Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

INFANT ASPHYXIA DETECTION USING AUTOENCODERS TRAINED ON LOCALLY LINEAR EMBEDDED-REDUCED MEL FREQUENCY CEPSTRUM COEFFICIENT (MFCC) FEATURES

I. M. Yassin[1,*], A. Zabidi[1], N. Ismail[1], F. H. K. Zaman[1], M. F. Shafie[1] and Z. I. Rizman[2]

[1]Faculty of Electrical Engineering, UniversitiTeknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[2]Faculty of Electrical Engineering, UniversitiTeknologi MARA, 23000 Dungun, Terengganu Malaysia

Published online: 10 September 2017

ABSTRACT

Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learned and solved. The objective of this paper is to explore one of the Deep Learning paradigms called autoencoders to perform diagnosis of infant asphyxia. LLE was used to reduce size of feature representation while enhancing them. Two stacked autoencoders were then trained to extract the necessary features for clasification. Extensive tests performed showed that the best-performing autoencoder network could produce 92.82% classification accuracy.

Keywords: autoencoder; locally linear embedding (LLE); Mel frequency cepstral coefficient (MFCC); pattern classification; deep learning.

Author Correspondence, e-mail: ihsan.yassin@gmail.com

doi: http://dx.doi.org/10.4314/jfas.v9i3s.56

1. INTRODUCTION

1.1. Background

Infant cries carry useful information regarding its health and physical status [1]. Many research has been done to determine the pathologic conditions related to infant cries, among them in diagnosing asphyxia [2-7]. Newborn asphyxia is defined as respiratory failure in infants, and is caused by insufficient oxygen intake before, during or after birth. This causes a medical condition called hypoxia (insufficient oxygen supply to important organs and tissues) leading to damaged organs or death if improperly treated [1].

The use of neural network-based techniques for successful diagnosis of infant diseases has been widely reported in literature [8-11]. However, most of these methods rely on the use of fourth generation neural networks to perform classification. The fifth-generation of neural networks has been recently introduced with the explosion of technology in the fields of Artificial Intelligence (AI) and computation hardware. Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learned and solved [12].

The objective of this paper is to explore one of the Deep Learning paradigms called autoencoders to perform diagnosis of infant asphyxia. Autoencoders are unsupervised neural networks trained using the backpropagation algorithm that has been extensively used for feature extraction /reduction [13]. Unlike other types of neural networks that have different inputs and targets, autoencoders train themselves on targets that are equal to the inputs [14]. Although this process may seem trivial, by constraining the network into only several hidden units, the output will not be exactly like the inputs but rather an approximation of the inputs in terms of its structure and interesting features. This is because under the abovementioned constraints, the autoencoder network is forced to learn a compressed version of the inputs by considering the intercorrelations between the input data [15], thus producing some low dimensional features like that of Principal Component Analysis (PCA). Several autoencoders

can be nested to learn deeper features [14].

1.2. Recent Relevant Works

Autoencoders have been extensively used in various applications. This section gives a summary of recent works in the area.

In [16] developed a novel spatio-temporal autoencoder for unsupervised training for motion prediction in videos. The spatio-temporal autoencoder was based on the classic image autoencoder combined with a temporal autoencoder constructed using convolutional long short-term memory (LSTM) that allows richer temporal features to be extracted from video frames. The proposed system was tested on several synthetic and supervised datasets, with the ability to segment various objects automatically.

In the area of image processing, reference [17] used a combination of deep variationalautoencoders and a mixture density network to estimate the saturation and shading field of input images. The estimated saturation and shading field were then used to enhance the lighting and shading of the original input image. The method used two deep variationalautoencoders to embed the conditioning image and generated image into a cascade of low-dimensional latent variables and the mixture density network was used to fit a multimodal Gaussian Mixture Model (GMM) between the latent variables. The autoencoders and mixture density network were trained together allowing them to adapt and complementary benefit each other. Tests in the MS-COCO image database proved that the method could correct severe image saturation and shading deficiencies and produce a more realistic color distribution.

Works by [18] used autoencoders to discover three-dimensional shapes by aggregating learned features extracted from two-dimensional images. The Deep Belief Network (DBN) autoencoder is an unsupervised network that consists of many hidden units (latent variables) which are exclusively connected between layers. The reconstruction of the three-dimensional mode was done by implementing a cascade of Restricted Boltzmann Machines. The method was tested on two benchmark datasets demonstrating that the proposed method was more efficient at three-dimensional shape retrieval than other global features based methods. In another similar work [19], a stacked local convolutional autoencoder was used for

three-dimensional object retrieval. Experiments were performed on three benchmark databases, demonstrating advantage of the model over several recent and advanced methods.

In [20], the Variational Fair Autoencoder (VFAE), an extension of the variationalautoencoder was used as feature extractors. The proposed VFAE was very suited for learning representations that are explicitly invariant while retaining as much information as possible. Further, a regularizer called Maximum Mean Discrepancy was also introduced to help the discovery of invariant features. Experiments performed suggested that the proposed method was effective in removing unwanted variation sources while maintaining information-carrying latent data representations.

