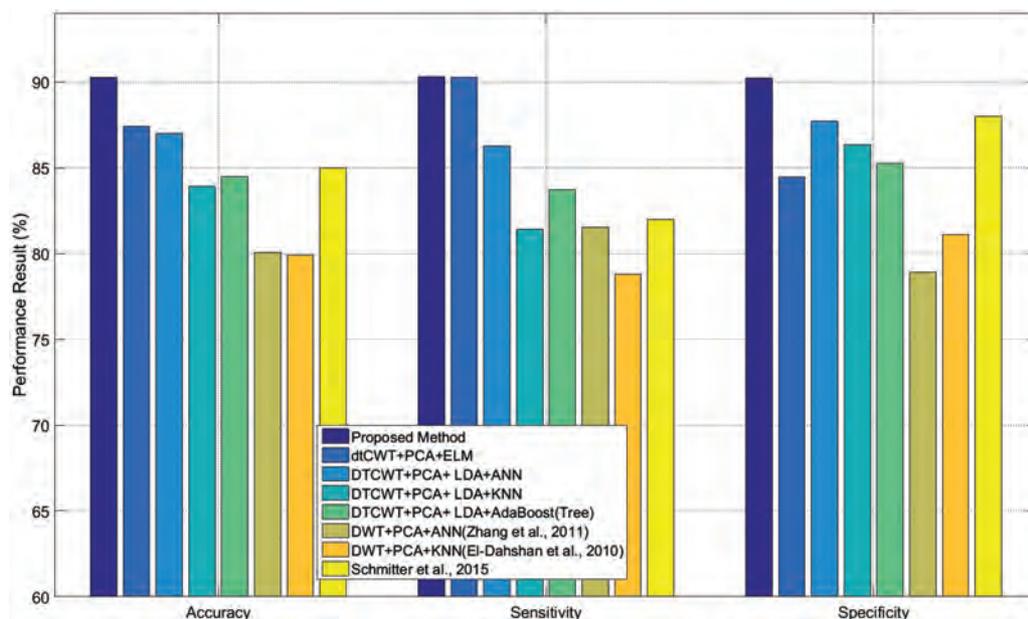
The above measures may give an enhanced assessment of the complete performance of a classifier.

### 3.1.3. Performance of Classification

In this research, the suggested proposed method has been implemented for OASIS and ADNI data to distinguish control subjects from AD subjects. The achieved classification performance about accuracy (acc), sensitivity (sens), and specificity (spec) has been shown in a bar diagram in Figures 6 and 7. Performance varies depending on the principal components used for training and testing. After testing this result with different PC values, it was concluded that optimal classification performance was achieved with  $PC = 20$ . To run a stringent statistical analysis, stratified cross-validation (SCV) is processed. In this analysis, we have applied 5-fold CV to OASIS data and as well as in ADNI data, and as the number of subjects in both the datasets are (is) not large. 5-fold CV splits the dataset into five consecutive folds, whereas the 10-fold CV divides the dataset into ten folds. The accuracies, sensitivities, specificities, and other statistical performance measures obtained with the runs of 10–20, runs of 5-fold of SCV are shown in Tables V and VII.

Although comparison with conventional approach can be difficult, also we have compared our methods with some new conventional disease detection algorithms using both datasets.

To examine the performance of the ADNI dataset, the classification performance has been documented with both run-wise fold-wise classification, as shown in Table VI. Distinct columns and rows signify the classification accuracy of corresponding



