International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Dimensionality Reduction Methods for Alzheimer's Disease Classification

Aow Yong Li Yew[1], Mohd Shafry Mohd Rahim[1]

[1] *School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia*

E-mail address: ylyaow2@live.utm.my, shafry@utm.my

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Alzheimer's Disease (AD) is the most common type of dementia in the world. There is no cure for the disease. The early detection of the disease is the goal for better AD treatment. Computer-aided diagnosis serves as a supportive tool in AD diagnosis, which classifies the stages of AD from the three-dimensional (3D) brain images. In image processing, a 3D image will result in millions features. Therefore, apart from extracting significant features for AD classification, the feature extraction step also involves reducing the dimensions of the data. This paper aims to investigate the suitability of the current dimensionality reduction methods on AD classification. In addition, this paper also examines the impact of various intrinsic dimension estimation techniques on the dimensionality reduction techniques. The contribution of this paper is to conduct the comparative study with same dataset, which allows a comparison of the strengths of existing methods on AD classification. A total of 200 subjects with T1-weighted images were obtained at different time points from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The performance measurement of this paper is classification accuracy. The best approach among the methods discussed in this paper is the combination of discrete wavelet transform and principal component analysis, and it achieved 87% accuracy in average for the dataset collected at different time points. It reveals that the current techniques have the strength in extracting significant features for AD classification. However, the classification performance is influenced by the intrinsic dimension on dimensionality reduction. Therefore, the rationale and suggestions for improvement of the methods are discussed.

Keywords: Alzheimer's Disease, Dimensionality reduction, Intrinsic dimension estimation, Feature extraction

1. INTRODUCTION

According to the World Alzheimer Report 2015, it reported that there were 46.8 million people suffered for dementia in 2015 and the number of patients is expected to increase almost double every 20 years in the worldwide [1]. AD is the most common cause of dementia which even leads to death due to complications. The patients undergo brain changes which results in memory loss, change of behaviors and language problem. It influences the patient, the family members and the caregivers in heath, work, finance, and social life. Therefore, the early detection of AD can help in improving their quality of life.

Magnetic Resonance Imaging (MRI) is one of the brain imaging techniques commonly used in AD diagnosis. It produces three-dimensional images for the brain which causes millions of features in image processing. Therefore, AD classification involves two main processes, which are feature extraction and classification. Feature extraction is the process in extracting the significant features for classification while reducing the dimensions of the data. For this reason, the technique used is also called dimensionality reduction technique. Dimensionality reduction is very important to prevent overfitting and reduce the computation cost. Classification is a process to categorize the group of the disease. Machine learning methods are commonly applied for AD classification, especially Support Vector Machine (SVM) and Neural Network nowadays. On the other hand, there are different kinds of dimensionality reduction methods have been applied.

The dimensionality reduction methods can be grouped into linear approach and non-linear approach. Principal component analysis (PCA) and partial least square (PLS) are widely used linear transformation approaches [2]–[4]. The result in [3] supported PLS performed better compared to PCA in extracting discriminative

E-mail: ylyaow2@live.utm.my

information. The authors in [5], [6] applied discrete wavelet transform (DWT) and PCA to extract and reduce the features. The authors in [6] claimed that the performance of DWT alone was better than the combination of DWT and PCA. H.Park [7] applied a non-linear approach, called ISOMAP to quantify the shape information through embedding the data to lower dimensions according to the distance measure. The author stated that Isomap may perform well in time series analysis. Xin Liu, et al. [8] implemented local linear embedding (LLE) for AD classification to preserve the local properties of the volume and cortical thickness, at the same time, it transformed the data to lower dimensional space. The authors mentioned that LLE performed well with different classifier.

Besides this, the author in [9] employed a deep learning approach called autoencoder. The author claimed that autoencoder outperforms PCA for non-linear data. Autoencoder uses artificial neural network architecture with a bottleneck layer to force the model to reduce the dimension or compress the data. The stacked autoencoders was applied in the multiclass AD classification [10]. The top layers of stacked autoencoders are unsupervised feature extraction and the target layer is supervised classification. Nevertheless, computational time is the concern of applying stacked autoencoder [11]. Apart from computational time, memory bottleneck is also the issue in deep learning. To tackle the problem, we suggest a two-tier autoencoders dimensionality reduction. The difference of traditional autoencoders with the proposed technique is mentioned in methodology section.

Intrinsic dimension refers to the minimum number of dimensions needed to interpret the data in low dimensional space. It is crucial to decide the intrinsic dimension to be transformed. However, most of the research studies in AD classification did not raise this issue. The intrinsic dimension estimation methods can be served as a problem solver in this case. It can be grouped into three classes, which are projection approach, probabilistic approach and geometric approach [12]. The typical methods for each category are eigenvalue-based estimation (EigValue), maximum likelihood estimator (MLE) and correlation dimension estimator (CorrDim) [13]–[15]. This paper compared aforementioned methods and the more recent method which is dimensionality from angle and norm concentration (DANCo) [16].

The motivation of this paper is to identify the suitable dimensionality reduction method on magnetic resonance images that can improve the accuracy of AD classification. Therefore, the contribution of this paper is to compare different dimensionality reduction techniques by using the same study population. In previous studies, most of the researchers did not apply the methods on the same study population, which makes the benchmarking of the methods is impossible. This paper enhances the understanding towards the strength of existing dimensionality reduction methods on AD classification. Besides, this paper also contributes to explore the use of intrinsic dimension estimation with different dimensionality reduction methods. We reveal the impact of the intrinsic dimensions computed from intrinsic dimension estimation methods towards the classification results. This paper is organized as follows: Section 2 describes the methodology of this paper, Section 3 explains the implementation and evaluation, Section 4 presents and discusses the results, and Section 5 draws a conclusion for this study.