Another LSTM autoencoder was used for natural language generation application [21]. In this work, the hierarchical autoencoder was trained to construct an embedding for a paragraph from embeddings of sentences and words. It then reconstructs the original paragraph by decoding the embedding. Test using standard metrics (ROUGE and Entity Grid) proved that the autoencoder could preserve the syntactic, semantic and discourse coherence integrity in its generated output.

In [22], a recurrent connectionist autoencoder called TRACX2 was used to model the brain behavior of newborn infants in extracting structure from their senses. The autoencoder learns to extract the structure by constructing data chunks and distributing them across the neural network's synaptic weights. The stored information can then be used to identify similar chunks when they reoccur in the inputs. The autoencoder managed to successfully model several tasks, namely forward and backward transitions, low salience features, part-sequences and illusory items-suggesting that a newborn's learning patterns are governed by similar audio based domain-general learning mechanisms.

2. METHODOLOGY

The interested reader is referred to [1] for details on the data collection process. The initial signal for this dataset was sampled at 11,025 Hz. The Short Time Fourier Transform (STFT) was applied to the signals to determine the fundamental frequencies for both asphyxiated and non-asphyxiated cries. One-second segmentation was then performed producing 600

segmented signals, from which 284 were normal cries while 316 were asphyxiated cries.

The Mel Frequency Cepstral Coefficient (MFCC) algorithm was used for feature extraction. To perform MFCC on the segments, they were first multiplied with a 25 ms Hamming window with 50% overlap. The energy spectrum was then calculated using the Fast Fourier Transform (FFT) method. This energy spectrum was then converted to melcepstrum by passing the spectrum through 27 triangular filter banks.

We used the Locally Linear Embedding (LLE) method for feature selection [23]. LLE is a nonlinear dimensionality reduction algorithm which has several advantages: 1) faster optimization on sparse matrices, and 2) proven record of accomplishment over many dimensionality problems. The algorithm works by first finding a set of nearest neighbors of each point in the feature space. Then, LLE computes a set of weights for each of the points that effectively describes each of the points as a linear combination of its neighbours. Finally, an eigenvector-based optimization was performed to discover low-dimensional embedding of the points while maintaining the same linear combination of its neighbors. Embedding construction depends on two parameters namely the number of neighbors (k) and maximum embedding dimension (d_{max}). In our experiments, we varied both the values of k and d_{max} to between 1 and 30 to discover the optimal parameters.

We stacked two autoencoders to learn important features from the LLE-reduced dataset before transferring the features to a softmax layer for classification. It is important to optimize the number of hidden units since they affect how the autoencoder learns and compresses important features. In our experiments, we set the number of hidden units to between 5 and 15. The softmax layer uses the cross-entropy method as its performance function. Additional parameter settings are listed in Table 1.

Table 1. Additional autoencoder parameter settings

Variable	Autoencoder 1	Autoencoder 2
Weight Regularization	0.001	0.001
Sparsity Regularization	4	4
Sparsity Proportion	0.05	0.05
Decoder activation function	purelin	purelin
Scale outputs?	false	true

3. RESULTS AND ANALYSIS

3.1. Choosing the Optimal Parameters

Using brute-force method, the total combination of parameters tested generated 31,892 solutions. We evaluated the effectiveness of each parameter combinations based on the accuracy of the training and testing sets. The top ten result is displayed in Table 2.

Table 2. Top ten best solutions of brute-force testing with corresponding

$\mathbf{k}$	$\mathbf{d_{max}}$	$\mathbf{h_{AC1}}$	$\mathbf{h_{AC2}}$	thr	$\mathbf{acc_{training}}$ (%)	$\mathbf{acc_{testing}}$ (%)
16	19	10	15	0.99	100	100
16	30	15	10	0.58	94.12	100
4	25	15	15	0.45	100	98.89
8	26	15	10	0.62	94.71	98.89
10	25	10	10	0.99	100	98.89
12	14	10	10	0.95	100	98.89
1	26	10	10	0.66	95.10	98.89
12	29	10	5	0.99	100	98.89

The top ten solutions showed significant accuracy both in the training and testing sets. The values of k and d_{max} does not appear to have any significant patterns as k varied between one and 16, while d_{max} ranged from 14 to 30 (the value of d_{max} needs to be sufficiently high to represent the MFCC features satisfactorily). Comparing between the LLE and non-LLE features, LLE features had managed to significantly reduce the number of features to represent the data. For example, for the best case in Table 2 ($d_{max} = 19$), a whopping

98.72% reduction in the number of features (original number of MFCC features = 1488).

Many factors influence the results of Multi-Layer Perceptron (MLP)-type neural networks [25], among them the value of the initialization weights. The initial weights are typically set to random values during the beginning of training, and the MLP's training algorithm modifies the weight values usually according to some second-order information such as gradient and others [24]. Because of this, the MLP's accuracy changes after each training run as different starting points are provided. Therefore, we decided to train the best network ten times and average its results. This would provide a better indication of the network performance under different initialization weights.

