

Research Article

Diagnosis of Alzheimer's Disease Using Dual-Tree Complex Wavelet Transform, PCA, and Feed-Forward Neural Network

Debesh Jha, Ji-In Kim, and Goo-Rak Kwon

Department of Information and Communication Engineering, Chosun University, 309 Pilmun-Daero, Dong-Gu, Gwangju 61452, Republic of Korea

Correspondence should be addressed to Goo-Rak Kwon; grkwon@chosun.ac.kr

Received 30 December 2016; Revised 22 March 2017; Accepted 30 April 2017; Published 21 June 2017

Academic Editor: Jose M. Juarez

Copyright © 2017 Debesh Jha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Error-free diagnosis of Alzheimer's disease (AD) from healthy control (HC) patients at an early stage of the disease is a major concern, because information about the condition's severity and developmental risks present allows AD sufferer to take precautionary measures before irreversible brain damage occurs. Recently, there has been great interest in computer-aided diagnosis in magnetic resonance image (MRI) classification. However, distinguishing between Alzheimer's brain data and healthy brain data in older adults (age > 60) is challenging because of their highly similar brain patterns and image intensities. Recently, cutting-edge feature extraction technologies have found extensive application in numerous fields, including medical image analysis. Here, we propose a dual-tree complex wavelet transform (DTCWT) for extracting features from an image. The dimensionality of feature vector is reduced by using principal component analysis (PCA). The reduced feature vector is sent to feed-forward neural network (FNN) to distinguish AD and HC from the input MR images. These proposed and implemented pipelines, which demonstrate improvements in classification output when compared to that of recent studies, resulted in high and reproducible accuracy rates of $90.06 \pm 0.01\%$ with a sensitivity of $92.00 \pm 0.04\%$, a specificity of $87.78 \pm 0.04\%$, and a precision of $89.6 \pm 0.03\%$ with 10-fold cross-validation.

1. Introduction

Alzheimer's disease (AD) is an irremediable neurodegenerative disorder that causes dementia in elderly people around the globe. It has been predicted that the pervasiveness of AD will double within the next 2 decades and that one out of every 85 people will be afflicted with the disease by 2050 [1]. Therefore, there is a need to identify neuroimaging biomarkers that can grant accurate and early diagnoses of dementia. In addition, to diagnose an AD sufferer clinically at a primitive disease stage, many imaging biomarkers must be identified using different imaging modalities, such as MRI [2], position emission tomography (PET) [3], functional magnetic resonance imaging (fMRI) [4], single-photon emission computed tomography (SPECT) [5], and magnetic resonance spectral imaging (MRSI) [6].

An accurate and early diagnosis of AD and identification of the risk of progression from mild cognitive impairment (MCI) to AD provide AD sufferers with awareness of the

condition's severity and allow them to take preventative measures, such as making lifestyle changes and taking medications [7]. Currently, many neurologists and medical analysts have been spending significant time to researching technique to allow for early diagnosis of AD, and encouraging results have been frequently achieved [8]. MRI is an influential, noninvasive brain imaging technique that provides higher-quality information about the shape and volume of the brain than computed tomography (CT), SPECT, and PET scans. It provides superior soft tissue differentiation, high spatial resolution, and better contrast and can even identify tiny irregularities in the brain [9]. Moreover, the diagnostic use of MRI has been tremendously improved due to the automated and precise labeling of MR images, which performs an important role in identifying AD in related patients from healthy and elderly controls (HC) [10].

Earlier, majority of diagnosis work was accomplished manually or semimanually for measuring a priori region of interest (ROI) of MRI, based on the reality that subjects with

AD experience have more cerebral atrophy when compared to HCs [11, 12]. Most of this ROI-based examination focused on the contracting of the cortex and hippocampus and amplified ventricles. Nevertheless, ROI-based approaches are not practicable in hospitals because of few shortcomings: (i) ROI technique needs a priori data and expert knowledge. (ii) The manual diagnosis accuracy is dependent on the knowledge of physicians and interpreter [13]. (iii) The interaction among the voxels was troublesome to enforce. (iv) It was essential to explore other potential areas that may be linked to AD [14]. (v) Automatic segmentation of ROI was not beneficial in practice, and investigator needed to segment the brain using hand [15]. Therefore, automated methods can assist physician in diagnosing diseases from images such as those produced by MRI, for which many slices are extracted from the tissues and long periods of may be necessary for the evaluation of the images.

The aim of this article is to present an automated approach for diagnosing AD by using the “whole brain analysis” method. It has achieved popularity, since it examines entire voxels of the brain. It is not essential to segment the brain as earlier, and it does not require any biomarker for the classification purpose. The main drawback is dimensionality that can be resolved through high-speed computers, which is comparably inexpensive [16]. The whole-brain investigation laboriously relies on true computation, and it can only be finished by a computer researcher after a physician assisted in labeling the input data as either AD or HC. Usually, the whole-brain inspection labels the entire brain as a ROI, where two stages are involved, namely, feature extraction and classification.

Scholars have presented different methods to extract effective features for the detection of AD and other types of pathological brain disease. Additionally, classification models and methods survive; nevertheless, not all of them are suitable for the processing of MR brain images. Based on latest literature, we found two drawbacks with the previous work: (i) The discrete wavelet transform (DWT) is usually utilized for feature extraction. The DWT has better directional selectivity in horizontal, vertical, and diagonal directions and has better image representation than Fourier transform, but its major drawbacks are that it has poor directionality, is sensitive to shifts, and lacks phase information. (ii) Most of the state-of-the-art mechanisms consider only single slice-based detection (SSD) per patient. The obtained slices may not contain the foci of the disease.

To tackle above problems, we suggested two improvements. First, we propose a DTCWT that possesses attractive properties for image processing, including shift invariance and high directionality. Second, we consider multiple slices for each patient unlike previous studies, so that information gain is more consistent, reliable, and accurate. In hospitals, multiple slice-based detection is utilized because of its inexpensiveness. Research has clearly showed that the DTCWT is more suitable than the traditional wavelet domain for feature extraction [17].

Our contribution aims to introduce a novel method for AD detection with higher accuracy than state-of-the-art methods, on the basis of DTCWT, PCA, and ANN

technique. Furthermore, we build a computer-aided diagnosis (CAD) system, which can be utilized in the early diagnosis of AD-related brain area and subjects. Our objective is to develop assisting tool for clinicians.

All of the preprocessing methods are used to obtain good results. To show effectiveness of our proposed system, we have evaluated performance measures including accuracy, sensitivity, specificity, precision, and bar plot for the comparison of the proposed method with the existing systems. The paper is arranged as follows. Section 2 offers background knowledge on materials and methods. In Section 3, the experiments, results, and discussion are presented. Finally, conclusion and plan for future studies are presented in Section 4.

2. Materials and Methods

2.1. Materials

2.1.1. Dataset. In our study, the dataset is accessed from Open Access Series of Imaging Studies (OASIS). OASIS is a project, for compiling and sharing MRI datasets of the brain to make such data accessible to the scientific community. The data are accessible at <http://www.oasis-brains.org>. A sample of the MR brain image is shown in Figure 1.

OASIS provides two types of data: cross-sectional and longitudinal MRI data. In this study, we used cross-sectional MRI data because we aimed to develop an automatic system for detecting AD, which would not require longitudinal data that had been gathered from AD patients over long periods of time.

The dataset consists of 416 subjects whose ages are between 18 and 96. In our study, we consider 126 samples (including 28 ADs and 98 HCs). Table 1 shows statistical information about the subjects included in the experiment. Only right-handed subjects are included in the study, consisting of both men and women. The exclusion criterion is patients less than 60 years of age or any of their reports are missing. The unbalanced data may cause difficulty in future recognition; we fine-tune the cost matrix to resolve this issue.