2. METHODOLOGY

The evaluation of the intrinsic estimation methods and dimensionality reduction methods were conducted as Fig. 1. We have grouped the steps into four main processes: data collection and preparation, intrinsic dimension estimation, dimensionality reduction and classification. The details of the processes are discussed in the following sub-sections.

A. Data collection and preparation

The data collection was based on the information provided by Salvatore, et al [4]. The data were obtained from 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

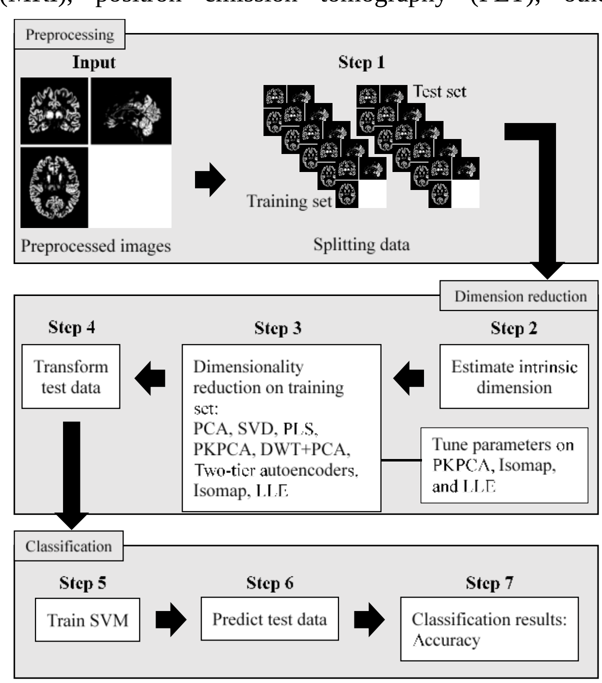

Figure 2. The framework of comparative study on dimensionality reduction approaches

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 data were collected across four time points to allow the prediction on the conversion to AD in future research studies. The brain images of each subject were obtained at the time point 24 months before stable diagnosis, 18 months before stable diagnosis, 12 months before stable diagnosis and during stable diagnosis time point. The data consisted of 50 healthy control (CN), 50 stable mild cognitive impairment (SMCI), 50 progressive mild cognitive impairment (PMCI) and 50 AD. The subjects were grouped into two classes, which were AD with PMCI and CN with SMCI in this paper.

The images collected from ADNI dataset were preprocessed T1-weighted structural MR images. The images had undergone 3D gradient inhomogeneity correction [17] to correct the gradient nonlinearity and B1 non-uniformity correction to correct the non-uniformity of intensity [18]. The images were downloaded in 3D Neuroimaging Informatics Technology Initiative (NIfTI) format in a single file. Furthermore, the images were applied with several preprocessing steps using Computational Anatomy Toolbox (CAT12) with the default parameters. The preprocessing steps included image alignment, skull-stripping and spatial normalization. After the normalization, the images had been transformed into the size of 121 x 145 x 121 voxels. The data were segmented to gray matter (GM) and white matter (WM). This paper used GM as the features for AD classification. After the preprocessing, the data were partitioned into two subsets due to the supervised learning were implemented in this paper. It is a must to implement data splitting so that it could prevent the test data being exposed to the trained model while allowing the assessment on model accuracy. In this paper, the data were divided equally for training set and test set.

B. Intrinsic dimension estimation

The process of dimensionality reduction started with computing the intrinsic dimension. The training set was the input for every intrinsic dimension estimation approach. MLE maximizes the likelihood to the distances between close neighbors. The distances between close neighbors are computed through k nearest neighbor (KNN). It is crucial to decide k value because a smaller k value may focus on the noisy data, and it gives a high impact towards the result. On the other hand, larger k value may cause bias and increase computation time. One of the suggestions on k value is k equals to the square root of the training set size [19]. In addition, by setting the k values in a range rather than a single k will help to tackle the noise and bias issue. Therefore, the k values used in this paper were in the range of 6 to 14, where the square root of the size of training set was 10.

EigValue method determines the intrinsic dimension by comparing the eigenvalues with the threshold. The threshold was set to 2.5% in this paper, which required the eigenvalues that were more than 97.5% of the total variance of the data. The value of the threshold should not be too high because the purpose for dimensionality reduction is extracting significant features while minimizing the loss of information. CorrDim finds the distance of the nearest neighborhood through KNN. By using the median value and maximum value of KNN, CorrDim estimates the intrinsic dimension by calculating the slope of the curve.

DANCo is specified design for the high dimensionality data through considering the concentration of distance and the angle effects in high dimensionality data. The concentration and angle effects are calculated through identifying the value of the set of neighbors. DANCo applies KNN to find the nearest neighbors, the k value was set to 10 in this paper. The rest of the parameters were used according to the original paper [16].

C. Dimensionality reduction

The intrinsic dimensions obtained from the previous section were used as the parameter for each dimensionality reduction technique. PCA decomposes the data to uncorrelated data, which is called principal components. The first principal component has the highest variance, and it presents most of the data. The decomposition process is done through eigen decomposition or singular values decomposition (SVD) [20]. SVD is proved that it is numerically stable for the matrix [21]. Therefore, this paper included this method in the comparison to compare the ability in extracting significant AD features, named SVD-PCA. PLS works similar to PCA but PLS maximizes the covariance between factors of data and response variables [22].