Table 3. Best network configuration trained using different initialization weights

k	d_{max}	h_{AC1}	h_{AC2}	Initial Random Seed	$acc_{training}$ (%)	$acc_{testing}$ (%)
16	19	10	15	0	100	92.22
16	19	10	15	5,000	100	91.11
16	19	10	15	10,000	93.53	93.33
16	19	10	15	15,000	93.73	93.33
16	19	10	15	20,000	100	93.33
16	19	10	15	25,000	100	91.11
16	19	10	15	30,000	92.35	92.22
16	19	10	15	35,000	92.55	93.33
				Average	96.35	92.82

Table 3 shows the training and testing accuracies of the autoencoder when trained ten time under different initialization seeds using optimal settings from Table 2. Due to the averaging effect, the average accuracies were lower than that of Table 2. However, as can be seen, the autoencoder is still very accurate with above 90% accuracy. Additionally, the results appear to be consistently good under different initialization parameters, demonstrating the strength of the feature representation, selection and classification approach proposed in this paper. Please refer to Fig. 1 to Fig. 11 for a breakdown of the classification results for each record in Table 3.

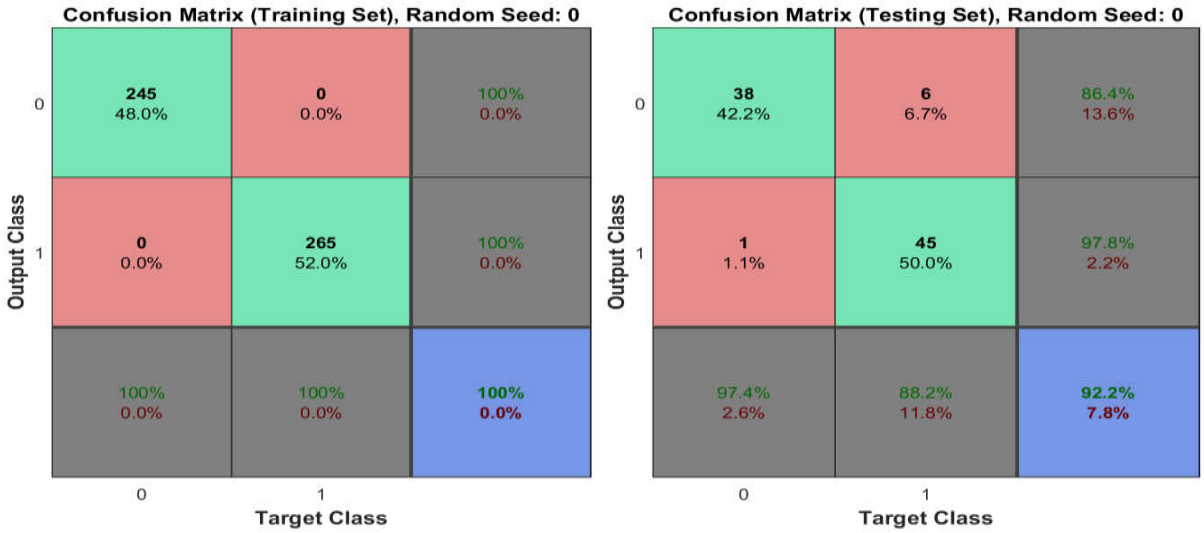

Fig.1. Confusion matrix for autoencoder with optimal parameters (Initial seed: 0)

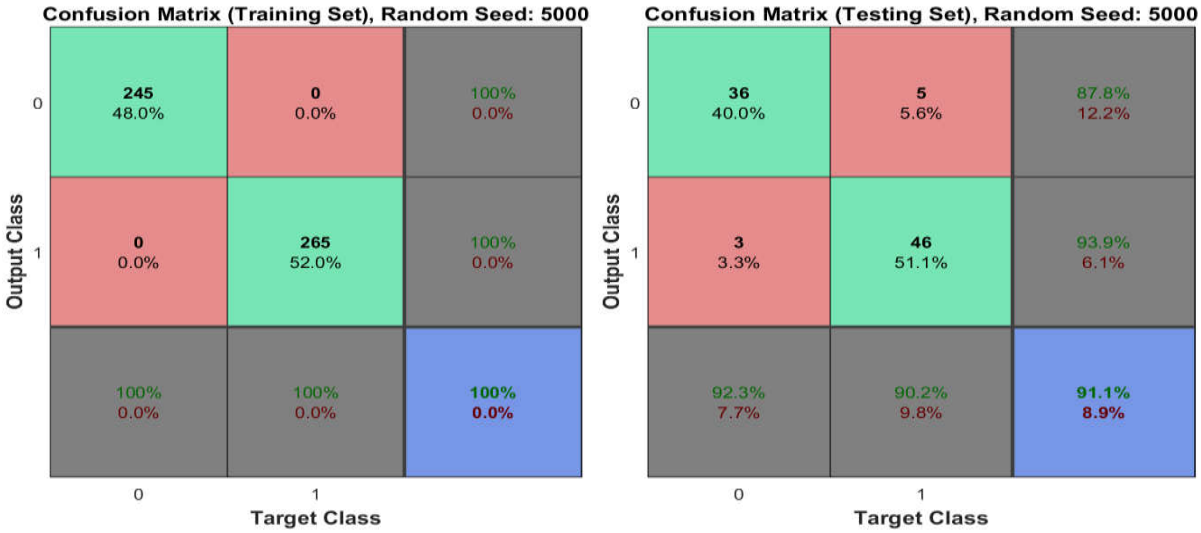

Fig.2. Confusion matrix for autoencoder with optimal parameters (Initial seed: 5,000)