The dataset contains information about the patient’s demographics. The demographic features contain gender (M/F), age, education, socioeconomic status, and handedness. The mini mental state examination (MMSE) is a short 30-point questionnaire test utilized to monitor for cognitive impairment and dementia. The MMSE test comprises simple questions and problems in numeral areas: the time and place, repeating list of words, arithmetic, language utilization, and comprehension, in addition to basic motor skills. Clinical dementia rating (CDR) is a numeric scale measuring the severity of symptoms of dementia. The patients’ cognitive and functional performances were accessed in six areas: memory, orientation, judgement and analytical, community affairs, residence and hobbies, and individual care. The patients’ CDR ranks and education level are listed in Tables 2 and 3, respectively.

2.2. Proposed Method. The proposed method consists of three important stages, namely, feature extraction using DTCWT,

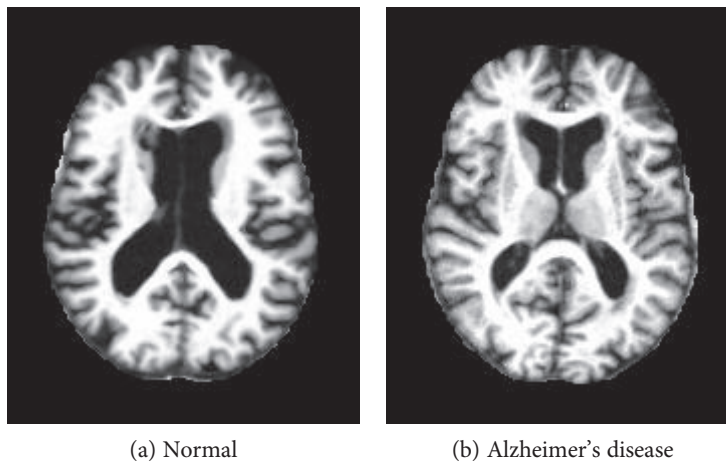


FIGURE 1: Dataset sample (axial view after preprocessing).

TABLE 1: Statistical data of the participants.

Factor	HC	AD
No. of patients	98	28
Age (years)	75.91 ± 8.98	77.75 ± 6.99
Education	3.26 ± 1.31	2.57 ± 1.31
Socioeconomic status	2.51 ± 1.09	2.87 ± 1.29
CDR	0	1
MMSE score	28.95 ± 1.20	21.67 ± 3.75
Gender (M/F)	26/72	9/19

feature dimensionality reduction using PCA, and classification using feed-forward artificial neural network. The overall block diagram of the suggested method is shown in Figure 2. Normalization of image is included in preprocessing section. All of these individual techniques have been proven outstanding, so we strongly believe that the proposed method can also achieve excellent results.

2.3. Image Preprocessing and Normalization. For each patient, each scanning session involves the MR of three or four T1-weighted image scans. In order to add the signal-to-noise ratio (SNR), all indicated MRI scans with the identical protocol of the same individual are motion-corrected and spatially coregistered, to the Talairach coordinate space to produce an averaged image, and then are brain-masked. The motion correction recorded the 3D images of all scans and then developed an average 3D image in initial acquisition space. Also, the scans are then resampled to $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$. The obtained image is converted from acquisition space to Talairach coordinate space. Lastly, the brain extraction is achieved.

We used MRIcro software (which can be downloaded from <http://www.cabiatl.com/mricro/mricro/>) and imported the image from the backup folder and then extracted the 2D MR image slices of each subject. In this paper, we only choose 32 important center slices from each subject manually based on our experience. These slices are used for

TABLE 2: Clinical dementia rating scale.

CDR	Rank
0	Nondementia
0.5	Very mild dementia
1	Mild dementia
2	Moderate dementia

TABLE 3: Education codes.

Code	Description
1	Beneath high school graduate
2	Secondary school graduate
3	Some college
4	College graduate
5	Above college

preprocessing. The reason behind picking center slice from all slices is that it retains more relevant information about the brain tissues as compared to earlier slices and later slices in the group of 1–256 slices. The direction of the slice possibly may be sagittal, coronal, or axial. In this research, we chose axial direction by knowledge. The same process is applied to all the subjects (126 including both ADs and HCs). All images are in PNG format, and the dimensions of the slices are 176×208 . The image is resized to 256×256 before being used for further processing.

2.4. Discrete Wavelet Transform. The discrete wavelet transform (DWT) is an image processing method [18] that gives multiscale representation of a stated signal or image [19]. Standard DWT is helpless to shift variance issue and only has horizontal and vertical directional selectivity [20]. Suppose s denotes a particular signal, n symbolizes the sampling point, h and g denote a high-pass filter and low-pass filter, respectively, and H and L depict the coefficients of high-pass and low-pass subbands. We have

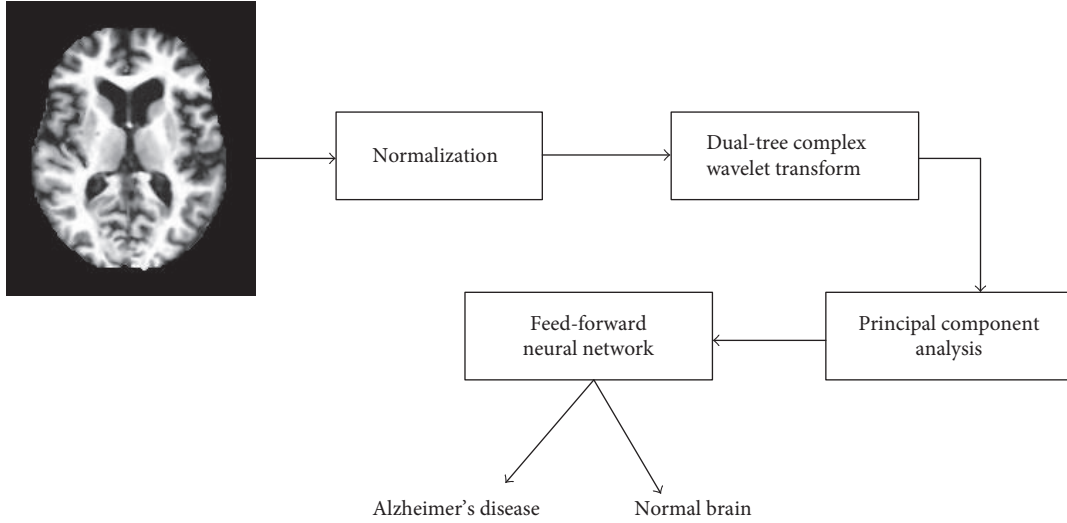


FIGURE 2: Block diagram of the proposed system.

$$\begin{aligned}
 H(n) &= \sum_m h(2n-m)s(m), \\
 L(n) &= \sum_m g(2n-m)s(m).
 \end{aligned} \tag{1}$$

The LH represents a low-pass filter along x -axis and high-pass filter along y -axis. HL represents a high-pass filter along x -axis and low-pass filter along y -axis. The LL represents low-pass filters along both directions, and HH represents high-pass filters along both directions.

Here, the HL and LH have clear-cut for both vertical and horizontal orientations. For the HH , it combines directions of both -45 and $+45$ degrees jointly, which stems from the utilization of real-valued filters in DWT. This combining also hinders the direction check [21].

