Prediction of Conversion to Alzheimer's Disease Using 3D-DWT and PCA

Li Yew Aow Yong[1](✉), Mohd Shafry Mohd Rahim[1], and Chi Wee Tan[2]

[1] School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
ylyaow2@live.utm.my

[2] Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia

Abstract. Alzheimer' Disease (AD) is the most common form of dementia worldwide. Structural Magnetic Resonance Imaging (sMRI) is the supportive tool for the diagnosis of this disease. Even, it can be used to predict the conversion of the disease from the mild cognitive impairment (MCI) to AD stage. Nevertheless, the 3D image produced by sMRI is high dimensional data, which raises the risk of overfitting in the classification model. For this reason, the combination of Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) was proposed as the feature extraction techniques to reduce the dimensional and extract significant features concurrently. The issues of DWT are the selection of level of decomposition and wavelet filter to decompose the image. In order to deal with these issues, a series of experiments were conducted to find the suitable parameters. By using 2D-DWT, spatial information of 3D data cannot be captured. The connection between the slices is neglected. Hence, 3D-DWT has been adopted instead of 2D-DWT in this paper. In the classification step, Support Vector Machine (SVM) was used as the classifier to predict the conversion of normal control (NC) and stable MCI (SMCI) to progressive MCI (PMCI) and AD for datasets collected up to 2 years before the progression. The dataset used in this paper was collected from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In the validation, the proposed method outperformed the other methods by attaining 79%, 79%, 82% and 82% in accuracy for the datasets collected at different time points, which were 1% to 4% higher than the model adopted 2D-DWT and PCA.

Keywords: Alzheimer's Disease · Structural MRI · DWT · PCA

1 Introduction

Alzheimer's Disease (AD) is a degenerative brain sickness which eventually leads to death due to complications [1]. The death rate of AD was increased 146% from year of 2000 to 2018, while other diseases such as heart's disease had decreased in United States. The patient suffers AD experiences cognitive and behavioral impairment such as memory loss, difficulty in thinking and reasoning, and personality change. Hence, it

S. Spinsante et al. (Eds.): HealthyIoT 2021, LNICST 432, pp. 199–213, 2022.
https://doi.org/10.1007/978-3-030-99197-5_16

requires a lot of time and money in taking care the patient suffers AD. In light of the AD facts and figures, it is becoming extremely difficult to ignore the existence of AD. The early prediction of the conversion to AD is getting more attention to provide a more comprehensive treatment to the patient. Also, it prepares the heart of the caregiver to provide emotional support and daily support to the patient.

There have been a number of longitudinal studies involving Structural Magnetic Resonance Imaging (sMRI) that have been reported to be the supportive tool in AD diagnosis [2]. MRI is a non-invasive brain imaging technique [3], therefore, it is more preferable compared to other brain imaging techniques as well. SMRI reveals the internal structure of the brain. Hence, the atrophy of the brain can be detected through examining the 3D image produced from the device. The longitudinal study of MRI images of a patient provides more information to the doctor on the formation of the disease [4]. On the other hand, a single scan also allows the doctor to identify the existence of AD with the help of brain imaging and neuropsychological tests.

There are several stages in the development of AD. The healthy patients are categorized as normal control (NC), the patients suffer mild cognitive impairment (MCI) but they do not convert to AD after a period of observation are categorized as stable mild cognitive impairment (SMCI), the patients suffer MCI and they converts to AD after a period of observation are categorized as progressive MCI (PMCI), and AD refers to the patients are having AD at the baseline scan. The prediction of the conversion to AD in computer-aided diagnosis involves classifying the different stages of the disease, especially on segregating PMCI from SMCI to prescribe the right medicine to the patient.

Nevertheless, the brain image produced from sMRI is high dimensional data. Each 3D image contains millions of features, and it causes overfitting. Hence, feature extraction is an important procedure before classification. This paper has proposed Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) to perform feature extraction. Some of the researchers had adopted 2D-DWT instead of 3D-DWT in their studies. However, 2D-DWT faces the issue of losing spatial information. It does not consider the connection between the slices of 3D image during the data compression [5]. Besides, the main challenges of DWT are selecting suitable wavelet and decomposition level. Different kinds of datasets require different parameter selection. Hence, 3D-DWT was adopted in this paper. In order to deal with the parameter selection issue, experiments were conducted to examine the performance of different parameters towards the AD classification.

The main contribution of this paper is to identify the best parameters for 3D-DWT on the AD image. It ensures the extracted features increase the accuracy of classification. The results in the comparison demonstrate that the proposed method has achieved better performance compared to other models. The rest of the paper is organized as follows: Sect. 2 presents the prior study, Sect. 3 describes the proposed method, Sect. 4 discusses the implementation and evaluation, Sect. 5 provides the results and discussion, and Sect. 6 draws the conclusion for this paper.