**Fig. 6.** Bar diagram of DTCWT based classification measurement of the AD from HC over ADNI dataset.

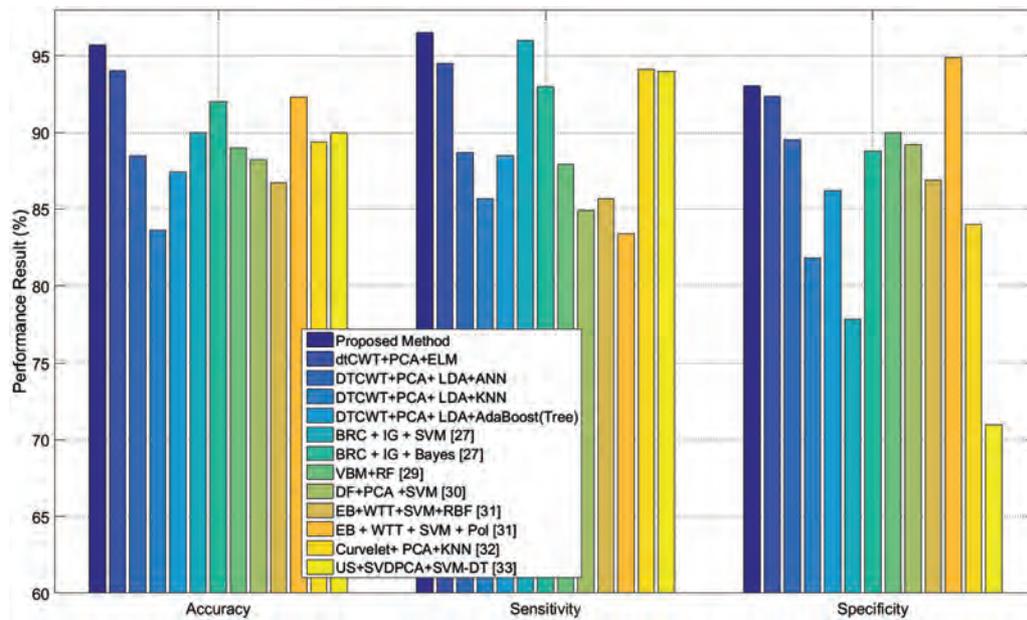


Fig. 7. Bar diagram of DTCWT based classification measurement of the AD from HC over OASIS dataset.

runs and folds. As a result, accuracy is calculated by taking the average value of all folds and runs. The classification performance is carried out in all five folds of each run can be analyzed with that.

We have compared numerous recently used sets of algorithms and approaches,<sup>24–26</sup> using the same datasets as in this article. We have achieved a  $90.26 \pm 1.17\%$  accuracy with comparable sensitivity and specificity, which outperforms the DWT based method proposed by Zhang et al.,<sup>24</sup> and El-Dahshan et al.,<sup>25</sup> shows that in Table VI and Figure 6. The proposed technique was also carried out by applying traditional DWT principal coefficients. We have seen that the DTCWT based approach outperforms DWT

based method. In addition, performance is recorded without using LDA for both types of features. However, classification of performance has become more effective when LDA operated features are considered, as shown in Tables V, IX, and Figure 6. Our approach has been influential from the volumetric feature-based research study suggested by Schmitter et al.,<sup>26</sup> and it outperforms the results there, as shown in Figure 6. Moreover, our outcome compared with other classification methods and produced superior performance.

Similarly, to examine and stratify OASIS dataset, identical approach have been used, namely run-wise and fold-wise classification, as shown in Table VIII.

Table V. Performance evaluation over ADNI dataset.

Methods	Accuracy	Sensitivity	Specificity	Precision	Recall	F-measure	G-mean
Proposed	$90.26 \pm 1.17$	$90.27 \pm 1.29$	$90.20 \pm 1.56$	$90.56 \pm 1.27$	$90.27 \pm 1.27$	$90.32 \pm 1.19$	$90.17 \pm 1.24$
DTCWT + PCA + ELM	$87.41 \pm 0.95$	$90.26 \pm 0.89$	$84.45 \pm 1.31$	$85.74 \pm 0.95$	$90.26 \pm 0.89$	$87.70 \pm 0.77$	$87.14 \pm 0.91$

Table VI. Run and fold wise classification performance of proposed approach over ADNI dataset.

Folds	Runs										Average accuracy
	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10	
Fold 1	94.11	88.23	91.17	91.17	94.11	97.05	97.14	88.57	88.23	97.14	90.25
Fold 2	82.85	91.17	85.71	85.71	94.11	85.71	85.71	88.23	91.42	94.28	
Fold 3	100	80	91.17	94.11	82.85	94.28	94.11	91.42	94.11	88.23	
Fold 4	94.11	94.11	91.17	88.57	97.14	91.17	76.47	88.23	79.41	88.23	
Fold 5	80	91.42	94.28	97.05	94.11	85.29	88.23	91.17	88.57	91.17	
<b>Fold-wise accuracy</b>	90.21	88.99	90.70	91.32	92.47	90.70	88.33	89.52	88.35	91.81	

Table VII. Performance evaluation over OASIS dataset.