2.5. Dual-Tree Complex Wavelet Transform. The dual-tree complex wavelet transform (DTCWT) is a modified version of the traditional DWT. To help boost the directional selectivity impaired by DWT, DTCWT is proposed. The traditional DWT is shift variant because of the decimation operation used in the transform. As a consequence, a small shift in the input signal can create a very dissimilar set of wavelet coefficients formed at the output. It utilizes two real DWTs processing input data in parallel [22]. The first DWT symbolizes the real component of the transform, whereas the second DWT depicts the imaginary component together forming a complex transform.

The DTCWT provides a solution for “shift-invariant problems” as well as for “directional selectivity in two or more dimensions,” which are both shortcomings of the ordinary DWT [23]. It obtains directional selectivity by utilizing wavelets that are approximately analytic. It also has the ability to produce a total of six directionally discriminating subbands oriented in the ± 15 , ± 45 , and ± 75 directions, for both the real (R) and imaginary (I) parts. Figure 3 illustrates the DTCWT. Let $h_i(n)$ and $g_i(n)$ be the filters in the first stage as in Figure 3. Let the new k th stage response of the first filter

bank be $H_{\text{new}}^{(k)}(e^{jw})$ and second filter bank be $H'_{\text{new}}(e^{jw})$; we now have the following result as a corollary of Lemma 1.

Corollary 1. Suppose one is provided with CQF pairs $\{h_o(n), h_1(n)\}$, $\{h'_o(n), h'_1(n)\}$. For $k > 1$,

$$H_{\text{new}}^{(k)}(e^{jw}) = H \left\{ H'_{\text{new}}(e^{jw}) \right\}, \tag{2}$$

if and only if

$$h'_0(1)(n) = h_0(1)(n-1). \tag{3}$$

A 2D image $f(x, y)$ is decomposed by 2D DTCWT over a series of dilations and translations of a complicated scaling function and six complex wavelet functions $\varphi_{j,l}^\theta$; that is,

$$f(x, y) = \sum_{l \in \mathbb{Z}^2} s_{j_o, l^{\theta}} l^{(x,y)} + \sum_{\theta \in \Theta} \sum_{j \geq j_o} \sum_{l \in \mathbb{Z}^2} c_j^\theta l^{\theta} l^{(x,y)}, \tag{4}$$

where $\theta \in \Theta = \{\pm 15^\circ, \pm 45^\circ, \pm 75^\circ\}$ gives the directionality of the complex wavelet function. This is to say that the decomposition of $f(x, y)$ by utilizing the DTCWT creates one complex valued low-pass subband and six complex valued high-pass subbands at every level of decomposition, where every high-pass subband corresponds to one particular direction θ .

A study was carried out in [24] to compare the DTCWT’s directional selectivity to that of the DWT. The simulation results showed that the edges detected by the DTCWT had clear contours, and nearly all directions could be detected clearly and perfectly. However, the edges detected by the DWT were discontinuous, and only horizontal and vertical edges could be successfully detected. The results verified the effectiveness of the DTCWT over the DWT. By utilizing DTCWT, we are extracting DTCWT coefficients from the preprocessed images. The additional features include the information about the demographics of the patients such as age, gender, handedness, education, SES, and clinical

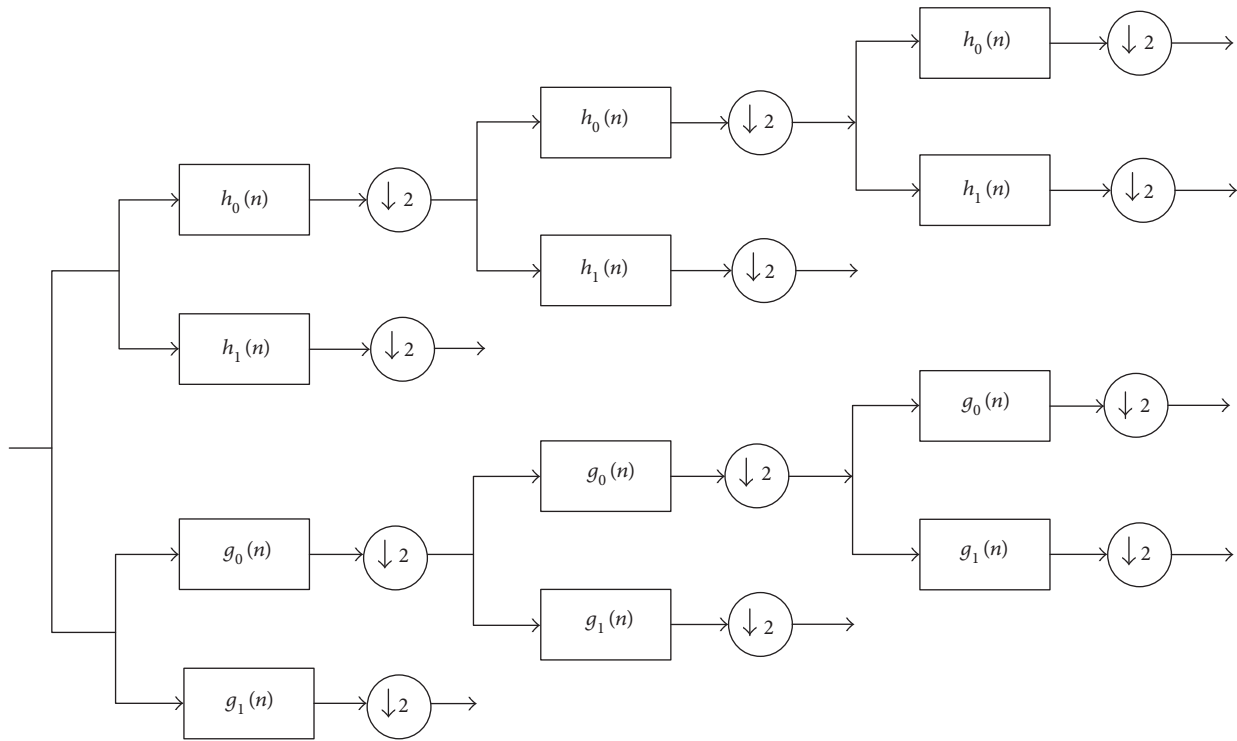


FIGURE 3: The DTCWT is implemented utilizing two wavelet filter banks functioning in parallel.

Let X be an input text file (X : matrix of dimensions $M \times N$)
Accomplish the following steps:

Step 1. Estimate the empirical mean: $u[m] = \frac{1}{N} \sum_{n=1}^N X[m, n]$.

Step 2. Compute the deviations from the mean and save the data in the matrix $B[M \times N]$: $B = X - u \cdot h$, here, h is a $1 \times N$ row vector of all 1's: $h[n] = 1$ for $n = 1, \dots, N$.

Step 3. Obtain the covariance matrix C : $C = \frac{1}{N} B \cdot B^*$.

Step 4. Get the eigenvectors and eigenvalues of the covariance matrix $V^{-1} C V = D$: V -the eigenvectors matrix; D -the diagonal matrix of eigenvalues of C , $D[p, q] = \lambda_m$ for $p = q = m$ is the m th eigenvalues of the covariance matrix C .

Step 5. Rearrange the eigenvectors and eigenvalues: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \dots \lambda_N$.

Step 6. Selecting components and developing a feature vector: save the first L columns or V as the $M \times L$ matrix W , $W[p, q] = V[p, q]$, for $p = 1, \dots, M$, $q = 1, \dots, L$ where $1 \leq L \leq M$.

Step 7. Obtaining the fresh data set: The eigenvectors with the leading eigenvalues are forecasted into space, this projection appears in a vector depicted by fewer dimension ($L < M$) accommodating the essential coefficients.