2 Prior Study

The selection of feature extraction approaches in AD classification is based on the features used in the studies. The Gray Level Co-occurrence Matrix (GLCM) was widely

used in extracting texture features of the brain image [6–8]. Tooba Altaf et al. [6] computed the texture features such as entropy, contrast, correlation and homogeneity by applying GLCM in each slice of the 3D MR image. Then, the average of the features was obtained to form a feature vector for each MR image. Besides, the authors also adopted scale invariant feature transform, local binary pattern and histogram of oriented gradient to extract texture features. Besides extracting texture features, Arpita Raut and Vipul Dalal [7] also extracted shape and area features by using moment invariants. A total of 13 features were extracted from the image. C.V. Dolph et al. [8] had extracted the features such as volume, surface area and cortical thickness by using FreeSurfer software. The texture feature was extracted through the proposed novel fractal texture feature extraction, which was the combination of fractal dimension and GLCM.

Apart from the methods aforementioned, voxel-based morphometry (VBM) is also another approach to compare the differences between the groups in voxel-wise [9]. The VBM approach proved that grey matter alternation is consistent with the neuropsychological test. It also found that the healthy patient who faces cognitive impairment in future has reduced grey matter density compared to the patient who remains normal cognitive [10]. Thus, the voxel-wise comparison was also widely used in AD classification. The VBM analysis can be conducted by using different software, such as FSL-VBM and Statistic Parameter Mapping (SPM) [10–12]. The number of VBM features is usually higher than the samples, therefore, Principal Component Analysis (PCA) and Partial Least Square (PLS) were commonly used to further extract and reduce the features [13, 14]. Both techniques were always compared, and PLS outperformed PCA in most studies. L. Khedher et al. [14] drew the graphs to compare the classification result by using different numbers of features. It showed that the number of features is the key to determine the classification results for both techniques.

Besides, it is noticed that the combination of PCA with DWT has improved the classification result. Yudong Zhang et al. [15] proposed 3D-DWT to extract the wavelet coefficients of the image. Then, the volume descriptors were calculated through the subbands obtained from each level of decomposition. PCA was used to find the uncorrelated features. This study showed that 3D-DWT outperformed 2D-DWT by preserving the spatial information of 3D images. Luis Javier Herrera et al. [5] performed 2D-DWT and PCA in their study. However, the results showed that the combination with PCA decreased the classification result. This might be due to the number of features selected for PCA having impacts on the results. Apart from that, both studies adopted DWT with different wavelets. In view of the controversial claims, this study aims to give a comprehensive view on the selection of suitable wavelet and level of decomposition for AD classification. However, it is impossible to compare the results with the previous studies because different features and different population studies have been used in the study. By using different scope, it brings great impact on the classification results. This study has conducted a voxel-wise comparison between the groups without using any descriptors. As a result, this study validated the proposed method by comparing the classification results of different parameters, and the evaluation of different models was done on the collected datasets.

There were several classifiers that have been used by the research in AD classification. Support Vector Machine (SVM) was one of the widely used classifiers [13, 14, 16]. It

involved linear SVM, kernel SVM or other variations of SVM. It has been proved that it is powerful in dealing with the AD classification by achieving good results, which were 80% and above accuracy in most of the cases. SVM was popularly used in binary classification, in contrast, deep learning approach especially convolutional neural network was commonly used for multiclass classification [8, 17]. However, parameter tuning is the main concern of deep learning approaches because there is lack of guidelines on determining the layers in between [18, 19]. A trial-and-error process is required to determine the number of hidden layers and neurons. Deep learning approach is treated as an one stop solution for AD classification, which has included feature extraction in the layers. Therefore, this paper chose to adopt SVM as classifier to perform binary classification.

3 Proposed Method

In this section, the proposed feature extraction techniques of this paper are presented. The AD classification scheme involved three steps, which were pre-processing, feature extraction and classification (see Fig. 1). The pre-processing was done through Computational Anatomy Toolbox (CAT12) software to segment the gray matter (GM) tissue with default parameters. The images were normalized to Montreal Neurological Institute (MNI) space by using Diffeomorphic Anatomic Registration Through Exponentiated Lie algebra algorithm (DARTEL) template to allow further analysis. The 3D images which contained the segmented GM tissue were transformed to $121 \times 145 \times 121$ voxels. After this, the 3D images underwent feature extraction by using 3D-DWT and PCA. The low dimensional data obtained from the feature extraction was the input of linear SVM classifier. The details of feature extraction and classification methods are described in the following sub-sections.