PCA is a linear transformation method but the real-world data is not all distributed linearly. Therefore, Kernel PCA is introduced to handle non-linear data [23]. Polynomial kernel PCA (PKPCA) and Gaussian kernel PCA were tested in this paper. However, the classification results of Gaussian kernel PCA showed that the classifier grouped all test set into one class. Hence, this paper has excluded Gaussian kernel PCA in the comparison. Instead of using covariance matrix as input, Kernel PCA computes eigen decomposition on the kernel matrix [24]. Kernel PCA transforms the data to a higher dimensional feature space, to make the data distributes linearly.

Autoencoder is also treated as a non-linear PCA approach. It utilizes a bottleneck layer to achieve dimensionality reduction [9]. The bottleneck layer contains smaller number of nodes than both input and output layer. Autoencoder is an unsupervised learning, which consists of encoder and decoder. The encoder of the trained autoencoder will generate the compressed features while the decoder reconstructs the data for examining the quality of the compression. Two or more autoencoders can be applied to further reduce the dimension. The output of first autoencoders are directly fed into second autoencoders. In this paper, we have adopted a two-tier autoencoders. The traditional autoencoders train all the features at the same time but this is impossible for those datasets that have a large number of features. This study contains of 2122945 features, it requires lots of computational resources to compute the neural network. Therefore, we have divided the computation into two levels. The first autoencoder compresses each slice of the sample to a low dimensional representation. The significant features of each slice are captured through first autoencoder. Then, all the low dimensional slices for each sample are concatenated and become a feature vector. The feature vectors of all the samples are the input of the second autoencoder. The second autoencoder captures the global properties and to further reduce the dimension. The size of the hidden layer for first autoencoder used in this paper was the number of slices for each sample and the size of the hidden layer for second autoencoder was based on the intrinsic dimension.

Furthermore, this paper also explored another two non-linear transformation approaches, which were Isomap and LLE. Isomap considers the distribution of neighboring datapoints through the incorporation of geodesic distances between the datapoints [25]. The eigen decomposition is applied on the geodesic distance matrix. LLE works similarly to Isomap but LLE preserves the local properties within local neighborhoods. The contribution of each neighbors of the datapoint determines the reconstruction of the data [26]. Most of the non-linear dimensionality reduction techniques require parameter tuning. Therefore, we implemented 10-fold cross validation for the non-linear dimensionality techniques. The optimal parameter is the value yields the lowest error rate.

The 2-dimensional discrete wavelet transform (2D-DWT) decomposes the data into different frequency bands which results in a smaller size of data [6]. Therefore, it can be considered as a data reduction tool indirectly. In this paper, we employed Haar wavelet up to level 3 on each slice of the sample instead of the whole training set to solve the memory issue. It has extracted 36784 features for each sample. Due to the large amounts of features, we further performed PCA on the extracted data. Besides this, we also applied 2D-DWT with Mutual Information (MI) because of the results in [6] showed that the combination of 2D-DWT+MI obtained higher results than 2D-DWT+PCA.

To put it simply, the transformation of test set is done through multiplying the test set datapoints with the eigenvectors for most of the methods. Eigenvectors in this context are the directions of the linear transformation apply to the training set datapoints. PCA, SVD-PCA and PLS are linear dimensionality reduction approaches, they transform the test set as simple as the description above. However, Kernel PCA, Isomap and LLE require extra steps to compute the non-linear transformation. PKPCA computes the gram matrix based on the polynomial kernel function before applying the multiplication of gram matrix with eigenvectors. On the other hand, Isomap and LLE require the measurement of the distance between test set datapoints and the training set datapoints. The test set datapoints are arranged according to the distance with training set datapoints in ascending order for LLE. After that, LLE computes the gram matrix and calculates the contribution of the local gram matrix. The embedding of test set is based on the contribution of the datapoints. The encoder from trained autoencoder generates the compressed features for test set while 2D-DWT is applied to test set as training set.

D. *Classification*

Linear support vector machine (SVM) was employed as the classifier in this paper. SVM is a supervised classifier which builds a trained model that maximizes the hyperplane of two diagnostic groups to predict the new data [27]. The low dimensional training set is the input for the trained model while the low dimensional test set is used for the prediction.

3. IMPLEMENTATION AND EVALUATION

A. *K-fold cross validation*

Based on the procedures described above, the non-linear dimensionality reduction methods applied 10-fold cross validation to ensure the selected parameter transforms the data to the features that have lower prediction error to the target label. Before the implementation of cross-validation, the neighborhoods' distance for LLE and Isomap was set from 6 to 30, and the degree of PKPCA was set from 1 to 10 to examine the impact of different values towards the data. Then, the 10-fold cross validation divides the training set to 10 subsets. Nine of the subsets were used as training set and the remaining subset was treated as validation set in parameter tuning. The value that obtained the lowest error rate in cross-validation was selected as the best parameter value for each technique.

B. *Performance mesasurement*

The performance measurement of the classification was accuracy. The accuracy tells us the ability of the model to differentiate two diagnostic groups correctly [28]. The accuracy is derived from confusion matrix through calculating true positive (TP), true negative (TN), false positive (FP) and false negative (FN) as shown in Table I. The equation for calculating the accuracy is given as in (1).

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN) \quad (1)$$

TABLE V. CONFUSION MATRIX FOR CLASSIFICATION

		Predicted Class	
		AD+PMCI	CN+SMCI
Actual Class	AD+PMCI	True Positive	False Negative
	CN+SMCI	False Positive	True Negative

4. RESULTS AND DISCUSSION

A. *Intrinsic dimensions based on different intrinsic dimension estimation approaches*

From Table II, we notice that the estimated intrinsic dimensions were very different based on different approaches. MLE considered more dimensions needed compared to other methods at all time points, which was at least 30 intrinsic dimensions. The second highest intrinsic dimension were computed by using DANCo. However, the intrinsic dimensions were significantly lower compared to MLE. CorrDim obtained the lowest intrinsic dimension, and the results were the same for all time points.