ALGORITHM 1: PCA algorithm.

examination. The handedness features are not included in the work since all the patients are right-handedness.

2.6. Principal Component Analysis. The coefficient from the DTCWT enlarges the dimensionality of feature space that makes the classification job more complicated.

Additionally, it leads to excessive computational overhead and enormous memory storage. As a result, it is essential to lower the dimension of the feature set and get the significant features to boost the classification result. Since the last two decades, a method called PCA has earned much more attention for data visualization and reduction of dimensionality. It systematically projects the initial input

data to a lower-dimensional space, well-known as principal subspace through an orthogonal transformation while preserving most of the data variations. For a stated set of likely correlated variables, these transformation outcomes in a set of values of linearly uncorrelated variables are called as principal components (PCs). All of the steps to implement PCA are demonstrated in Algorithm 1. The additional information on PCA and its implementations can be viewed in literature [25, 26].

Let us consider a set of data. PCA is employed to find a linear lower-dimensional reduction of the dataset. In this case, the variance of the constructed data is preserved. PCA limits the feature vectors to the component it selects, which

leads to an effective classification algorithm. The main idea behind implementing PCA is reduction of the dimensionality of the DTCWT coefficients, which results in more adequate and accurate classification.

The following algorithm is utilized to obtain the principal components from the input matrix and finally fed to the feed-forward neural network. Now, the input matrix possesses only these PCs. Hence, the size of the matrix is reduced. Therefore, feature extraction is done in two steps: DTCWT extracts the wavelet coefficients, and essential coefficients are later selected by the PCA as described in Algorithm 1.

2.7. Feed-Forward Neural Networks

2.7.1. Structure. Feed-forward neural networks (FNN) are broadly used in pattern classification because they do not need any information regarding the probability distribution or a priori probabilities of distinct classes. Neural networks (NN) harness power from their densely parallel structure and their ability to acquire information from experience. As a result, they can be utilized for accurate classification of input data into different classes, provided that they are pre-trained. The architecture of a multilayer feed-forward neural network is shown in Figure 4.

Three factors need to be considered in designing an ANN for a specific application: (i) the topology of the network, (ii) the training algorithm, and (iii) the neuron activation function. A network may have many layers of neurons, and its complete architecture may possess either a feed-forward or a back propagation structure. A multilayer back propagation NN with sigmoid neurons in its hidden layer is chosen. Similarly, linear neurons are selected for the output layer. The training vector is provided to the NN, which is instructed batch mode [27]. The NN is a two-layer network, and its configuration is $N_I \times N_H \times N_O$; N_I represents the input neurons, N_H is the hidden layer, and N_O indicates that the brain under observation is either HC or AD.

2.7.2. Training Method. Mathematicians have already proven that a conjugate gradient (CG) algorithm, probing along conjugate gradient directions, produces a faster convergence than the steepest descent directions do. Among CG algorithm, the scaled conjugate gradient (SCG) method is the most powerful [28]. Thus, we utilize the SCG to train our network.

Let ω_1 and ω_2 be the connection weight matrix linking the input layer and hidden layer, and the hidden layer and the output layer, respectively. Later, we can deduce the training process reported by the following equations to improve these weighted values that can be divided into four subsequent steps [29].

- (1) The calculation of the outputs of all neurons in the hidden layer is done by

$$y_j = f_H \left(\sum_{i=1}^{N_I} \omega_1(i, j) x_i \right) \quad j = 1, 2, \dots, N_H. \quad (5)$$

Here, x_i stands for the i th input value, y_j stands for the j th output of the hidden layer, and f_H refers to the activation function of hidden layer, commonly a sigmoid function as observed:

$$f_H(x) = \frac{1}{1 + \exp(-x)}. \quad (6)$$

- (2) The outputs of all neurons in the output layer are stated as follows:

$$O_k = f_o \left(\sum_{j=1}^{N_H} \omega_2(j, k) y_j \right) \quad k = 1, 2, \dots, N_O. \quad (7)$$

Here, f_o represents the activation function of output layer that is usually a line function. At first, all weights are accredited with random values and amended by the delta rule on the basis to the learning samples.

- (3) The error is articulated as the MSE of the distinction among output and target value [30].

$$E_l = \text{MSE} \left(\sum_{k=1}^{N_O} O_k - T_k \right) \quad l = 1, 2, \dots, N_s, \quad (8)$$

where T_k depicts the k th value of the genuine labels which is already well-known to users and N_s denotes the number of samples [31].

- (4) Let us consider that there are N_s samples; therefore, the fitness value can be written as

$$F(\omega) = \sum_{l=1}^{N_s} E_l, \quad (9)$$

where ω designates the vectorization of the (ω_1, ω_2) . The aim is to minimize the fitness function $F(\omega)$, namely, force the output values of every sample appropriate to equivalent target values.

The hidden layer or the output layer j is depicted in Figure 5. The inputs, weighted sum, and activation function of output layer are shown in Figure 5. The connection weight between the input layer and hidden layer and hidden layer and output layer is shown in Figure 6. The connection weights can also be represented in the matrix form known as connection weight matrix.

3. Experiment, Results, and Discussion

The proposed method is implemented using the 32-bit Matlab 2015b environment on Intel(R) Core (TM) i3-2120, with a processing speed of 3.30 GHz and 2 GB of RAM running Microsoft Windows 7. Readers can repeat our results on any computer with which MATLAB is compatible.

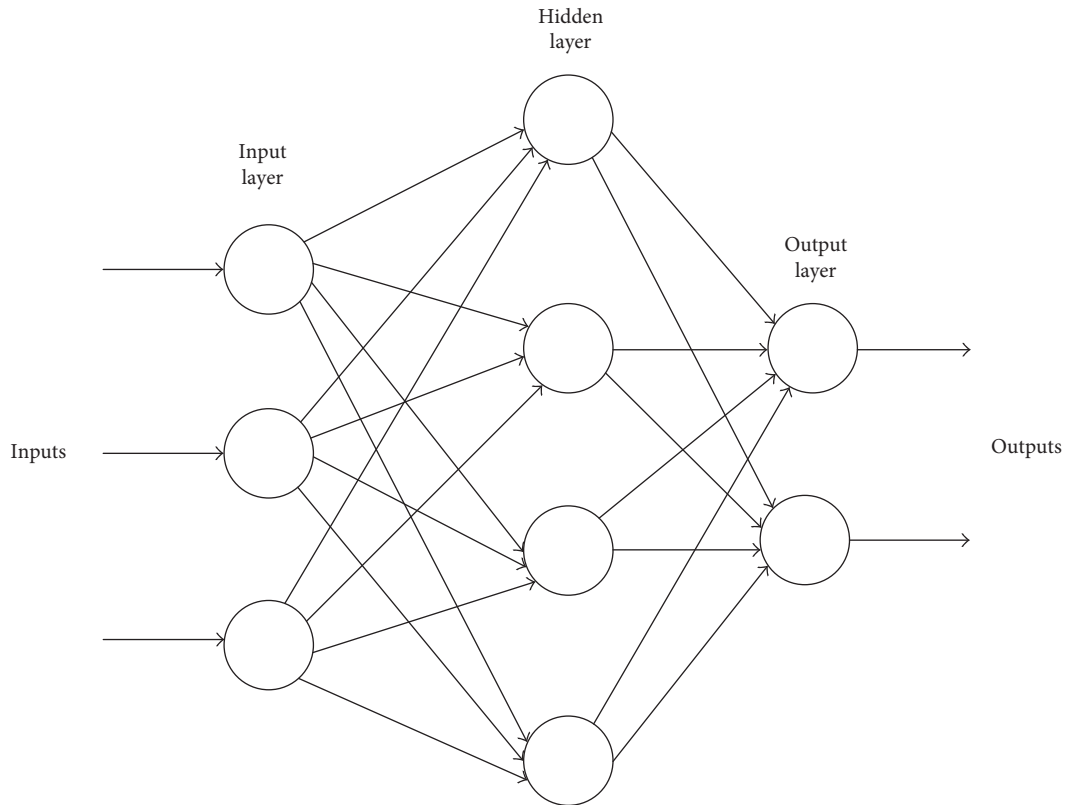


FIGURE 4: Architecture of a multilayer feed-forward neural network.