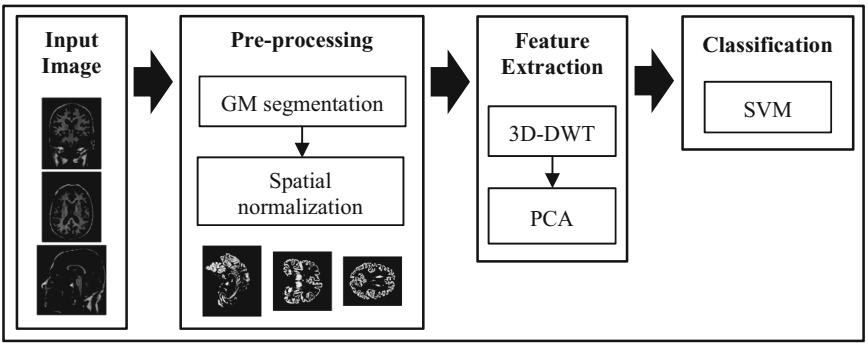

Fig. 1. Alzheimer's disease classification scheme

3.1 3D-Discrete Wavelet Transform

DWT is a time-frequency analysis which is used for data denoising or data compression by choosing appropriate frequency bands. In image processing, the image is treated as

signal and edges, which provide the frequency information to the algorithm [20]. The image is passed into a low pass filter and a high pass filter to find the image details. DWT compresses the data, at the same time, it extracts the significant features in the down sampling process. In order to decide the level of decomposition and wavelet for 3D-DWT, experiments were done to examine the results by using different parameters. The results are reported in Sect. 5. At last, 3-level 3D-DWT with Haar wavelet was adopted in this paper as the first step of feature extraction.

The 3-level 3D-DWT indicates that it involves three times decomposition at the approximation coefficients obtained from previous level of decomposition, and it takes part of three axes of the data. Figure 2 shows the diagram of 3-level 3D-DWT, where L refers to low pass filter, H refers to high pass filter, and downward arrow refers to down sampling process.

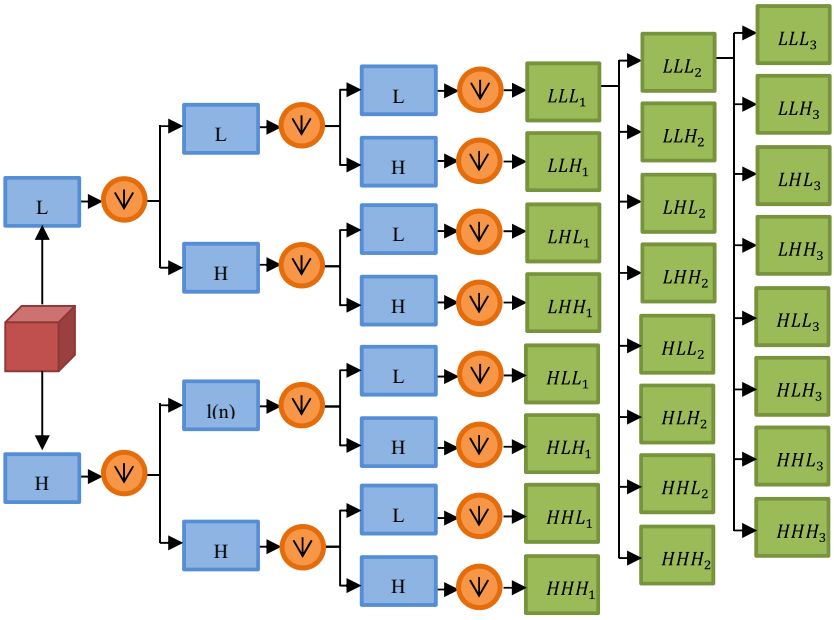

Fig. 2. Diagram of 3-level 3D-DWT

The 3D-DWT can be done through passing the 3D image in row, column and slice direction to the filter. Therefore, it produces 8 subbands named as LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH. The subband LLH is obtained through passing the image to low pass filter along x-axis, low pass filter along y-axis and high pass filter along z-axis. The other subbands also can be interpreted through this concept. Figure 3 provides the examples of the subbands obtained from 1-level 3D-DWT on the sagittal image. The subband LLL produces an approximation coefficient of the image, and it is used for further decomposition. On the other hand, the other subbands capture the detailed coefficients of the image by passing through a high pass filter in any of the direction.

The image is down sampled by 2, therefore, the size of the image in each dimension from each level of decomposition is half of the previous level.