B. *Analysis of intrinsic dimension estimation approaches*

In this section, PCA, SVD-PCA and PKPCA obtained the same result with the estimated intrinsic dimension. Therefore, the results are tabulated in Table III named as PCA-based approaches. From the result in Table III, it did not indicate the most suitable intrinsic dimension estimation technique for PCA-based approaches since all the intrinsic dimension estimation methods were not able to achieve the highest result at all time points. However, the average accuracies of the techniques at different time points gave the ranking as following: MLE, EigValue, DANCo and CorrDim. Contrary to PCA-based approaches, PLS achieved highest results at all time point by using CorrDim as shown in Table IV.

Table V reveals that the selection of intrinsic dimension had a great impact on the performance of Isomap. The range of the accuracies obtained by Isomap with different intrinsic dimension estimation approaches was big. It was 18% difference at the time point 24 months before stable diagnosis. The ranking of the intrinsic dimension estimation methods in the average accuracies at different time points was as following: DANCo, EigValue, CorrDim and MLE. Table VI summarizes the performance of LLE. It reports that LLE obtained a smaller difference in the range of accuracy compared to Isomap, which was 12% difference at time point of 18 months before stable diagnosis. The ranking of the average accuracies at different time points was DANCo, EigValue, CorrDim and MLE.

DWT+PCA obtained similar ranking with PCA on the combination with intrinsic dimension estimation methods as shown in Table VII. This is because DWT+PCA inherits the property of PCA. MLE and EigValue obtained the same result in the average accuracies at different time points. It was followed by DANCo and CorrDim. On the other hand, the ranking of the intrinsic dimension estimation methods on DWT+MI was DANCo, MLE, EigValue and CorrDim as illustrated in Table VIII. At last, there was no indication on the suitable intrinsic dimension estimation technique for two-tier autoencoders

TABLE I. THE INTRINSIC DIMENSIONS BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION APPROACHES

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	31	38	33	30
EigValue	8	5	6	7
CorrDim	2	2	2	2
DANCo	10	10	10	10

TABLE II. THE ACCURACIES OF PCA-BASED APPROACHES BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION METHODS

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	0.79	0.81	0.83	0.87
EigValue	0.79	0.81	0.84	0.84
CorrDim	0.78	0.83	0.84	0.82
DANCo	0.80	0.82	0.83	0.82

TABLE III. THE ACCURACIES OF PARTIAL LEAST SQUARE BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION METHODS

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	0.79	0.84	0.83	0.85
EigValue	0.79	0.84	0.83	0.85
CorrDim	0.81	0.84	0.86	0.85
DANCo	0.79	0.84	0.83	0.85

TABLE IV. THE ACCURACIES OF ISOMAP BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION METHODS

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	0.69	0.64	0.70	0.68
EigValue	0.59	0.80	0.77	0.79
CorrDim	0.72	0.68	0.70	0.75
DANCo	0.77	0.73	0.81	0.77

TABLE VI. THE ACCURACIES OF LOCAL LINEAR EMBEDDING BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION METHODS

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	0.76	0.71	0.76	0.80
EigValue	0.73	0.69	0.78	0.85
CorrDim	0.70	0.79	0.74	0.81
DANCo	0.76	0.81	0.79	0.82

TABLE VII. THE ACCURACIES OF DWT+PCA BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION METHODS

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	0.78	0.81	0.83	0.87
EigValue	0.79	0.80	0.86	0.84
CorrDim	0.75	0.77	0.83	0.80
DANCo	0.77	0.81	0.83	0.82

TABLE VIII. THE ACCURACIES OF DWT+MI BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION METHODS

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	0.63	0.79	0.78	0.74
EigValue	0.66	0.74	0.72	0.77
CorrDim	0.66	0.76	0.78	0.72
DANCo	0.65	0.80	0.74	0.76

TABLE IX. THE ACCURACIES OF TWO-TIER AUTOENCODER BASED ON DIFFERENT INTRINSIC DIMENSION ESTIMATION METHODS

Methods	24 m before stable diagnosis	18 m before stable diagnosis	12 m before stable diagnosis	Stable diagnosis time point
MLE	0.64	0.77	0.78	0.83
EigValue	0.70	0.70	0.68	0.75
CorrDim	0.71	0.73	0.67	0.84
DANCo	0.69	0.72	0.74	0.79

from Table IX. MLE obtained the highest and lowest results in different time points. The ranking of the techniques in average accuracies at different time points was as following: MLE, CorrDim, DANCo and EigValue.

From the performance of different dimensionality reduction approaches, we can conclude that it is not all intrinsic dimension estimation approaches suit all the dimensionality reduction approaches. Different dimensionality reduction approaches require different intrinsic dimensions. The overestimation and underestimation of intrinsic dimension always occur in the combination of the current techniques. The statement is supported from a research which mentioned that the underestimation occurs when the required intrinsic dimension is more than 10 [29]. In addition, the threshold of EigValue was set manually in this paper. It is arbitrary and might cause a big loss of data unknowingly. CorrDim will underestimate the intrinsic dimension for the data when the data is non-uniform distribution. CorrDim treats the significance of every point as identical [12]. DANCo suffers the same issue as CorrDim, it considers the neighborhood of each point or manifold is distributed uniformly [16].