Methods	Accuracy	Sensitivity	Specificity	Precision	Recall	F-measure	G-mean
Proposed	$95.72 \pm 1.54$	$96.59 \pm 2.34$	$93.03 \pm 1.67$	$97.63 \pm 1.47$	$96.59 \pm 2.44$	$97.07 \pm 1.61$	$94.72 \pm 1.74$
DTCWT + PCA + ELM	$94.02 \pm 1.59$	$94.53 \pm 1.36$	$92.35 \pm 2.39$	$97.69 \pm 1.91$	$94.53 \pm 1.36$	$95.95 \pm 1.23$	$93.22 \pm 1.32$

**Table VIII.** Run and fold-wise classification performance of proposed approach over OASIS dataset.

Folds	Runs										Average accuracy
	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10	
Fold 1	96	92.30	92.30	96	96.15	91.66	96	96	96	100	95.72
Fold 2	100	95.83	100	88.46	92.30	96.15	100	96	96.15	88	
Fold 3	96.15	96.15	92	95.83	95.83	92	96	92	92	96.15	
Fold 4	96.15	95.83	96	96	96.15	100	92.30	92.30	96.15	100	
Fold 5	91.66	100	96.15	100	100	100	92	100	100	96	
<b>Fold-wise accuracy</b>	95.99	96.02	95.292	95.25	96.08	95.96	95.26	95.26	96.06	96.03	

**Table IX.** Classification performance of AD from HC over ADNI data.

Methods	Accuracy	Sensitivity	Specificity
<b>Proposed</b>	90.26 ± 1.17	90.27 ± 1.29	90.20 ± 1.56
DTCWT + PCA + ELM	87.41 ± 0.95	90.26 ± 0.89	84.45 ± 1.31
DTCWT + PCA + LDA + ANN	86.97 ± 1.30	86.25 ± 1.78	87.72 ± 3.51
DTCWT + PCA + LDA + KNN	83.89 ± 0.75	81.41 ± 1.33	86.34 ± 1.08
DTCWT + PCA + LDA + AdaBoost(Tree)	84.48	83.72	85.26
DWT + PCA + ANN <sup>25</sup>	80.05 ± 0.72	81.538 ± 1.41	78.974 ± 1.09
DWT + PCA + KNN <sup>12</sup>	79.964 ± 1.19	78.771 ± 2.37	81.08 ± 1.67
Schmitter et al., <sup>26</sup>	85	82	88

We observed, as shown in Tables VII, X, and Figure 7 that our approach yielded an accuracy of  $95.72 \pm 1.54$ , a sensitivity of  $96.59 \pm 2.34$ , and specificity of  $93.03 \pm 1.67$ . This classification performance has also been recognized without using LDA, however, outcomes improve when LDA processed on principal dual-tree complex wavelet transform coefficients or principal DWT coefficients and ELM is used as a classifier. The outcome is effective when DTCWT principal coefficients are utilized over DWT approach.

To further verify the validity of the proposed suggested approach, we evaluated it with 12 state-of-the-art methods, as shown in Table X for comparison, which employed different statistical settings.

The outcome shows that US + SVD-PCA + SVM-DT<sup>33</sup> gained an accuracy of 90%, a sensitivity of 94%, and a specificity of 71%. BRC + IG + SVM<sup>27</sup> attained an accuracy of 90.00%, a sensitivity of 96.88%, and a specificity of 77.78%, and curvelet + PCA + KNN<sup>32</sup> achieved stratification accuracy of 89.47%,

a sensitivity of 94.12%, and a specificity of 84.09%. We noticed that these approaches have lower specificity compared to the other approach, which was stated earlier. In contrast, BRC + IG + Bayes<sup>27</sup> gained higher specificity. Likewise, BRC + IG + VFI<sup>27</sup> generated a classification accuracy of 78%, sensitivity of 65.63% and specificity of 100%. Although it gained high specificity, accuracy and sensitivity generated by this algorithm is comparatively poor.