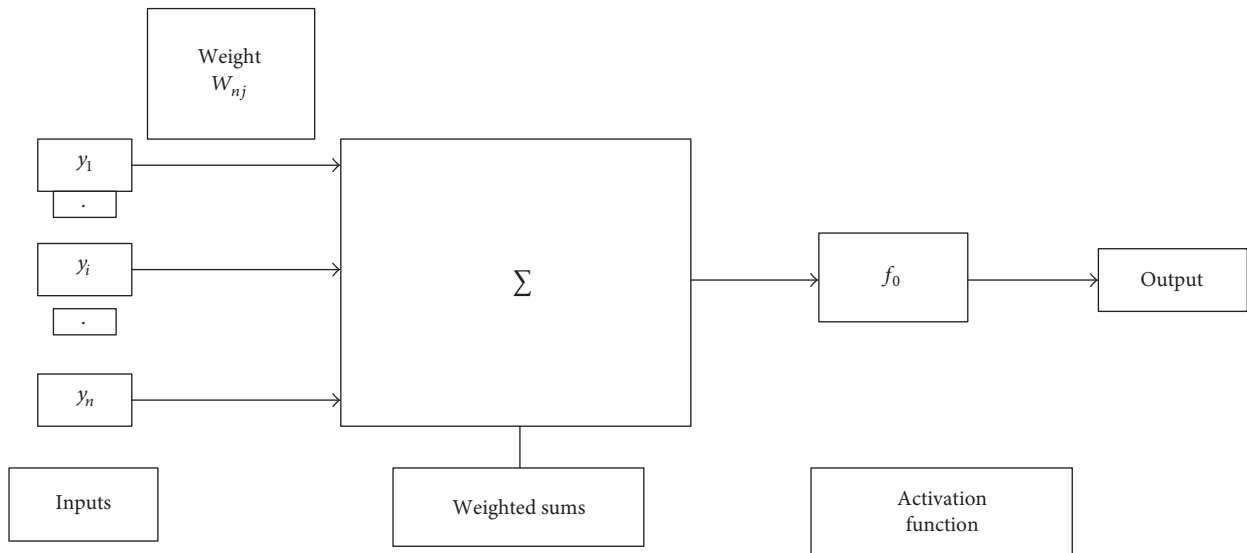


FIGURE 5: Hidden or output layer j . The input j are outputs from the previous layers. These are multiplied by their corresponding weights to configure a weighted sum. A nonlinear activation function is applied to the net input. [The inputs to input j are labeled y_1, y_2, \dots, y_n . If unit j was in the first hidden layer, then these inputs would correspond to the input tuple $I_1, I_2, I_3, \dots, I_n$.]

This article aims at developing a CAD of AD brain system with better performance. The pseudocode is listed in Table 4.

3.1. *Parameter Estimation for s .* It is always a major concern to find the optimum value of decomposition level s . We know

that a smaller s provides less information whereas a larger s provides more information to the classifier. In order to avoid overfitting problem, a smaller s is used. Here, we change the value of s from 1 to 5 with increment of 1 and check up the corresponding average accuracies with FNN. The one which gives the highest accuracy is the optimal value of s .

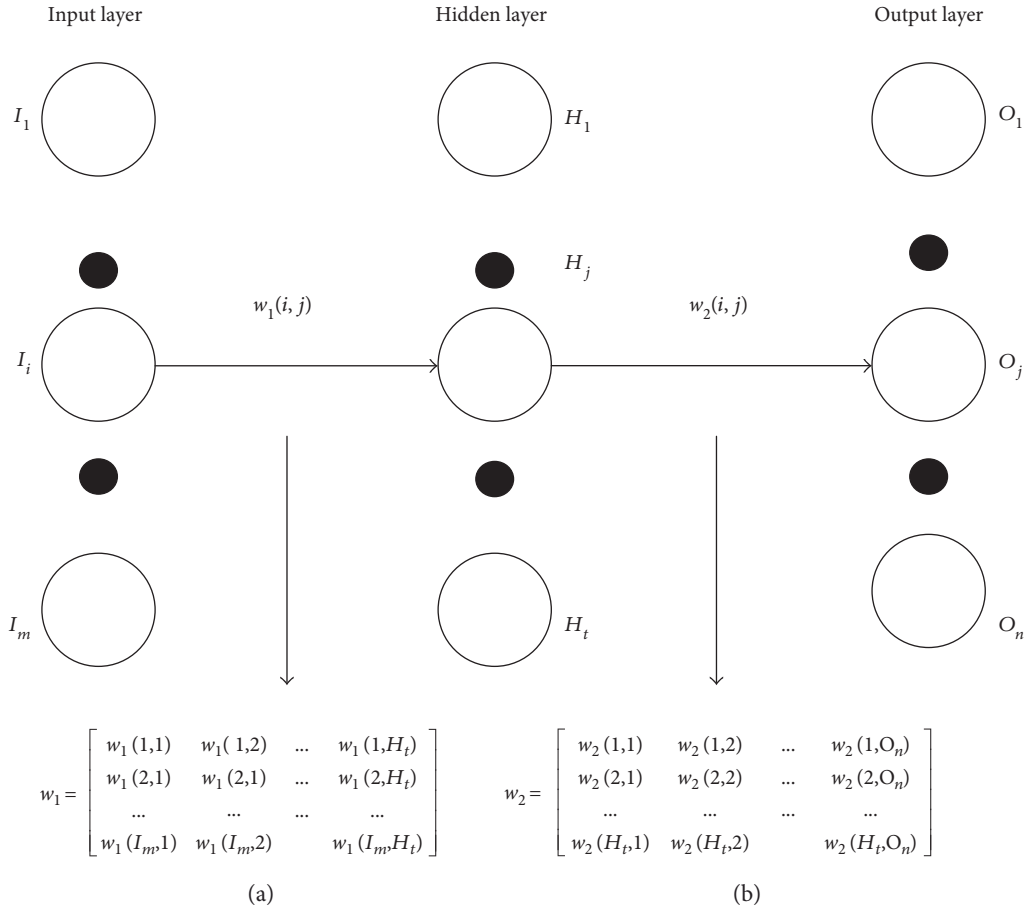


FIGURE 6: Connection weight matrix between (a) input layer and hidden layer and (b) hidden layer and output layer.

TABLE 4: Pseudocode of the proposed system.

Step 1: Import.

- (a) Import the OASIS dataset.
- (b) Ensure MRI as normal or abnormal brain.

Step 2: Resample the image into 256×256 .

Step 3: Compute 5-level DTCWT on the preprocessed images.

Step 4: Perform PCA on the obtained matrix. The selected number of principal component (PC) should preserve at least 90% of total variances.

Step 5: Train feed-forward neural network by taking input as reduced set of feature vectors and their corresponding class labels.

Step 6: Evaluation

- (a) Obtain the confusion matrix.
- (b) Calculate the classification accuracy and other essential parameters.