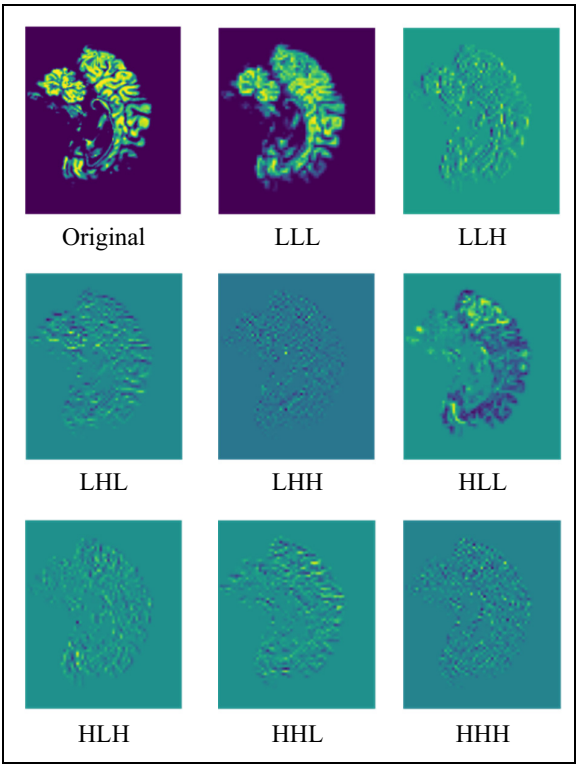

Fig. 3. The subbands of 3D-DWT on medical image

The Haar wavelet is considered as Daubechies 1 wavelet as well. It is the simplest wavelet but it is suitable for the AD images as compared to other wavelet filters [5]. The Haar wavelet takes the sum of successive pairs of pixels' values in each axis of the image, then multiplying it with the normalization constant, $1/\sqrt{2}$. This forms the 3D low pass filter matrix, LLL as shown in Eq. (1). The low pass filter in this paper had size of ($2 \times 2 \times 2$), which produced the coarser image by examining 8 pixels' values at a time until the whole image was decomposed. The subband LLL was used as the input of PCA. As a result, the 3-level 3D-DWT reduced the size of pre-processed image to 16 $\times$ 19 $\times$ 16 dimensions, which was 4,864 features.

$$LLL = \left[\left[\left[\frac{1}{2\sqrt{2}}\frac{1}{2\sqrt{2}}\right]\left[\frac{1}{2\sqrt{2}}\frac{1}{2\sqrt{2}}\right]\right]\left[\left[\frac{1}{2\sqrt{2}}\frac{1}{2\sqrt{2}}\right]\left[\frac{1}{2\sqrt{2}}\frac{1}{2\sqrt{2}}\right]\right]\right] \tag{1}$$

3.2 Principal Component Analysis

PCA is a linear transformation approach which projects the image data points to lower dimensional space. PCA finds the lines that maximize the variance of the data to obtain the minimal representation of the original image. The lines are called principal axes. Principal components are obtained through projecting the data points to the principal axes. The first principal component has the largest variance of the data, and the variance goes down to the last principal component.

Eigen-decomposition is the classical way to find the direction and magnitude to project the data. However, it might face accuracy loss due to eigenvalues are sensitive to perturbations in some matrices [21]. Any changes of the matrices will lead to great changes in eigenvalues. Therefore, singular value decomposition (SVD) was used as a replacement for eigen-decomposition in this paper due to its numerically stable in computation.

The 3D image from DWT was transformed to 1D long feature vector, and all the training set samples were stacked together to form the input of PCA. The input data has undergone mean-centring to decompose to $U\Sigma V^T$. The columns of U are called left singular vector, which contains the unit vectors, the diagonal of Σ are singular values, which contains the sum of projection lengths, and the columns of V are called singular vector, which contains the projection directions. The rank of mean-centring data, X_{mean} is $r = \min(m, n)$ to gain non-zero singular values, where m is number of samples and n is the number of dimensions of each sample. Consequently, the full SVD can be simplified to economy SVD as illustrated in Fig. 4.

The equation for economy SVD is derived as Eq. (2), where U is m-by-m matrix, $\widehat{\Sigma}$ is m-by-m matrix and $\widehat{V}^T$ is m-by-n matrix. Then, the covariance matrix can be computed from Eq. (3). By obtaining the covariance matrix through SVD, the projection of the input data was done through multiplying with $\widehat{V}^T$ to obtain the low-dimensional uncorrelated data.

$$X_{mean} = U\Sigma V^T = U\begin{bmatrix}\widehat{\Sigma}\\ 0\end{bmatrix}V^T = U\widehat{\Sigma}\begin{bmatrix}\widehat{V}^T\widehat{V}^{\perp^T}\end{bmatrix} \tag{2}$$

$$X_{mean}{}^T X_{mean} = V\Sigma U^T U\Sigma V^T = \widehat{V}\widehat{\Sigma}^2\widehat{V}^T \tag{3}$$

Each principal component refers to the transformed feature for classification. The dataset used in this paper had 160 samples for training set, as a result, the number of features after PCA was 159. All the features were kept in this paper to avoid eliminating the useful features, since low variance principal components also contributed to segregate the classification groups.