All other approaches attained satisfying outcomes. VBM + RF<sup>29</sup> gained an accuracy of  $89.0 \pm 0.7\%$ , a sensitivity of  $87.9 \pm 1.2\%$ , and a specificity of  $90.0 \pm 1.1$ . These encouraging outcomes largely due to voxel-based morphometry (VBM). DF + PCA + SVM<sup>30</sup> gained an accuracy of  $88.27 \pm 1.89\%$ , a sensitivity of  $84.93 \pm 1.21\%$ , a specificity of  $89.21 \pm 1.63\%$ . This approach is based on a novel method called displacement field (DF). EB + WTT + SVM + RBF<sup>31</sup> achieved an accuracy of  $86.71 \pm 1.93\%$ , a sensitivity of  $85.71 \pm 1.91\%$ , a specificity of  $86.99 \pm 2.30\%$ , though EB + WTT + SVM + Pol<sup>31</sup> yields improved classification

**Table X.** Algorithm performance comparison over OASIS MRI data.

Algorithm	Accuracy	Sensitivity	Specificity	Precision
<b>Proposed</b>	95.72 ± 1.54	96.59 ± 2.34	93.03 ± 1.67	<b>96.13 ± 1.57</b>
DTCWT + PCA + ELM	94.02 ± 1.59	94.53 ± 1.36	92.35 ± 2.39	94.15 ± 2.01
DTCWT + PCA + LDA + ANN	88.59 ± 2.08	88.75 ± 2.75	89.55 ± 3.96	NA
DTCWT + PCA + LDA + KNN	83.69 ± 1.57	85.7 ± 1.94	81.8 ± 1.45	NA
DTCWT + PCA + LDA + AdaBoost (Tree)	87.45	88.59	86.26	NA
BRC + IG + SVM <sup>27</sup>	90.00 [77.41, 96.26]	96.88 [82.01, 99.84]	77.78 [51.92, 92.63]	NA
BRC + IG + Bayes <sup>27</sup>	92.00 [79.89, 97.41]	93.75 [77.78, 98.27]	88.89 [63.93, 98.05]	NA
BRC + IG + VFI <sup>27</sup>	78.00 [63.67, 88.01]	65.63 [46.78, 80.83]	100.00 [78.12, 100]	NA
MGM + PEC + SVM <sup>28</sup>	92.07 ± 1.12	86.67 ± 4.71	N/A	95.83 ± 5.89
GEODAN + BD + SVM <sup>28</sup>	92.09 ± 2.60	80.00 ± 4.00	NA	88.09 ± 5.33
TJM + WTT + SVM <sup>28</sup>	92.83 ± 0.91	86.33 ± 3.73	N/A	85.62 ± 0.85
VBM + RF <sup>29</sup>	89.0 ± 0.7	87.9 ± 1.2	90.0 ± 1.1	NA
DF + PCA + SVM <sup>30</sup>	88.27 ± 1.9	84.93 ± 1.21	89.21 ± 1.6	69.30 ± 1.91
EB + WTT + SVM + RBF <sup>31</sup>	86.71 ± 1.93	85.71 ± 1.91	86.99 ± 2.30	66.12 ± 4.16
EB + WTT + SVM + Pol <sup>31</sup>	92.36 ± 0.94	83.48 ± 3.27	94.90 ± 1.09	82.28 ± 2.78
Curvelet + PCA + KNN <sup>32</sup>	89.47	94.12	84.09	NA
US + SVDPCA + SVM-DT <sup>33</sup>	90	94	71	NA
DTCWT + PCA + FFNN <sup>36</sup>	90.6 ± 0.01	92.00 ± 0.04	87.78 ± 0.04	89.6 ± 0.03

performance. In addition, MGM + PEC + SVM,<sup>28</sup> GEODAN + BD + SVM,<sup>28</sup> and TJM + WTT + SVM<sup>28</sup> each obtained roughly 92% accuracy with relatively high sensitivity and precision; specificity was not taken into account for these approach.

Lastly, taking classification performance into consideration, our method outperforms entirely other methods analyzed here. We have also formed encouraging performance evaluation for sensitivity and specificity. Hence, we submit our outcomes, which are either finest or equivalent to the other compared techniques.