3.2. Feature Extraction. In this paper, we extract the DTCWT coefficients from the input images. The features of 5th resolution scales are selected because they provide higher classification performance than other resolution level scales. The DTCWT has a multiresolution representation as the wavelet transform does. For disease detection, it is preferable to use a few intermediate coefficient scales as the classifier input. The

lowest scales have lost fine signal details whereas the most highly detailed scales contain mostly noise. Therefore, we prefer to choose only a few intermediate scales for the DTCWT coefficients. These obtained coefficients are sent as input to the PCA.

3.3. Feature Reduction. Excessive features increase calculation time as well as memory storage. In addition, they sometimes make classification much more complicated, which is known as curse of dimensionality. In this article, we utilized PCA to decrease the number of features.

Therefore, the extracted feature from DTCWT is sent to the PCA for the feature reduction. For each image, there are 768 features after 5th level of decomposition. As we have employed 32 slices for each patient, the total number of features becomes 32×768 . Now, the image is reformed into a row vector of $1 \times 24,576$. The row vectors of 126 subjects are arranged into an "input matrix" with dimensions of $126 \times 24,576$. It is still too large for calculation. So, the input data matrix is now decomposed into the principal component "score matrix" and the "coefficient matrix." The score matrix size after decomposition is 126×125 . Here, the rows and columns of "score matrix" correspond to subjects and components, respectively.

The variance with the number of principal components from 1 to 18 is listed in Table 5. Experimenting with different

TABLE 5: Detailed data of PCA.

No. of prin. comp.	1	2	3	4	5	6	7	8	9
Variance (%)	63.18	72.15	77.08	80.28	83.05	84.55	85.68	86.59	87.47
No. of prin. comp.	10	11	12	13	14	15	16	17	18
Variance (%)	88.18	88.28	89.41	89.96	90.44	90.86	91.23	91.58	91.91

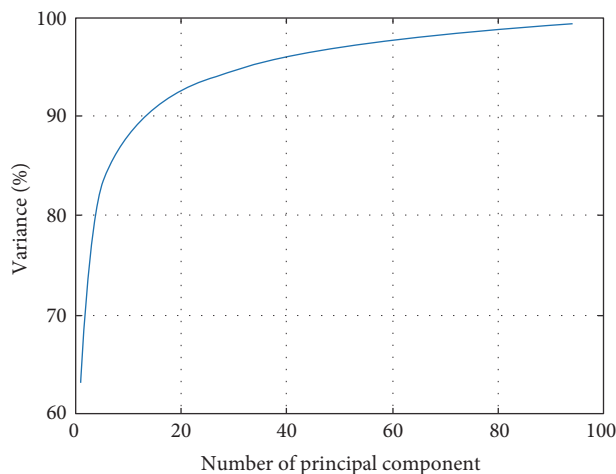


FIGURE 7: Variances versus number of principal component.

numbers of principal components (PCs) revealed that accuracy with PC=14 provided the best classification accuracy preserving 90.44% of the total variance. The curve of cumulative sum of variances with the number of principal component is shown in Figure 7. We did not set the energy threshold as 95% because that would cost too many features, along with computational burden.

3.4. BPNN Training. The 14 PCs are directly sent to BPNN. Thus, the number of input neurons N_I is 14. Then, the number of hidden layer neurons (N_H) is determined as 10 according to the information entropy method [32]. Therefore, the architecture of the neural network becomes 14-10-1. The SCG method is employed because it is extremely faster than BP, MBP, and ABP [28].

3.5. Performance Measures. There are several techniques to evaluate the efficiency of classifiers. The performance is calculated on the essence of the overall confusion matrix. It holds the correct and incorrect classification results. Table 6 shows a confusion matrix for binary classification, where TP, TN, FP, and FN depict true positive, true negative, false positive, and false negative, respectively, as illustrated in Table 7.

Here, AD brains are assumed to hold the value “true” and NC ones are assumed to hold the value “false” following normal convention.

The accuracy is the most accepted empirical measure to access effectiveness of classifier. It is formulated by

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (10)$$

TABLE 6: Confusion matrix for a binary classifier to discriminate between two classes (A_1 and A_2).

True class	Predicted class	
	A_1 (patients)	A_2 (controls)
A_1 (patients)	TP	FN
A_2 (controls)	FP	TN

TABLE 7: Evaluation indicators.

Indicator	Explanation
TP	True positive, anticipating an AD to AD
FP	False positive, anticipating an HC to AD
TN	True negative, anticipating an HC to HC
FN	False negative, anticipating an AD to HC

Sensitivity is the measure of the proportion of true positives that are correctly classified, and specificity is the measure of the proportion of negatives which are correctly classified. These are calculated by

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (11)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

The precision and the recall are formulated by

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

3.6. Statistical Analysis. In order to execute a strict statistical analysis, stratified cross-validation (SCV) is used. We apply a 10-fold CV technique in this experiment because of two reasons: (1) to make balance between reliable estimate and computational cost and (2) for providing a fair comparison because the common convention was to take the value of K equal to 10 [33].

A 10-fold CV means we have to divide our dataset randomly into ten mutually exclusively folds of approximately equal size and almost the same distribution. In each run, 9 subsets will be used for training, and the remaining one will be utilized for the validation. This process is repeated 10 times, in which every subset is utilized for validation once. The 10-fold CV is repeated 50 times; namely, a 50x 10-fold CV is implemented.

The accuracies, sensitivities, and specificities obtained from the 50 runs of 10-fold CV are presented in Table 8.

TABLE 8: Algorithm performance comparison for MRI brain image.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
DTCWT + PCA + FNN (proposed)	90.06 ± 0.01	92.00 ± 0.04	87.78 ± 0.04	89.6 ± 0.03
VBM + RF [38]	89.0 ± 0.7	87.9 ± 1.2	90.0 ± 1.1	N/A
DF + PCA + SVM [40]	88.27 ± 1.9	84.93 ± 1.21	89.21 ± 1.6	69.30 ± 1.91
EB + WTT + SVM + RBF [56]	86.71 ± 1.93	85.71 ± 1.91	86.99 ± 2.30	66.12 ± 4.16
BRC + IG + SVM [34]	90.00	96.88	77.78	N/A
BRC + IG + VFI [34]	78.00	65.63	100.00	N/A
Curvelet + PCA + KNN [35]	89.47	94.12	84.09	N/A
US + SVD-PCA + SVM-DT [36]	90.00	94.00	71.00	N/A

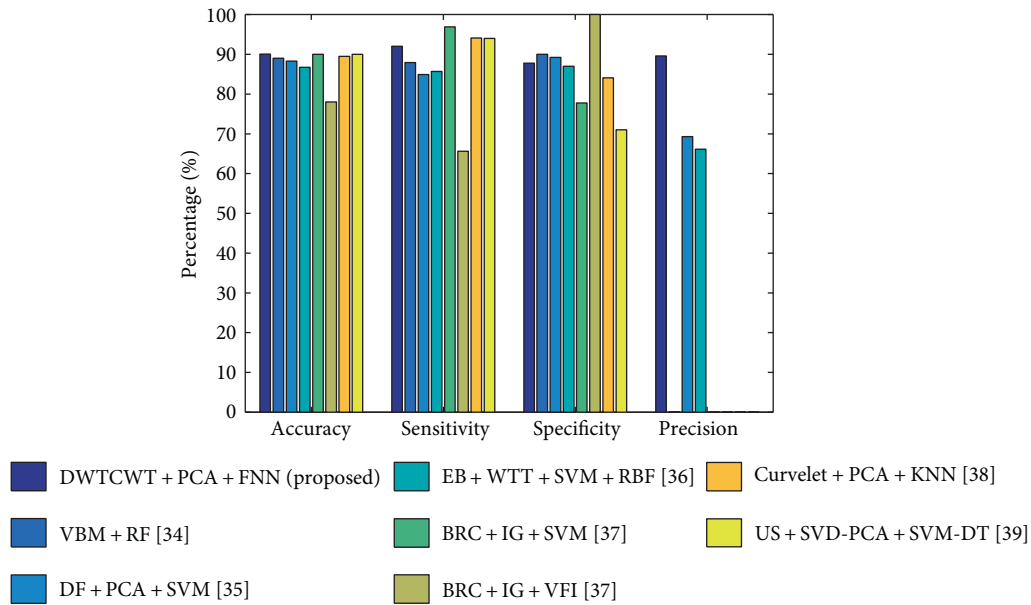


FIGURE 8: Bar plot of the algorithm comparison ([34–36, 38] did not mention its precision).