3.3 Support Vector Machine

In this paper, the only classifier was soft margin linear SVM. The soft margin SVM can deal with the situation where there are outliers fall in the areas belong to different groups. This makes it more suitable compared to hard margin SVM when the classifier deals with real-world data because the data might not distribute linearly. SVM finds the

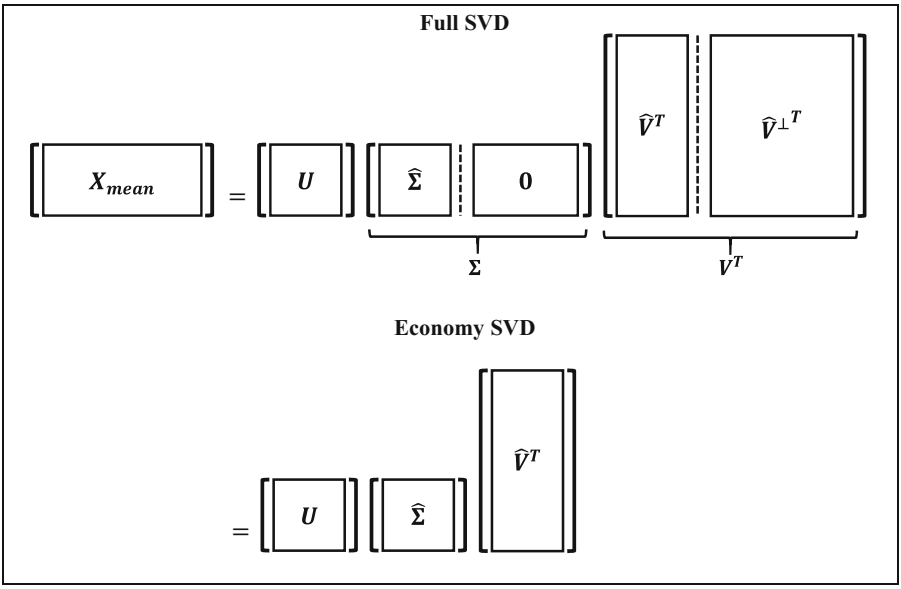

Fig. 4. Full and economy SVD

hyperplane which separates different groups. It needs to fulfill two requirements, which are having the largest distance with the nearest data and minimizing the classification loss. The hyperplane is found through following Eq. 4 subjects to Eq. 5, where the w denotes vector, x denotes features or variables of the image, b denotes the biased term, y denotes the predicted label, ξ refers to the distance between outlier and the correct margin, and n refers to the total number of features.

$$\min_{w,b} \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \xi_i \tag{4}$$

$$y_i\left(w^T x_i + b\right) \geq 1 - \xi_i \tag{5}$$

4 Implementation and Evaluation

Based on the proposed method in Sect. 3, the experiments were conducted to identify suitable wavelet and level of decomposition for 3D-DWT. The dataset, evaluation method and evaluation metric are described in this section.

4.1 Dataset Acquisition

The datasets were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The ADNI was launched in 2003 as a public-private partnership,

led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

The datasets were collected at different time points, which were 24 months before stable diagnosis, 18 months before stable diagnosis, 12 months before stable diagnosis and at the stable diagnosis time point. Each dataset consisted of 50 NC, 50 SMCI, 50 PMCI and 50 AD. The dataset collected during stable diagnosis time point refers to the patients who had ended the observation periods which lasted for 24 months. The subjects in the SMCI category refer to the patients remaining in MCI state at the end of the observation, while the subjects in the PMCI category refer to the patients convert to AD even though they had been diagnosed as MCI during baseline scan.

All the MR images obtained were pre-processed T1-weighted MR images, which had undergone 3D gradient inhomogeneity correction and B1 non-uniformity correction [22, 23]. After that, the images were further processed by using CAT12 as mentioned in Sect. 3. This paper conducted binary classification, which divided the different categories into two groups. The NC and SMCI was grouped together as (NC + SMCI) to represent the healthier patient, and the PMCI and AD was grouped together as (PMCI + AD) to represent the sick patient. Hence, the prediction on the conversion can be done by comparing the binary groups.

4.2 K-Fold Cross-Validation

Cross-validation allows assessing the model without bias. A 5-fold cross-validation was employed in this paper to perform out-of-sample assessment on the model. The assessment was done on the independent set. The 5-fold cross-validation divided the datasets to five subsets randomly. The training set was built with four subsets and the validation set was constructed with the last subset. This procedure was repeated for 5 times, and the different classification groups were divided equally in each cross-validation partition. All datasets were divided into training set and validation set before PCA instead of classification. This is because PCA is a machine learning approach which requires a training set to find the projection direction first. Then, the validation set will be projected based on the training set's projection direction.