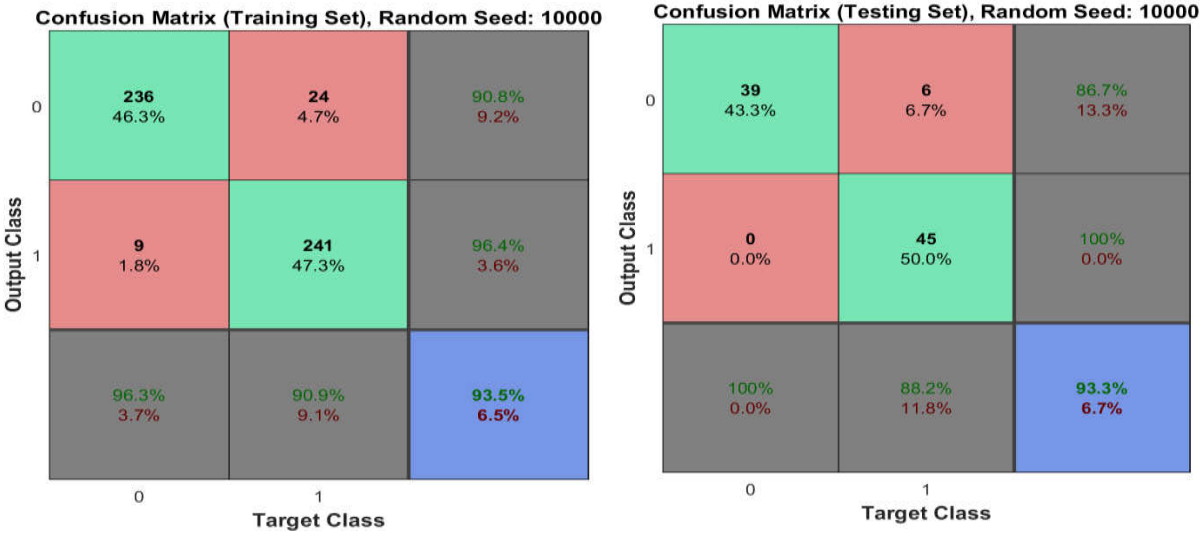

Fig.3. Confusion matrix for autoencoder with optimal parameters (Initial seed: 10,000)

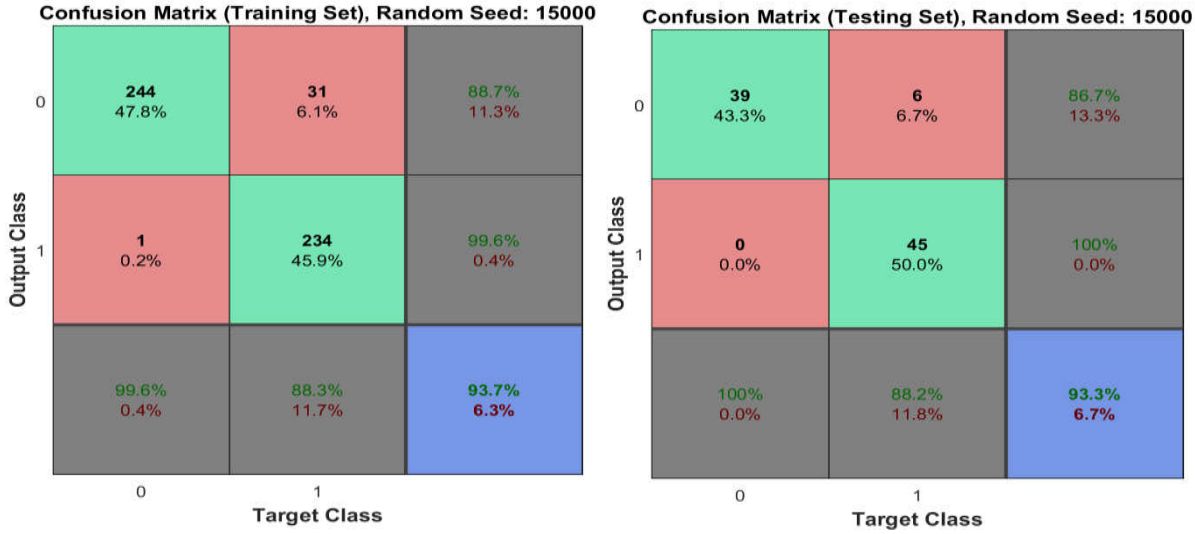

Fig.4. Confusion matrix for autoencoder with optimal parameters (Initial seed: 15,000)