Our method achieved an accuracy of $90.06 \pm 0.01\%$, a sensitivity of $92.00 \pm 0.04\%$, a specificity of $87.78 \pm 0.04\%$, and a precision of $89.6 \pm 0.03\%$.

3.7. Comparison to Other State-of-the-Art Approaches. To further determine the effectiveness of the proposed “DTCWT + PCA + FNN,” we compared it with seven state-of-the-art approaches in Table 8. Some of these approaches utilized different statistical settings, making direct comparison difficult. The results in Table 8 show that study [34–37] did not present standard deviations (SD) of three standards. The specificities of study [34–36] are lower than those demonstrated by other methods. Therefore, these three methods are not worthy for further study. Similarly, study [37] obtained a classification specificity of 100%. In spite of its high specificity, both the accuracy and sensitivity achieved by this algorithm are poor. Hence, this method is also not considered further for the study. Three other methods reported both mean values and standard deviation values. They also achieved satisfying results. Study [38] obtained promising results because of the voxel-based morphometry (VBM). Indeed, VBM has frequently been employed to study

brain changes. Study [37] demonstrated that a taxi driver will normally have a larger back section of the posterior hippocampus. Study [39] concluded that global gray matter decreases linearly with old age but global white matter remains in the same amount. Nevertheless, VBM requires an accurate spatial normalization, or the classification accuracy may decrease significantly. Study [40] was based on a novel approach called the displacement field (DF). This study measured and estimated the displacement field of various slices between AD and HC subjects. There are other methods that have distinguished AD from HC; however, they dealt with images formed by other modalities: PET, SPECT, DTI, and so forth. Hence, they are also not considered in this study.

Finally, the proposed “DTCWT + PCA + FNN” achieved an accuracy of $90.06 \pm 0.01\%$, a sensitivity of $92.00 \pm 0.04\%$, a specificity of $87.78 \pm 0.04\%$, and a precision of $89.60 \pm 0.03\%$. With respect to classification accuracy, our approach outperforms five other methods and is almost equal to the accuracies of the remaining two methods that did not account for means and standard deviations. We also achieved a promising sensitivity and a promising specificity.

TABLE 9: Acronyms list.

Acronym	Definition
AD	Alzheimer's disease
HC	Healthy control
MR(I)	Magnetic resonance (imaging)
DTCWT	Dual-tree complex wavelet transform
PCA	Principal component analysis
FNN	Feed-forward neural network
DWT	Discrete wavelet transform
OASIS	Open Access Series of Imaging Studies
MMSE	Mini mental state examination
CDR	Clinical dementia rating
SNR	Signal-to-noise ratio
VBM	Voxel-based morphometry
RF	Random forest
DF	Displacement field
SVM	Support vector machine
EB	Eigenbrain
WTT	Welch's <i>t</i> -test
RBF	Radial basis function
BRC	Brain region cluster
IG	Information gain
KNN	<i>k</i> -nearest neighbor
ANN	Artificial neural network
SCV	Stratified cross-validation

Hence, our results are either better than or comparable to those of the other methods. The bar plot of the algorithm comparison is shown in Figure 8. The acronyms list is depicted in Table 9.

4. Conclusions and Future Research

We presented an automated and accurate method for AD identification based on a DTCWT, PCA, and FNN. The results showed that the proposed method achieved an accuracy of $90.06 \pm 0.01\%$, a sensitivity of $92.00 \pm 0.04\%$, a specificity of $87.78 \pm 0.04\%$, and a precision of $89.6 \pm 0.03\%$ and outperformed 7 state-of-the-art algorithms.

We will focus our future research on the following aspects: (i) testing other advanced variants of wavelet such as 3D-DTCWT, wavelet packet analysis, and fractional calculus; (ii) utilizing different feature reduction techniques such as independent component analysis (ICA) [41], linear discriminant analysis (LDA) [42], probabilistic PCA [43], or sparse-autoencoders [44]; (iii) testing our data with least-square techniques [45], kernel support vector machine (k-SVM), such as fuzzy SVM [46], radial basis function neural network (RBFNN) [47], deep learning methods such as convolutional neural network (CNN) [48], and other alternative pattern recognition tool for classification; (iv) utilizing advanced swarm intelligence techniques such as particle swarm optimization [49], artificial bee colony [50], genetic pattern search [51], ant colony optimization [52], and biogeography-based optimization [53] to find the optimal kernel; (v) testing

the proposed method on images obtained from different modalities such as computed tomography (CT) [54], ultrasound, spectrum imaging [55], and 3D MRI; (vi) utilizing other advance image preprocessing technique to enhance the classification performance, such as image denoising, image enhancement, and image segmentation; and (vii) classification may be carried out on the sparsity domain.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C7A1046050). And this study was financed by the research fund of Chosun University, 2017.

References

- [1] R. Brookmeyer, E. Johnson, K. Ziegler-Grahamb, and H. M. Arrighic, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186–191, 2007.
- [2] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiology of Aging*, vol. 32, no. 12, pp. 2322.e19–2322.e27, 2011.
- [3] A. Nordberg, J. O. Rinne, A. Kadir, and B. Langstrom, "The use of PET in Alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, pp. 78–87, 2010.
- [4] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4637–4642, 2004.
- [5] T. Magnander, E. Wikberg, J. Svensson et al., "A novel statistical analysis method to improve the detection of hepatic foci of 111 In-octreotide in SPECT/CT imaging," *EJNMMI Physics*, vol. 3, no. 1, p. 1, 2016.
- [6] P. De Visschere, M. Nezzo, E. Pattyn, V. Fonteyne, C. Van Praet, and G. Villeirs, "Prostate magnetic resonance spectroscopic imaging at 1.5 tesla with endorectal coil versus 3.0 tesla without endorectal coil: comparison of spectral quality," *Clinical Imaging*, vol. 39, no. 4, pp. 636–641, 2015.
- [7] E. D. Roberson and L. Mucke, "100 years and counting: prospects for defeating Alzheimer's disease," *Science*, vol. 314, no. 5800, pp. 781–784, 2006.
- [8] M. Tabaton, P. Odetti, S. Cammarata et al., "Artificial neural networks identify the predictive values of risk factors on the conversion of amnesic mild cognitive impairment," *Journal of Alzheimer's Disease*, vol. 19, no. 3, pp. 1035–1040, 2010.
- [9] B. S. Mahanand, S. Suresh, N. Sundararajan, and K. M. Aswatha, "Identification of brain regions responsible for Alzheimer's disease using self-adaptive resource allocation network," *Neural Networks*, vol. 32, pp. 313–322, 2012.
- [10] B. Jeurissen, A. Leemans, and J. Sijbers, "Automated correction of improperly rotated diffusion gradient orientations in diffusion weighted MRI," *Medical Image Analysis*, vol. 18, no. 7, pp. 953–962, 2014.