In contrast, PLS looked working well with CorrDim, it achieved the highest result with CorrDim at all time points. The reason may not be because of CorrDim is good enough, but the nature of PLS. From the results of PLS, it proved that PLS obtains more stable results compared to other approaches. By adding the number of dimensions, it will not improve the accuracy of the classification. Therefore, the development of intrinsic dimension estimation approach shall consider the nature of each dimensionality reduction technique. Different dimensionality reduction approaches have different criteria to reduce and transform the data. It might also be the reason for underestimating or overestimating the intrinsic dimension of the current methods. Besides this reason, the author in [30] stated that the intrinsic dimension estimation approaches face the problem of curse of dimensionality. Therefore, we suggest that applying the intrinsic dimension estimation approach after conducting the dimensionality reduction approach with the maximum dimensions.

C. Analysis of dimensionality reduction approaches over dimensions

The previous section has analyzed the classification results based on the intrinsic dimensions, but it is necessary to get into bottom to see the impact of the chosen dimension numbers on classification results based on the dimensionality reduction methods. The maximum dimension number is (n-1), where n is number of training data. Hence, this section reports the classification results of the dimensionality reduction methods from 1 to 99 dimensions. The exception is the method of DWT+MI. The number of dimensions of DWT+MI was 36784 after decomposition. It is not valid to compare with the rest of the results in the same figure. Therefore, it is not included in the figures below, but the results are described in paragraph form. Fig. 2 until Fig. 5 reports the classification accuracies for the dataset collected at different time points. PCA and SVD-PCA obtained the same result over the dimensions. Therefore, we use the term "PCA" to represent both techniques in this section.

Fig. 2 demonstrates the accuracies of the dataset collected at time point of 24 months before stable diagnosis. DWT+PCA achieved 85% as the highest result across the dimensions. PCA and LLE achieved the same result, which was 84% in their highest results. It was followed by PKPCA, two-tier autoencoders DWT+MI, Isomap and PLS. Although PLS achieved the lowest accuracy, but the results over dimensions were more stable compared to the other methods. There was a huge difference for the results obtained by two-tier autoencoders, DWT+PCA, PCA, PKPCA, DWT+MI, Isomap and LLE over dimensions. This phenomenon also happened for the dataset collected at other time points.

Fig. 3 presents the accuracies of dataset collected at the time point 18 months before stable diagnosis. The ranking of the methods based on the highest results was two-tier autoencoders, DWT+PCA, PCA, PKPCA, DWT+MI, PLS, Isomap and LLE, where PCA, PKPCA and DWT+MI obtained the same result. Fig. 4 indicates the accuracies of dataset collected at time point of 12 months before stable diagnosis. DWT+PCA achieved 88% as highest accuracy across the dimensions. DWT+MI, PCA and PKPCA achieved 87% as their highest results. It was followed by PLS, Isomap, LLE and two-tier autoencoders. Fig. 5 shows the accuracies of dataset collected from stable diagnosis time point. DWT+PCA achieved 89% while DWT+MI, PCA, PKPCA, two-tier autoencoders and LLE achieved 88% as their highest results. Then, it was followed by PLS and Isomap.

From the observation, PKPCA obtained the same result with PCA with less numbers of dimensions. This is because of PKPCA is functioning as a linear approach during less dimensions. The degree of polynomial chosen with k-fold cross validation was 1. When PKPCA did not perform as a linear approach, the results dropped as shown in the figures. This can be proved that the data are extra complex, therefore it turns out that the linear approach works better than non-linear approach on PCA-based approaches. When involving more dimensions, it is possible to occur overfitting in the PKPCA trained model. The model chose higher degree of polynomial to fit the training data, but it increased the generalization error.

Besides, the results obtained by non-linear approaches like two-tier autoencoders, Isomap and LLE also have greater differences compared to PCA-based approaches and PLS over the dimensions. This is possible due to the ability of the methods is relying on the parameter and hyperparameter selection. The parameter for LLE and Isomap is the distance of the neighbors, while the main hyperparameters for two-tier autoencoders are the numbers of hidden layers and the regularization value. Apart from this, autoencoders also require a suitable hidden layer size to extract significant features through self-learning process. It will cause overfitting when using too little number of hidden layers. The number of dimensions influenced the results of Isomap the most. It was followed by two-tier autoencoders, DWT+MI and LLE. The results dropped or raised significantly from one dimension to another dimension. PCA-based approaches maximize the variance of the data and transform the data to principal component. Although the principal components with a higher variance explain most of the information, the principal components with smaller variance also contribute to differentiate the classes in AD classification. On the contrary, the results of PLS were remained the same from 2 or 4 dimensions onwards. However, PLS never achieved the highest accuracies at all time point.

Overall, DWT+PCA achieved the top results among the techniques. It was followed by PCA and PKPCA. They can extract significant features for AD classification. The issue of the method is to select a suitable intrinsic dimension. DWT+MI performed slightly poorer than PCA and it required more dimensions than other techniques to achieve higher results. PLS did a great job in terms of its stability, but it is necessary to increase the discriminative power to extract more meaningful data. Two-tier autoencoders, Isomap and LLE also performed well in extracting significant features for AD classification but the methods require more concern on the parameter and hyperparameter tuning. Besides, the combination of the extracted features from different dimensions might be a way to improve the current dimensionality reduction approach.

5. CONCLUSIONS

To the best of our knowledge, this paper is the first study to compare various dimensionality reduction techniques for AD classification with the same dataset. The main finding from the comparative study is the existing dimensionality reduction methods can extract significant features for classification. The results are consistent with the previous studies on the contribution of PCA and PLS towards AD classification. Moreover, the combination of PCA with other technique also boosts the accuracy for classification. Despite this, the deep learning approach, such as autoencoder outperforms PLS. However, hyperparameters tuning should not be neglected for deep learning approach. A deep autoencoder might help in improving the classification accuracy, but the computational resource limit always is the issue for deep learning. The second finding is about the influence of intrinsic dimension for dimensionality reduction approach towards the classification result. The classification accuracies of most of the dimensionality reduction approaches have a great difference over the dimensions. The current intrinsic dimension estimation methods do not exert the ability of the dimensionality reduction approaches. Therefore, we believe that further analysis on the development of intrinsic dimension estimation technique should follow the nature of the dimensionality reduction technique closely. We found that the different criteria in building the dimensionality reduction technique will influence the numbers of dimensions needed to interpret the data. In conclusion, a further research on dimensionality reduction framework is required to improve the accuracy of AD classification.