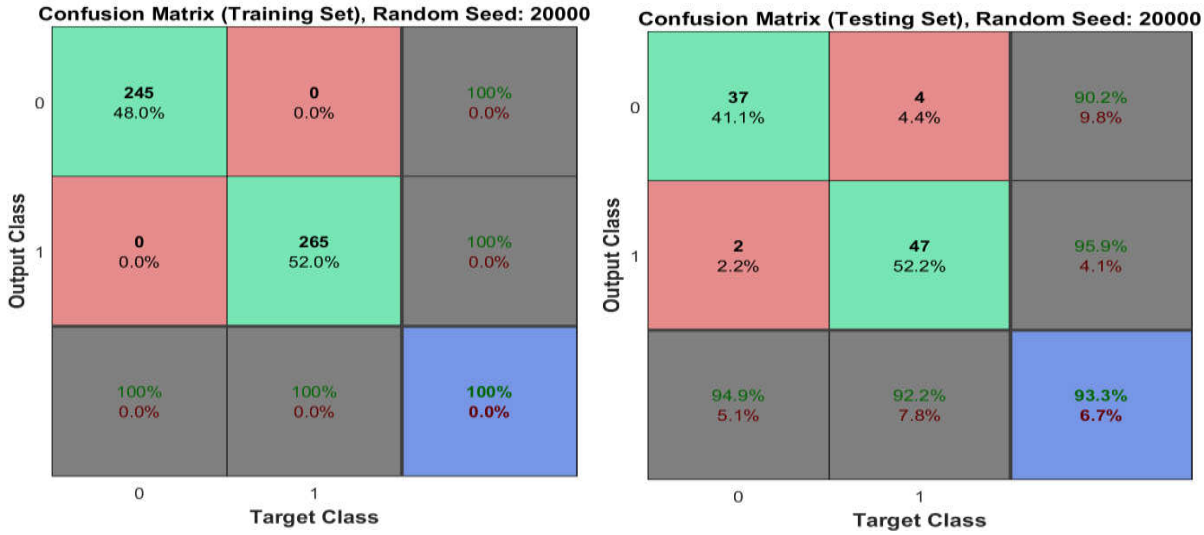

Fig.5. Confusion matrix for autoencoder with optimal parameters (Initial seed: 20,000)

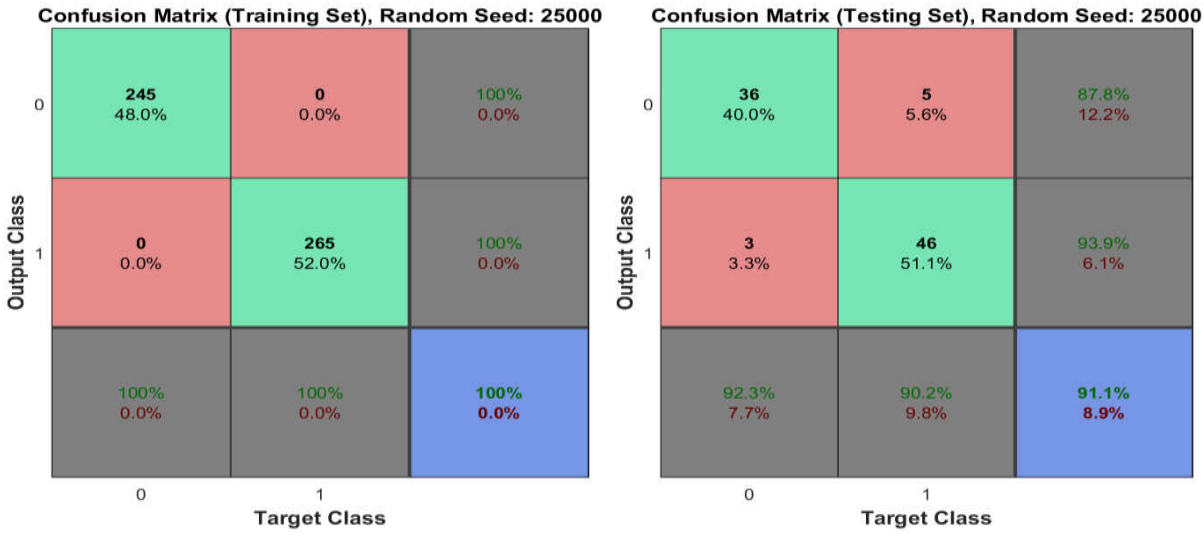

Fig.6. Confusion matrix for autoencoder with optimal parameters (Initial seed: 25,000)

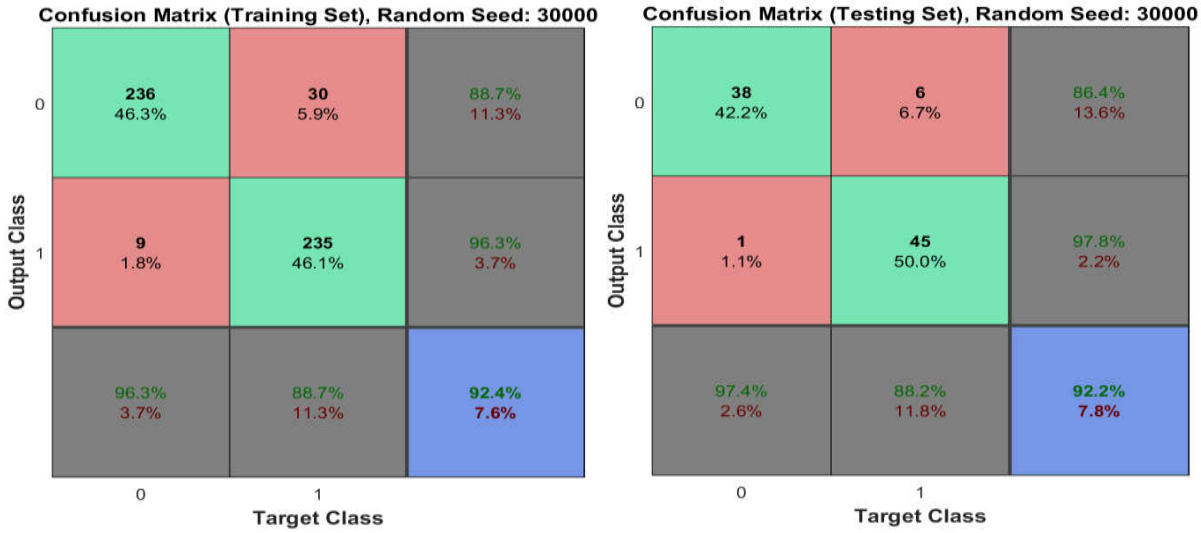

Fig.7. Confusion matrix for autoencoder with optimal parameters (Initial seed: 30,000)