#### 4. CONCLUSIONS

Our work presents the way of combining different methods to detect the severity of the disease. It also demonstrates the importance of an assembling method, which includes efficient feature extraction, dimensionality reduction, and classification to detect the stage of the disease. This work may lead to design an automated Computer aided diagnosis (CAD) framework of the disease, which has the potential to change the current diagnosis methods done by radiologists and clinicians manually. The proposed method applies LDA, principal component analysis on DTCWT coefficients, then ELM to detect AD during the training phase. Both PCA and LDA reduces the dimensionality reduction considering inside class variability and between class variability of extracted feature. The approach gives us the promising result which is comparable or superior to some state of the art algorithms mentioned in the manuscript.

In future, our research activity will focus on following directions: 3D DTCWT-based feature extraction with the concept of multi-resolution and classification, and Convolutional Neural Network (CNN) based stratification using 3D MRI.

**Acknowledgments:** This research was supported by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (NRF-2014M3C7A1046050). And this work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2017R1A2B4006533). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The funding details of ADNI can be found at: <http://adni.loni.usc.edu/about/funding/>.

#### References and Notes

1. J. Ashburner and S. Klöppel, Multivariate models of inter-subject anatomical variability. *NeuroImage* 56, 422 (2011).
2. N. Koutsouleris, E. M. Meisenzahl, C. Davatzikos, R. Bottlender, T. Frodl, J. Scheuerecker, G. Schmitt, T. Zetzsche, P. Decker, M. Reiser, H. J. Möller, and C. Gaser, Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch. Gen. Psychiatry* 66, 700 (2009).
3. M. Weygandt, K. Hackmack, C. P. Füller, J. Bellmann-Strobl, F. Paul, F. Zipp, and J. D. Haynes, MRI pattern recognition in multiple sclerosis normal-appearing brain areas. *PLoS One* 6, e21138 (2011).
4. M. Chupin, E. Gerardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehericy, H. Benali, L. Garnero, and O. Colliot, Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* 19, 579 (2009).
5. S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr., J. Ashburner, and R. S. Frackowiak, Automatic classification of MR scans in Alzheimer's disease. *Brain* 131, 681 (2008).
6. C. Ecker, A. Marquand, J. Mourão-Miranda, P. Johnston, E. M. Daly, M. J. Brammer, S. Maltezos, C. M. Murphy, D. Robertson, S. C. Williams, and D. G. Murphy, Describing the brain in autism in five dimensions-magnetic resonance imaging assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. *The Journal of Neuroscience* 30, 10612 (2010).
7. R. A. Yotter, I. Nenadic, G. Ziegler, P. M. Thompson, and C. Gaser, Local cortical surface complexity maps from spherical harmonic reconstructions. *NeuroImage* 56, 961 (2011).
8. Z. Sankari and H. Adeli, Probabilistic neural networks for EEG-based diagnosis of Alzheimer's disease using conventional and wavelet coherence. *Journal of Neuroscience Methods* 197, 165 (2011).
9. S. Chaplot, L. M. Patnaik, and N. R. Jagannathan, Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical Signal Processing and Control* 1, 86 (2006).
10. C. H. Moritz, V. M. Haughton, D. Cordes, M. Quigley, and M. E. Meyerand, Whole-brain functional MR imaging activation from finger tapping task examined with independent component analysis. *American Journal of Neuroradiology* 21, 1629 (2000).
11. R. N. Bracewell, *The Fourier Transform and its Applications*, 3rd edn., McGraw-Hill, New York (1999).
12. El-Dahshan, T. Hosney, and A. B. M. Salem, Hybrid intelligent techniques for MRI brain images classification digital. *Signal Process* 20, 433 (2010).
13. D. Jha, J.-I. Kim, and G.-R. Kwon, Pathological brain detection using weiner filtering, 2D-discrete wavelet transform, probabilistic PCA, and random subspace ensemble classifier. *Computat. Intelli. Neurosci.* 2017, 4205141 (2017).
14. Y. Zhang and S. Wang, Detection of Alzheimer's disease by displacement field and machine learning. *PeerJ* 3, e1251 (2015).
15. D. Tomar and Sonali Agarwal, Twin support vector machine: A review from 2007 to 2014. *Egyptian Informatics Journal* 16, 55 (2015).
16. A. Barri, A. Dooms, and P. Schelkens, The near shift-invariance of the dual-tree complex wavelet transform revisited. *Journal of Mathematical Analysis and Applications* 389, 1303 (2012).
17. H. Andersen, D. M. Gash, and M. J. Avison, Principal component analysis of the dynamic response measured by fMRI: A generalized linear systems framework. *Journal of Magnetic Resonance Imaging* 17, 795 (1999).
18. R. A. Fisher, The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7, 179 (1936).
19. G. B. Huang, Q. Y. Zhu, and C. K. Siew, Extreme learning machine: Theory and applications. *Neurocomputing* 70, 489 (2006).
20. S. Alam and G. R. Kwon, Classification of Alzheimer disease using dual tree complex wavelet transform and minimum redundant and maximum relevant feature. *The Journal of Korean Institute of Next Generation Computing* 12, 51 (2016).
21. G. B. Huang and L. Chen, Convex incremental extreme learning machine. *Neurocomputing* 70, 3056 (2007).
22. R. Fletcher, *Practical Methods of Optimization: Constrained Optimization*, 2, John Wiley and Sons, Chichester, UK (1981).
23. B. Mirza, Z. Lin, and N. Liu, Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift. *Neurocomputing* 149, 316 (2015).
24. El-Dahshan, T. Hosney, and A. B. M. Salem, Hybrid intelligent techniques for MRI brain images classification Digital. *Signal Process.* 20, 433 (2010).
25. Y. Zhang, Z. Dong, L. Wu, and S. Wang, A hybrid method for MRI brain image classification. *Exp. Syst. Appl.* 38, 10049 (2011).
26. D. Schmitter, A. Roche, B. Maréchal, et al., An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. *J. Neuroimage: Clinical* 7, 7 (2015).
27. C. Plant, S. J. Teipel, A. Oswald, C. Bohm, T. Meindl, J. M. Miranda, A. W. Bokde, H. Hampe, and M. Ewers, Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neu. Image* 50, 162 (2010).
28. A. Savio and M. Grana, Deformation based feature selection for computer aided diagnosis of Alzheimer's disease. *Expert Syst. Appl.* 40,1619 (2013).
29. K. R. Gray, P. Alijabar, R. A. Heckemann, A. Hammers, and D. Rueckert, Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neu. Ima.* 65, 167 (2013).
30. Y. Zhang and S. Wang, Detection of Alzheimer's disease by displacement field and machine learning. *PeerJ* 3 (2015).
31. Y. Zhang, Z. Dong, P. Phillips, S. Wang, J. Genlin, J. Yang, and T.-F. Yuan, Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Front. Comput. Neurosci.* 9, 66 (2015).