4.3 Evaluation Metric

The aim of this paper is to improve the accuracy of AD classification. Therefore, the evaluation metric used in this paper was mean accuracy from the cross-validation. After obtaining the accuracies from each cross-validation, the average of the accuracies was calculated to be the overall result. Accuracy gauges the ability of the model in correctly predicting the conversion of the disease in this paper context [24]. It calculates the number of correct predictions from the total number of predictions. Accuracy is derived from the confusion matrix, which calculates the true positive (TP), true negative (TN),

false positive (FP) and false negative (FN) in the predictions. The confusion matrix is demonstrated in Table 1, and the accuracy is calculated by using Eq. 6.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \tag{6}$$

Table 1. Confusion matrix for AD classification

		Predicted class	
		(PMCI + AD)	(NC + SMCI)
Actual class	(PMCI + AD)	TP	FN
	(NC + SMCI)	FP	TN

TP = the number of (PMCI + AD) subjects is identified as (PMCI + AD).
TN = the number of (NC + SMCI) subjects is correctly identified as (NC + SMCI).
FP = the number of (NC + SMCI) subjects is identified as (PMCI + AD).
FN = the number of (PMCI + AD) subjects is identified as (NC + SMCI).

5 Results and Discussion

The purpose of conducting the experiments was to identify the suitable wavelet filter and level of decomposition for AD classification. Therefore, the mean accuracy from the cross-validation was the measurement in these experiments. The experiments had compared different popular wavelets, which included Haar, Daubechies (Db) 2, Db4, Symlet (Sym) 2 and Sym4 with different levels of decomposition of 3D-DWT on the four datasets collected at different time points. The initial decomposition was set to 2-level. The criteria to stop the increment of the decomposition level was the dropping of the mean accuracy.

Table 2 shows the 2-level 3D-DWT with different wavelets. The results revealed that the selection of wavelets had a great impact on classification results especially on the dataset collected at time points of 24 months and 18 months before stable diagnosis. It is challenging to classify the datasets collected at the earlier time points. The mean accuracies of the datasets collected at time points of 24 months and 18 months before stable diagnosis were always lower than the datasets collected at later time points. Despite this, Sym4 wavelet achieved the best results compared to other wavelets in 2-level 3D-DWT.

Table 3 shows that 3-level 3D-DWT by using Sym4 wavelet and Haar wavelet achieved higher mean accuracies compared to 2-level 3D-DWT with Sym4 wavelet at all time points except the dataset collected at time point of 18 months before stable diagnosis. The lowest mean accuracy of 3-level 3D-DWT with Haar wavelet on different datasets was 1% higher compared to Sym4 wavelet. Nevertheless, the Sym4 wavelet achieved 1% higher in the highest accuracy.

Table 2. The classification results with 2-level 3D-DWT and PCA

	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
Haar	0.77	0.78	0.80	0.81
Db2	0.78	0.77	0.80	0.81
Db4	0.76	0.80	0.81	0.81
Sym2	0.78	0.77	0.80	0.81
Sym4	0.77	0.80	0.81	0.81

Table 3. The classification results with 3-level 3D-DWT and PCA

	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
Haar	0.79	0.79	0.82	0.82
Db2	0.78	0.78	0.82	0.80
Db4	0.76	0.77	0.83	0.80
Sym2	0.78	0.78	0.82	0.80
Sym4	0.80	0.78	0.81	0.83

Table 4 shows that the classification dropped significantly in the 4-level 3D-DWT on most of the wavelets except Db4. However, the overall classification results of 4-level 3D-DWT by using Db4 did not show improvement compared to 3-level 3D-DWT by using Haar or Sym4 wavelet. Therefore, the experiment was stopped at 4-level 3D-DWT. The possible reason for dropping the classification result is the significant features for AD classification will be omitted when it is up to 4-level decomposition. DWT is a fine-to-coarse method, it is a trade-off relationship between compression ratio and the accuracy, due to the details of the image are being eliminated in each level of decomposition.

Table 4. The classification results with 4-level 3D-DWT and PCA

	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
Haar	0.72	0.76	0.77	0.82
Db2	0.75	0.78	0.77	0.80
Db4	0.80	0.77	0.80	0.82
Sym2	0.75	0.78	0.77	0.80
Sym4	0.76	0.76	0.75	0.78

In view of the results, this paper proposes to use 3-level 3D-DWT with Haar wavelet. Even though Sym4 can achieve higher results in certain datasets, but it is more important to increase the lowest accuracy obtained from different datasets to ensure more cases can be predicted correctly. The output of 3-level 3D-DWT on the image is demonstrated in Fig. 5. The image is given with the axes to show the size of the image in pixels after each decomposition level. As aforementioned, the size of the output image in each dimension is half of the original size. Hence, the output of first level of decomposition was (61 × 73 × 61), as the original size was (121 × 145 × 121). The down-sampling process was repeated in the second and third level of decomposition. Eventually, there was 4,864 features left after DWT. The number of features of the image was further reduced by using PCA. There were 159 features to be fed into classifier.