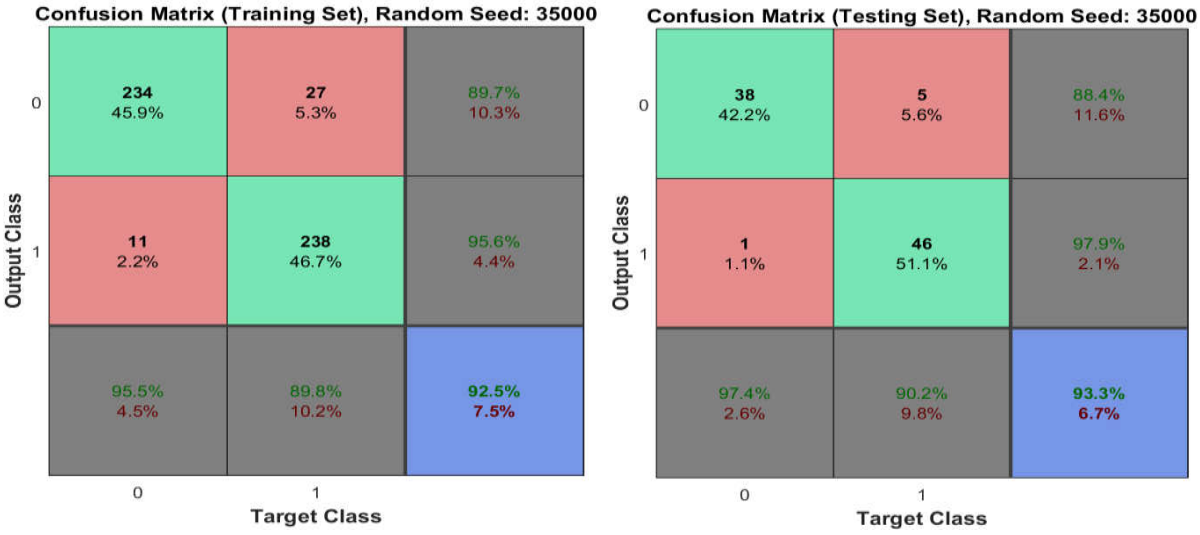

Fig.8. Confusion matrix for autoencoder with optimal parameters (Initial seed: 35,000)

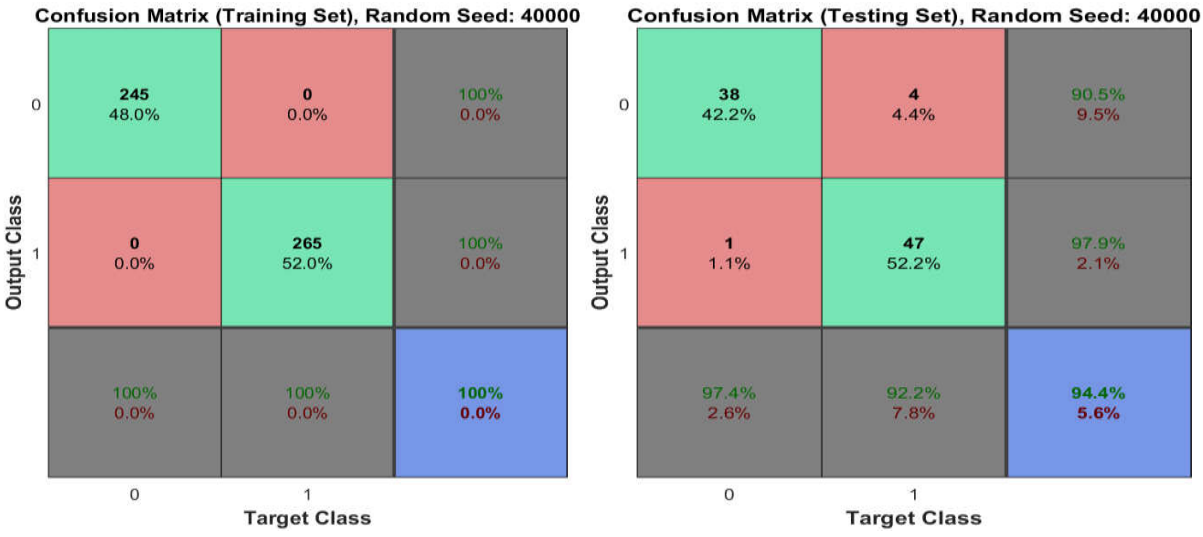

Fig.9. Confusion matrix for autoencoder with optimal parameters (Initial seed: 40,000)