32. D. Jha and G. R. Kwon, Alzheimer disease detection in MRI using curvelet transform with KNN. *The Jour. Korean Inst. Infor. Tech.* 14 (2016).
33. Y. Zhang, S. Wang, and Z. Dong, Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Prog. Electromag. Res.* 144, 171 (2014).
34. S. Lu and X. Qiu, A pathological brain detection system based on extreme learning machine optimized by bat algorithm. *CNS Neurol. Disord. Drug Targets* 16, 23 (2017).
35. B. Mirza, Z. Lin, and K.-A. Toh, Weighted online sequential extreme learning machine for class imbalance learning. *Neural Processing Letters* 38, 465 (2013).
36. D. Jha, J.-I. Kim, and G.-R. Kwon, Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and fee-forward neural network. *Journal of Healthcare Engineering in Hindawi* 2017, 9060124 (2017).
37. R. K. Lama, J. Gwak, and S. W. Lee, Diagnosis of Alzheimer's disease based on structural MRI images using a regularized extreme learning machine and PCA features. *Journal of Healthcare Engineering in Hindawi* 2017, 5485080 (2017).
38. D. Jha, Ji-Kim, B. Lee, and G.-R. Kwon, Efficient cascade model for pathological brain image detection by magnetic resonance imaging. *J. Med. Imaging Health Inf.* 7, 1744 (2017).

Received: xx Xxxx xxxx. Accepted: xx Xxxx xxxx.