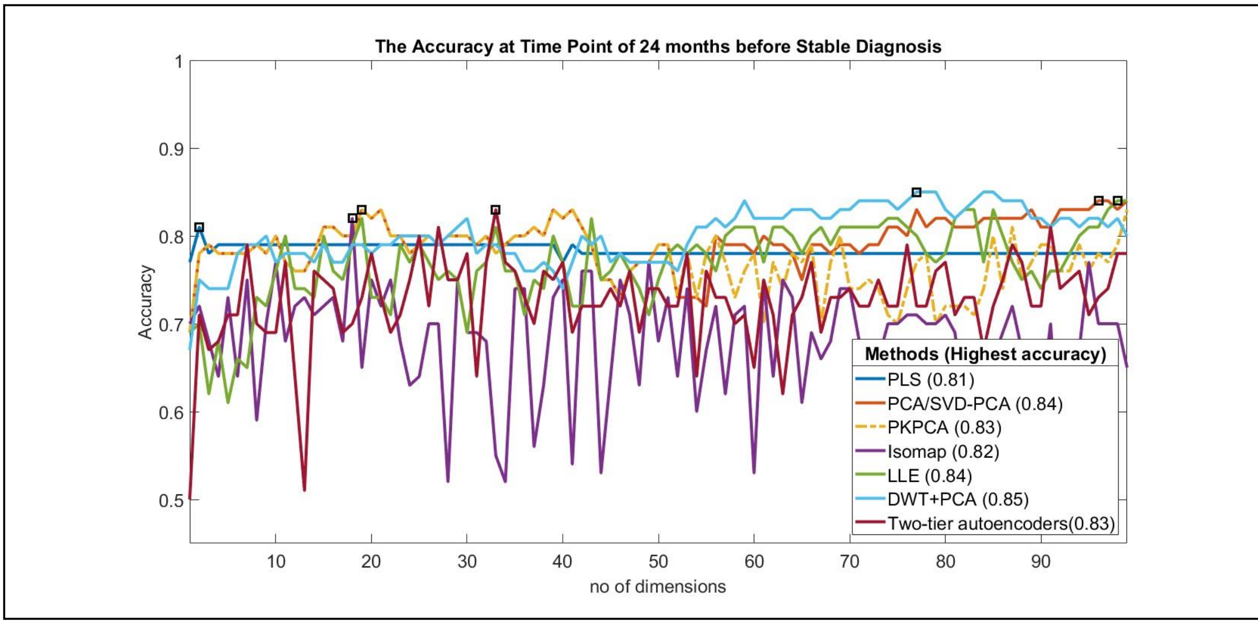

Figure 2. The accuracies based on different dimensionality reduction methods at time point of 24 months before stable diagnosis

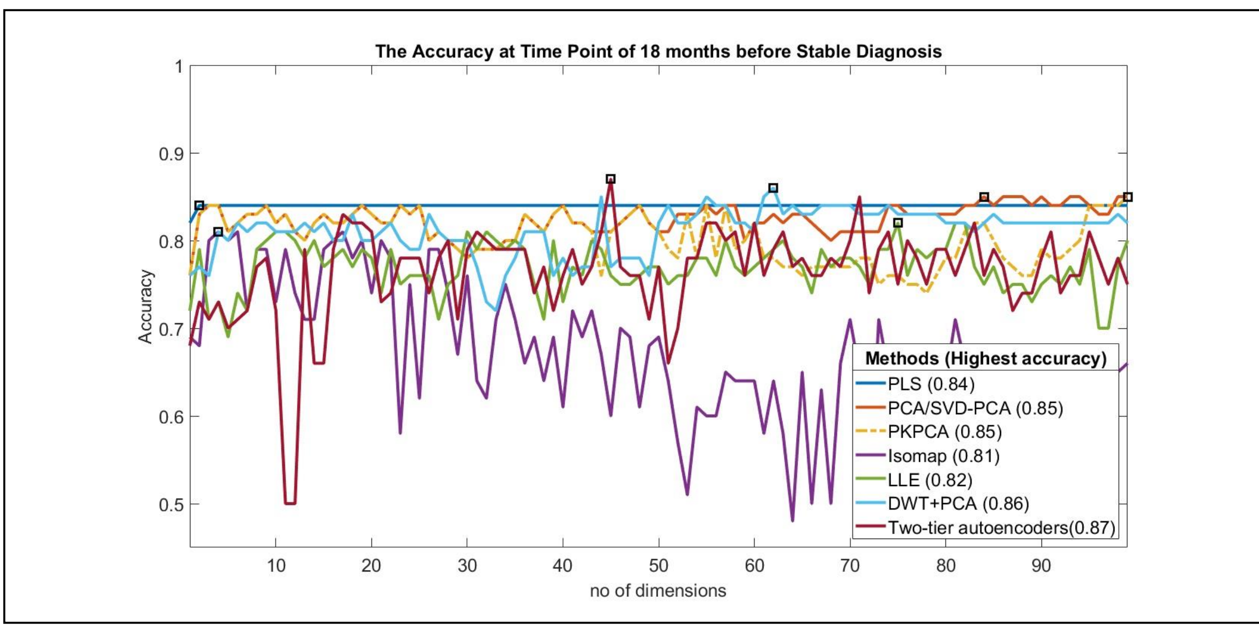

Figure 3. The accuracies based on different dimensionality reduction methods at time point of 18 months before stable diagnosis