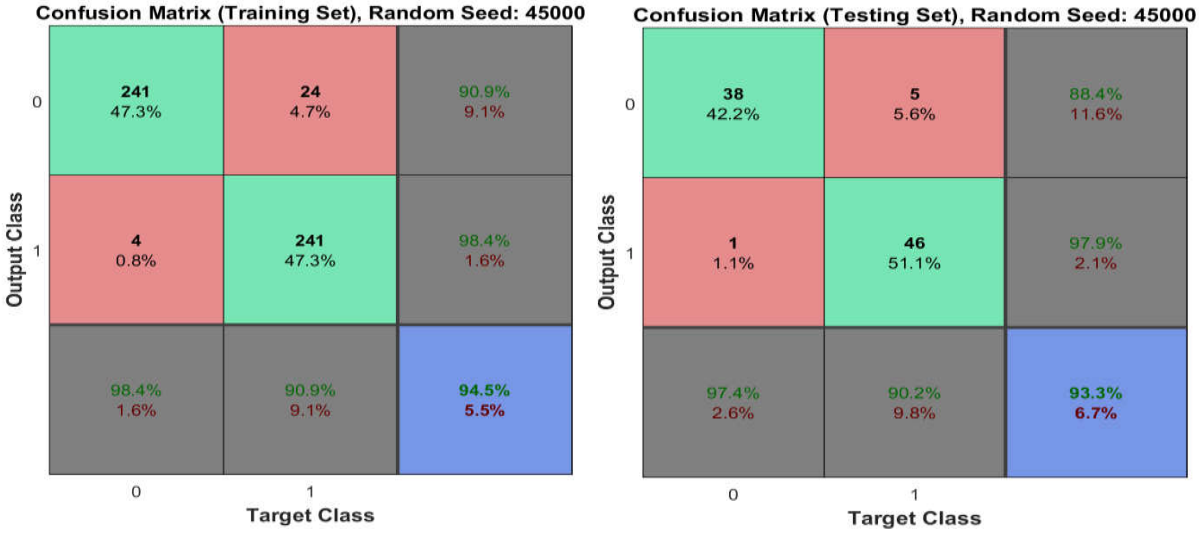

Fig.10. Confusion matrix for autoencoder with optimal parameters (Initial seed: 45,000)

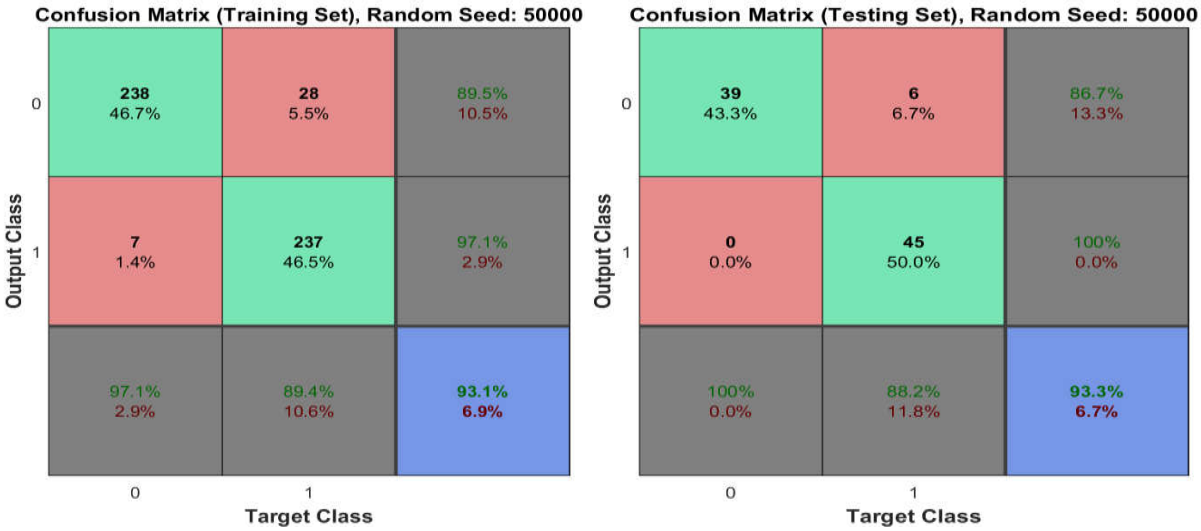

Fig.11. Confusion matrix for autoencoder with optimal parameters (Initial seed: 50,000)

4. CONCLUSION

A method for automated diagnosis of asphyxia has been presented in this paper. LLE was implemented on 600 MFCC features of normal and non-aspxiated newborns. LLE was used to reduce size of feature representation while enhancing them. Two stacked autoencoders were then trained to extract the necessary features for clasification. Extensive tests performed showed that the best-performing autoencoder network could produce 92.82% classification accuracy.

ACKNOWLEDGEMENTS

The authors would like to graciously acknowledge the Ministry of Higher Education and UniversitiTeknologi Mara for supporting this research work through Grant No: 600-RMI/DANA 5/3/ARAS(4/2015).