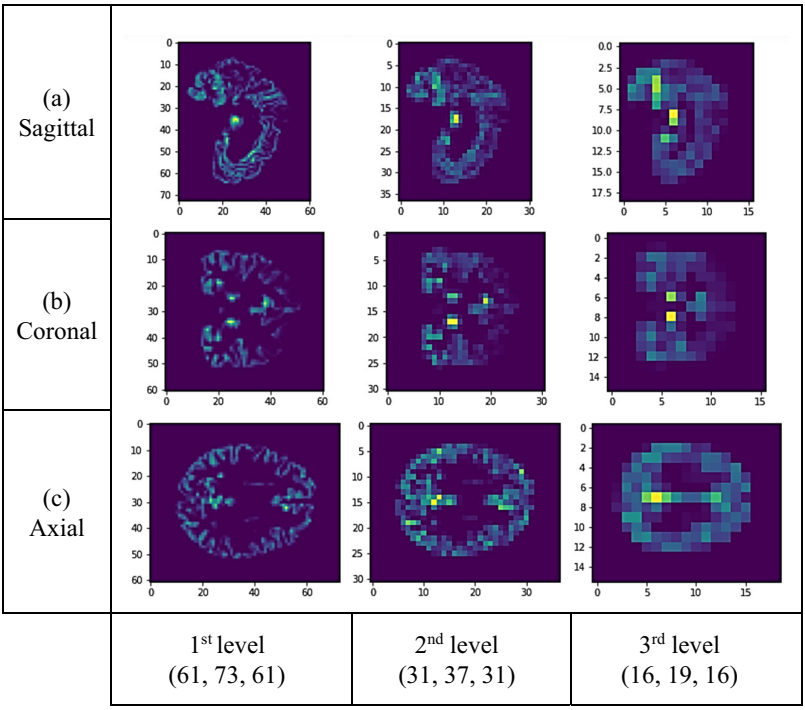

Fig. 5. The output of 3-level 3D-DWT

It is impossible to benchmark with the literature provided in this paper due to different features and different datasets were used in the studies. Therefore, a comparison of different models was conducted with the same datasets are demonstrated in Table 5. The first model adopted PCA and SVM, the second model applied 2D-DWT, PCA and SVM, and the proposed model implemented 3D-DWT, PCA and SVM. It is noticed that applying 2D-DWT and PCA as the feature extraction techniques did not increase the mean accuracy, but it dropped the results on the datasets collected at time point of 18 months and 12 months before stable diagnosis. On the other hand, 3D-DWT obtained

2%–4% higher compared to use PCA and SVM only on most of the datasets except the dataset collected at time point of 18 months before stable diagnosis. Even though there is an exception, but 3D-DWT still gives a promising result on most of the datasets. The results also indicated that 3D-DWT compresses the image more accurately compared to 2D-DWT. The significant features can be captured from 3D-DWT.

Table 5. The comparison of different models

	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
PCA	0.75	0.80	0.79	0.80
2D-DWT + PCA	0.76	0.78	0.78	0.80
Proposed	0.79	0.79	0.82	0.82

6 Conclusion

This paper has proposed to apply 3-level 3D-DWT with Haar wavelet and PCA as the feature extraction for AD classification. There were different wavelet families used in previous study, this paper has clarified that the most suitable wavelet is Haar wavelet. Besides, the maximum level of decomposition shall set to 3 because the significant features will be eliminated along the decomposition process. By using the suitable wavelet and level of decomposition, 3-level 3D-DWT and PCA improved the classification result in predicting the conversion to AD.

It is worth to note that all principal components obtained from PCA was used in this paper. Feature selection was not included in this model to avoid the high sensitivity issue, due to different number of features gives high impact on the classification result. Nevertheless, it is necessary to examine the scheme of AD classification to further improve the results. The additional data such as age, gender and neuropsychological test can be included to boost the classification result. Without surprise, the prediction of the conversion to AD is more challenging on the datasets collected in the earlier stage of the disease compared to the datasets collected at time point which are near to stable diagnosis. Therefore, more attention shall be focused on the early diagnosis of the disease to provide better treatment to the patients. The future work can be done on designing the AD classification scheme to enlarge the difference between the classification group, which shall give a more distinguishable boundary for the classifier.