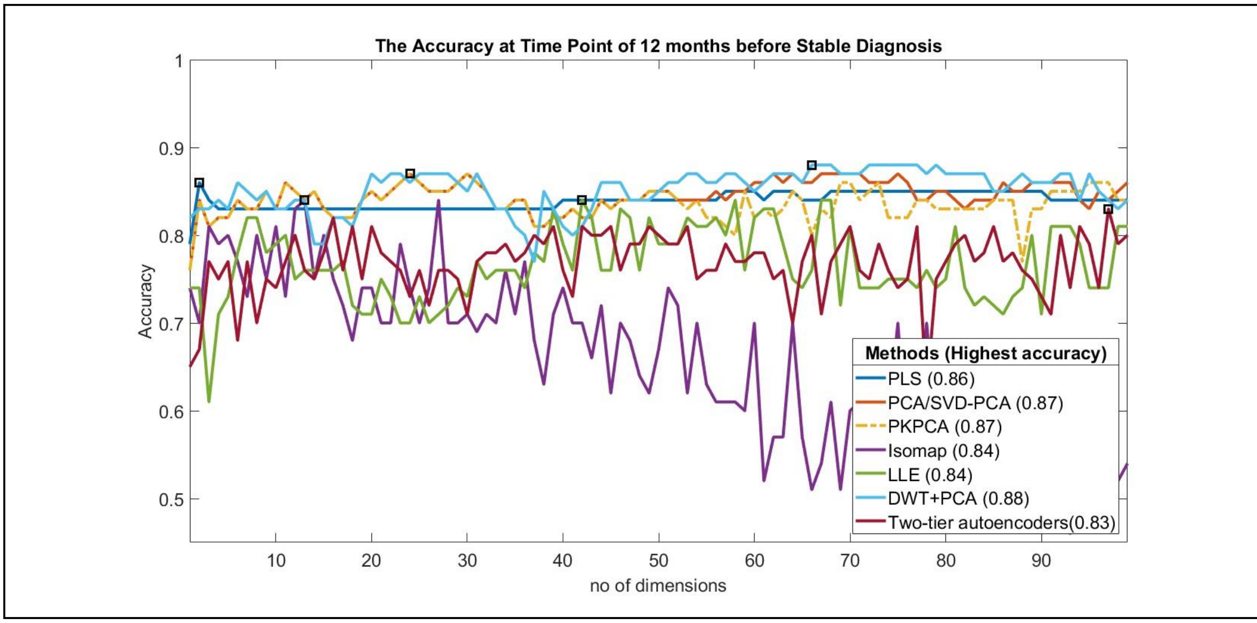

Figure 4. The accuracies based on different dimensionality reduction methods at time point of 12 months before stable diagnosis

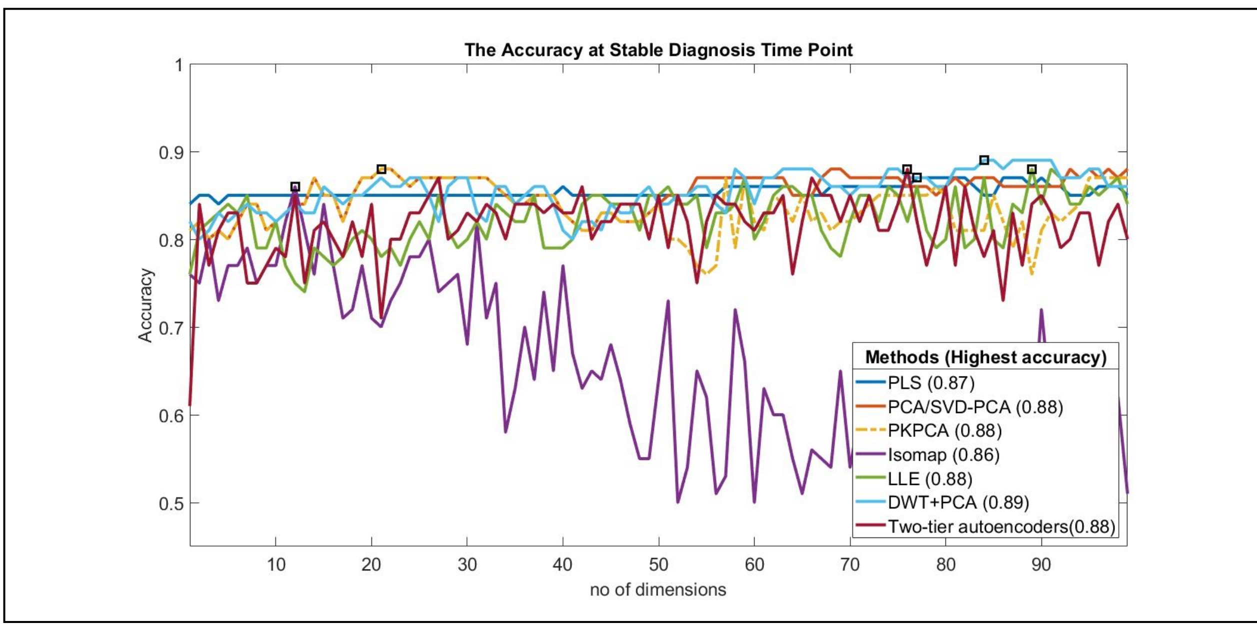

Figure 5. The accuracies based on different dimensionality reduction methods at stable diagnosis time point

ACKNOWLEDGMENT

The authors would like to thank Universiti Teknologi Malaysia (UTM) for supporting this research. The authors also acknowledge that the data used in this paper 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] M. Prince, A. Wimo, M. Guerchet, A. Gemma-Claire, Y.-T. Wu, and M. Prina, "World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends," 2015.