REFERENCES

[1] Sahak R, Mansor W, Khuan L Y, Yassin A I, Zabidi A, Rahman F Y. Choice for a support vector machine kernel function for recognizing asphyxia from infant cries. In IEEE Symposium onIndustrial Electronics and Applications, 2009, pp. 675-678

[2] Zabidi A, Mansor W, Lee Y K, Yassin A M, Sahak R. Particle swarm optimisation of mel-frequency cepstral coefficients computation for the classification of asphyxiated infant cry. In 3rd IEEE International Conference on Biomedical Engineering and Informatics, 2010, pp. 991-995

[3] Zabidi A, Mansor W, Lee Y K, Yassin I M, Sahak R. Binary particle swarm optimization for selection of features in the recognition of infants cries with asphyxia. In 7th IEEE International Colloquium on Signal Processing and its Applications, 2011, pp. 272-276

[4] Sahak R, Mansor W, Khuan L Y, Zabidi A, Yassin A I. Detection of asphyxia from infant cry using support vector machine and multilayer perceptron integrated with orthogonal least square. In IEEE-EMBS International Conference on Biomedical and Health Informatics, 2012, pp. 906-909

[5] Sahak R, Mansor W, Khuan L Y, Ihsan A Y M, Zabidi A. An orthogonal least square approach to select features of infant cry with asphyxia. In 6th IEEE International Colloquium on Signal Processing and Its Applications, 2010, pp. 21-23

[6] Zabidi A, Khuan L Y, Mansor W, Yassin I M, Sahak R. Classification of infant cries with asphyxia using multilayer perceptron neural network. In 2nd IEEE International Conference on Computer Engineering and Applications, 2010, pp. 204-208

[7] Zabidi A, Mansor W, Khuan L Y, Yassin I M, Sahak R. Three-dimensional particle swam optimisation of mel frequency cepstrum coefficient computation and multilayer perceptron neural network for classifying asphyxiated infant cry. In IEEE International Conference on

Computer Applications and Industrial Electronics, 2011, pp. 290-293

[8] Er O, Cetin O, Bascil M S, Temurtas F. A comparative study on Parkinson's disease diagnosis using neural networks and artificial immune system. Journal of Medical Imaging and Health Informatics, 2016, 6(1):264-268

[9] Narain R, Saxena S, Goyal A K. Cardiovascular risk prediction: A comparative study of Framingham and quantum neural network based approach. Patient Prefer Adherence, 2016, 10:1259-1270

[10] Gautam M K, Giri V K. An approach of neural network for electrocardiogram classification. APTIKOM Journal on Computer Science and Information Technologies, 2016, 1(3):115-123

[11] Gautam M K, Giri V K. A neural network approach and wavelet analysis for ECG classification. In IEEE International Conference on Engineering and Technology, 2016, pp. 1136-1141

[12] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553):436-444

[13] Romero A, Gatta C, Camps V G. Unsupervised deep feature extraction for remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(3):1349-1362

[14] Singhal V, Gogna A, Majumdar A. Deep dictionary learning vs deep belief network vs stacked autoencoder: An empirical analysis. In International Conference on Neural Information Processing, 2016, pp. 337-344

[15] Le Q V. A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks. 2015, http://www.cs.mcgill.ca/~dprecup/courses/ML/Materials/dl-tutorial2.pdf

[16] Patraucean V, Handa A, Cipolla R. Spatio-temporal video autoencoder with differentiable memory. In 4th International Conference on Learning Representations, 2016, pp. 1-13

[17] Lu J, Deshpande A, Forsyth D. CDVAE: Co-embedding deep variational auto encoder for conditional variational generation. 2016, https://arxiv.org/pdf/1612.00132.pdf

[18] Zhu Z, Wang X, Bai S, Yao C, Bai X. Deep learning representation using autoencoder for 3D shape retrieval. Neurocomputing, 2016, 204:41-50

[19] Leng B, Guo S, Zhang X, Xiong Z. 3D object retrieval with stacked local convolutional autoencoder. Signal Processing, 2015, 112:119-128

[20] Louizos C, Swersky K, Li Y, Welling M, Zemel R. The variational fair auto encoder. In 4th International Conference on Learning Representations, 2016, pp. 1-11

[21] Li J, Luong M T, Jurafsky D. A hierarchical neural autoencoder for paragraphs and documents. 2015, https://arxiv.org/pdf/1506.01057.pdf

[22] Mareschal D, French R M. TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. Philosophical Transactions of the Royal Society B: Biological Sciences, 2017, 372(1711):1-9

[23] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000, 290(5500):2323-2326

[24] Mohamad N, Zaini F, Johari A, Yassin I, Zabidi A. Comparison between levenberg-marquardt and scaled conjugate gradient training algorithms for breast cancer diagnosis using MLP. In 6th IEEE International Colloquium on Signal Processing and Its Applications, 2010, pp. 1-7

[25] Ihsan M Y, Azlee Z, Rozita J, Megat S A M A, Rahimi B, Abu H A H, Zairi I R, Comparison between cascade forward and multi-layer perceptron neural networks for NARX Functional electrical stimulation (FES)-based muscle model. International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(1):215-221

How to cite this article:
Yassin M I, Zabidi A, Ismail N, Zaman F H K, Shafie M F, Rizman Z I. Infant asphyxia detection using autoencoders trained on locally linear embedded-reduced mel frequency cepstrum coefficient (mfcc) features. J. Fundam. Appl. Sci., 2017, *9(3S), 716-729*