- [11] L. Hamelin, M. Bertoux, M. Bottlaender et al., "Sulcal morphology as a new imaging marker for the diagnosis of early onset Alzheimer's disease," *Neurobiology of Aging*, vol. 36, no. 11, pp. 2932–2939, 2015.
- [12] M. Á. Caballero, M. Brendel, A. Delker et al., "Mapping 3-year changes in gray matter and metabolism in Abeta-positive nondemented subjects," *Neurobiology of Aging*, vol. 36, no. 11, pp. 2913–2924, 2015.
- [13] G. Yang, Y. Zhang, J. Yang et al., "Automated classification of brain images using wavelet energy and biogeography-based optimization," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 15601–15617, 2016.
- [14] E. S. El-Dahshan, H. M. Mohsen, K. Revett, and A. B. Salem, "Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.
- [15] Y. Zhang, S. Wang, G. Ji, and Z. Dong, "An MR images classifier system via particle swarm optimization and kernel support vector machine," *The Scientific World Journal*, vol. 2013, Article ID 130134, 9 pages, 2013.
- [16] I. Álvarez, J. M. Górriz, J. Ramírez et al., "Alzheimer's diagnosis using eigenbrains and support vector machines," *Electronics Letters*, vol. 45, no. 7, pp. 973–980, 2009.
- [17] M. Z. Iqbal, A. Ghafoor, A. M. Siddhiqui, M. M. Riaz, and U. Khalid, "Dual-tree complex wavelet transform and SVD based medical image resolution enhancement," *Signal Processing*, vol. 105, pp. 430–437, 2014.
- [18] Y. Zhang, B. Peng, S. Wang et al., "Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection," *Scientific Reports*, vol. 6, article 21816, 2016.
- [19] D. K. Shin and Y. S. Moon, "Super-resolution image reconstruction using wavelet based patch and discrete wavelet transform," *Journal of Signal Processing Systems*, vol. 81, no. 1, pp. 71–81, 2015.
- [20] Y. Zhang, S. Wang, G. Ji, and Z. Dong, "Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 10, no. 1, pp. 116–117, 2015.
- [21] S. Beura, B. Majhi, and R. Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer," *Neurocomputing*, vol. 154, pp. 1–14, 2015.
- [22] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–353, 2001.
- [23] A. Barri, A. Doooms, and P. Schelkens, "The near shift-invariance of the dual-tree complex wavelet transform revisited," *Journal of Mathematical Analysis and Applications*, vol. 389, no. 2, pp. 1303–1314, 2012.
- [24] S. Wang, S. Lu, Z. Dong, J. Yang, M. Yang, and Y. Zhang, "Dual-tree wavelet transform and twin support vector machine for pathological brain detection," *Applied Science*, vol. 6, no. 6, p. 169, 2016.
- [25] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
- [26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2012.
- [27] D. Guo, Y. Zhang, Q. Xiang, and Z. Li, "Improved radio frequency identification method via radial basis function neural network," *Mathematical Problems in Engineering*, vol. 2014, Article ID 420482, 9 pages, 2014.
- [28] Y. Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10049–10053, 2011.
- [29] M. Manoochehri and F. Kolahan, "Integration of artificial neural network and simulated annealing algorithm to optimize deep drawing process," *The International Journal of Advanced Manufacturing Technology*, vol. 73, no. 1, pp. 241–249, 2014.
- [30] S. U. Aswathy, G. G. D. Dhas, and S. S. Kumar, "A survey on detection of brain tumor from MRI brain images," in *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pp. 871–877, Kanyakumari, 2014.
- [31] A. Poursamad, "Adaptive feedback linearization control of anticlock braking systems using neural networks," *Mechatronics*, vol. 19, no. 5, pp. 767–773, 2009.
- [32] H. C. Yuan, F. L. Xiong, and X. Y. Huai, "A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy," *Computers and Electronics Agriculture*, vol. 40, no. 1–3, pp. 57–64, 2003.
- [33] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 21–31, 2014.
- [34] C. Plant, S. J. Teipel, A. Oswald et al., "Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease," *NeuroImage*, vol. 50, no. 1, pp. 162–174, 2010.
- [35] D. Jha and G. R. Kwon, "Alzheimer disease detection in MRI using curvelet transform with KNN," *The Journal of Korean Institute of Information Technology*, vol. 14, no. 8, 2016.
- [36] Y. Zhang, S. Wang, and Z. Dong, "Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree," *Progress in Electromagnetic Research*, vol. 144, pp. 171–184, 2014.
- [37] E. A. Maguire, D. G. Gadian, I. S. Johnsrude et al., "Navigation-related structural change in the hippocampi of taxi drivers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 8, pp. 4398–4403, 2000.
- [38] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, 2013.
- [39] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston, and R. S. J. Frackowiak, "A voxel-based morphometric study of ageing in 465 normal adult human brains," *NeuroImage*, vol. 14, no. 1, Part 1, pp. 21–36, 2001.
- [40] Y. Zhang and S. Wang, "Detection of Alzheimer's disease by displacement field and machine learning," *PeerJ*, vol. 3, article e1251, 2015.
- [41] S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, and T. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 20, pp. 10979–10984, 1997.
- [42] A. J. Izenman, "Linear discriminant analysis," in *Modern Multivariate Statistical Techniques*, pp. 237–280, Springer, New York, 2013.

- [43] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.
- [44] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [45] K. Polat, S. Güneş, and A. Arslan, "A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine," *Expert Systems with Applications*, vol. 34, no. 1, pp. 482–487, 2008.
- [46] S. Wang, X. Yang, Y. Zhang, P. Phillips, J. Yang, and T.-F. Yuan, "Identification of green, oolong and black teas in china via wavelet packet entropy and fuzzy support vector machine," *Entropy*, vol. 17, no. 10, pp. 6663–6682, 2015.
- [47] M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 697–710, 2002.
- [48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.
- [49] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 22–31, 2014.
- [50] Y. Zhang, L. Wu, and S. Wang, "Magnetic resonance brain image classification by an improved artificial bee colony algorithm," *Progress in Electromagnetics Research*, vol. 116, pp. 65–79, 2011.
- [51] Y. Zhang, S. Wang, G. Ji, and Z. Dong, "Genetic pattern search and its application to brain image classification," *Mathematical Problems in Engineering*, vol. 2013, Article ID 580876, 8 pages, 2013.
- [52] C. Pang, G. Jiang, S. Wang et al., "Gene order computation using Alzheimer's DNA microarray gene expression data and the ant colony optimisation algorithm," *International Journal of Data Mining and Bioinformatics*, vol. 6, no. 6, pp. 617–632, 2012.
- [53] S. Wang, Y. Zhang, G.-L. Ji, J.-G. Wu, and L. Wei, "Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization," *Entropy*, vol. 17, no. 8, pp. 5711–5728, 2015.
- [54] Y. Chen, Y. Zhang, J. Yang et al., "Curve-like structure extraction using minimal path propagation with backtracking," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 988–1003, 2016.
- [55] Z. Dong, Y. Zhang, F. Liu, Y. Duan, A. Kangarlu, and B. S. Peterson, "Improving the spectral resolution and spectral fitting of H MRSI data from human calf muscle by the SPREAD technique," *NMR in Biomedicine*, vol. 27, no. 11, pp. 1325–1332, 2014.
- [56] Y. Zhang, Z. Dong, P. Phillips et al., "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Frontier in Computational Neuroscience*, vol. 9, p. 66, 2015.