[2] R. K. Lama, J. Gwak, J.-S. Park, and S.-W. Lee, "Diagnosis of Alzheimer's Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features," *J. Healthc. Eng.*, vol. 2017, no. 1, pp. 1–11, 2017.

[3] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer ' s disease based on partial least squares , principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139–150, 2015.

[4] C. Salvatore, A. Cerasa, and I. Castiglioni, "MRI Characterizes the Progressive Course of AD and Predicts Conversion to Alzheimer's Dementia 24 Months Before Probable Diagnosis," *Front. Aging Neurosci.*, vol. 10, May 2018.

[5] K. Ejaz *et al.*, "Segmentation method for pathological brain tumor and accurate detection using MRI," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 394–401, 2018.

[6] L. J. Herrera, I. Rojas, H. Pomares, A. Guillén, O. Valenzuela, and O. Baños, "Classification of MRI images for Alzheimer's disease detection," *Proc. - Soc. 2013*, pp. 846–851, 2013.

[7] H. Park, "ISOMAP induced manifold embedding and its application to Alzheimer's disease and mild cognitive impairment," *Neurosci. Lett.*, vol. 513, no. 2, pp. 141–145, 2012.

[8] X. Liu, D. Tosun, M. W. Weiner, and N. Schuff, "Locally linear embedding (LLE) for MRI based Alzheimer's disease classification," *Neuroimage*, vol. 83, pp. 148–157, Dec. 2013.

[9] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE J.*, vol. 37, no. 2, pp. 233–243, 1991.

[10] C. V. Dolph, M. Alam, Z. Shboul, M. D. Samad, and K. M. Iftekharuddin, "Deep learning of texture and structural features for multiclass Alzheimer's disease classification," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, no. 1310353, pp. 2259–2266.

[11] L. Vařeka and P. Mautner, "Stacked Autoencoders for the P300 Component Detection," *Front. Neurosci.*, vol. 11, no. MAY, pp. 1–9, May 2017.

[12] Y. Liu, Z. Yu, M. Zeng, and S. Wang, "Dimension Estimation Using Weighted Correlation Dimension Method," *Discret. Dyn. Nat. Soc.*, pp. 1–10, 2015.

[13] K. Fukunaga and D. R. Olsen, "An Algorithm for Finding Intrinsic Dimensionality of Data," *IEEE Trans. Comput.*, vol. C–20, no. 2, pp. 176–183, Feb. 1971.

[14] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Phys. D Nonlinear Phenom.*, vol. 9, no. 1–2, pp. 189–208, 1983.

[15] E. Levina and P. J. Bickel, "Maximum Likelihood Estimation of Intrinsic Dimension," in *Proceedings of the 17th International Conference on Neural Information Processing Systems*, 2004, pp. 777–784.

[16] C. Ceruti, S. Bassis, A. Rozza, G. Lombardi, E. Casiraghi, and P. Campadelli, "DANCo: An intrinsic dimensionality estimator exploiting angle and norm concentration," *Pattern Recognit.*, vol. 47, no. 8, pp. 2569–2581, 2014.

[17] J. Jovicich *et al.*, "Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data," *Neuroimage*, vol. 30, no. 2, pp. 436–443, 2006.

[18] C. R. Jack *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *J. Magn. Reson. Imaging*, vol. 27, no. 4, pp. 685–691, 2008.

[19] A. B. Hassanat, M. A. Abbadi, G. A. Altarawneh, and A. A. Alhasanat, "Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach," *Int. J. Comput. Sci. Inf. Secur.*, vol. 12, no. 8, pp. 33–39, Sep. 2014.

[20] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular Value Decomposition and Principal Component Analysis," in *A Practical Approach to Microarray Data Analysis*, Boston: Kluwer Academic Publishers, 2003, pp. 91–109.

[21] V. C. Klema and A. J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," *IEEE Trans. Automat. Contr.*, vol. 25, no. 2, pp. 164–176, 1980.

[22] S. Maitra and J. Yan, "Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression," *Casualty Actuar. Soc.*, pp. 79–90, 2008.

[23] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, 1992, pp. 144–152.

[24] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[25] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science (80-.).*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[26] R. Sam T. and Lawrence K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science (80-.).*, vol. 290, no. 22, pp. 2323–2326, 2000.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[28] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal, "Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity.," *Emergency*, vol. 3, no. 2, pp. 48–49, 2015.

[29] P. Campadelli, E. Casiraghi, C. Ceruti, and A. Rozza, "Intrinsic Dimension Estimation: Relevant Techniques and a Benchmark Framework," *Math. Probl. Eng.*, vol. 2015, pp. 1–21, 2015.

[30] V. Erba, M. Gherardi, and P. Rotondo, "Intrinsic dimension estimation for locally undersampled data," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Aow Yong Li Yew received bachelor's degree in computer science from Univeresiti Teknologi Malaysia (UTM) in 2013. Presently, she is pursuing PhD in UTM. She has interested in medical image processing.

Mohd Shafry Mohd Rahim received his BSc and MSc degrees in computer science from Universiti Technologi Malaysia (UTM) in 1999 and 2002. He achieved his PhD degree in computer science from University Putra Malaysia (UPM). Currently, he is a professor in the School of Computing, Faculty of Engineering, UTM. He is a research fellow in UTM-IRDA Digital Media and Game Innovation of Excellence (MaGICX). He has interested in computer graphics, spatial modelling, and image processing.

E-mail: ylyaow2@live.utm.my