Acknowledgements. The authors would like to thank Universiti Teknologi Malaysia (UTM) for supporting this research. The authors also acknowledge that the data used in this study was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

The data was funded by ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12–2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-la Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the Study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

1. Alzheimer's Association: 2020 Alzheimer's disease facts and figures. Alzheimer's & Dementia **16**(3), 391–460 (2020)
2. Dubois, B., et al.: Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. **6**(8), 734–736 (2007)
3. Soucy, J.-P., et al.: Clinical applications of neuroimaging in patients with Alzheimer's disease: a review from the fourth Canadian consensus conference on the diagnosis and treatment of demantia. Alzheimer's Res. Therapy **5**(1), 1 (2013)
4. Ledig, C., Schuh, A., Guerrero, R., Heckemann, R.A., Rueckert, D.: Structural brain imaging in Alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. Sci. Rep. **8**(1), 1–6 (2018)
5. Herrera, L.J., Rojas, I., Pomares, H., Guillén, A., Valenzuela, O., Baños, O.: Classification of MRI images for Alzheimer's disease detection. In: 2013 International Conference on Social Computing, pp. 846–851 (2013)
6. Altaf, T., Anwar, S.M., Gul, N., Majeed, M.N., Majid, M.: Multi-class Alzheimer's disease classification using image and clinical features. Biomed. Signal. Process. Control **43**, 64–74 (2018)
7. Raut, A., Dalal, V.: A machine learning based approach for detection of Alzheimer's disease using analysis of hippocampus region from MRI Scan. In: IEEE International Conference on Computing Methodologies and Communication, pp. 236–242 (2017)
8. Dolph, C.V., Alam, M., Shboul, Z., Samad, M.D., Iftekharuddin, K.M.: Deep learning of texture and structural features for multiclass Alzheimer's disease classification. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 2259–2266. IEEE (2017)
9. Margarida Matos A., Faria P., Patricio M.: Voxel-based morphometry analyses in Alzheimer's disease. In: 2013 IEEE 3rd Portuguese Meeting in Bioengineering (ENBENG), pp. 1–4. IEEE (2013)
10. Tondelli, M., Wilcock, G.K., Nichelli, P., De Jager, C.A., Jenkinson, M., Zamboni, G.: Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. Neurobiol. Aging **33**(4), 825-e25 (2012)

11. Beheshti, I., Demirel, H.: Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease. Comput. Biol. Med. **64**, 208–216 (2015)
12. Wang, W.-Y., et al.: Voxel-based meta-analysis of grey matter changes in Alzheimer's disease. Transl. Neurodegener. **4**(1), 1–9 (2015)
13. Salvatore, C., Cerasa, A., Castiglioni, I.: MRI Characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 months before probable diagnosis. Front. Aging. Neurosci. **10**, 135 (2018)
14. Khedher, L., Ramírez, J., Górriz, J.M., Brahim, A., Segovia, F.: Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. Neurocomputing **151**, 139–150 (2015)
15. Zhang, Y., Wang, S., Phillips, P., Dong, Z., Ji, G., Yang, J.: Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC. Biomed. Signal Proces. Control **21**, 58–73 (2015)
16. Jongkreangkrai, C., Vichianin, Y., Tocharoenchai, C., Arimura, H.: Computer-aided classification of Alzheimer's disease based on support vector machine with combination of cerebral image features in MRI. J. Phys. Conf. Ser. **694**, 012036 (2016)
17. Fulton, V.L., Dolezel, D., Harrop, J., Yan, Y., Fulton, C.P.: Classification of Alzheimer's Disease with and without Imagery using gradient boosted machines and ResNet-50. Brain Sci. **9**(9), 212 (2019)
18. Munteanu, C.R., et al.: Classification of mild cognitive impairment and Alzheimer's disease with machine-learning techniques using 1H magnetic resonance spectroscopy data. Expert. Syst. App. **42**(15–16), 6205–6214 (2015)
19. Ebrahimighahnavieh, M.A., Luo, S., Chiong, R.: Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. Comput. Methods Programs. Biomed. **187**, 105242 (2020)
20. Ejaz, K., et al.: Segmentation method for pathological brain tumor and accurate detection using MRI. Int. J. Adv. Comput. Sci. App. **9**(8), 394–401 (2018)
21. Moler, C.B.: Eigenvalues and singular values. In: Numerical Computing with Matlab, pp. 269–305. Society for Industrial and Applied Mathematics (2004)
22. Jovicich, J., et al.: Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. NeuroImage **30**(2), 436–443 (2006)
23. Jack, C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reason. Imaging. **27**(4), 685–691 (2008)
24. Baratloo, A., Hosseini, M., Negida, A., El Ashal, G.: Part 1: simple definition and calculation of accuracy sensitivity and specificity. Emergency **3**(2), 48–49 (2015)