VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

FEATURE EXTRACTION OF EEG SIGNAL USING WAVELET TRANSFORM FOR AUTISM CLASSIFICATION

Lung Chuin Cheong, Rubita Sudirman and Siti Suraya Hussin
Faculty of Electrical Engineering, Universiti Teknologi Malaysia UTM Johor Bahru, Johor, Malaysia
E-Mail: rubita@fke.utm.my

ABSTRACT

Feature extraction is a process to extract information from the electroencephalogram (EEG) signal to represent the large dataset before performing classification. This paper is intended to study the use of discrete wavelet transform (DWT) in extracting feature from EEG signal obtained by sensory response from autism children. In this study, DWT is used to decompose a filtered EEG signal into its frequency components and the statistical feature of the DWT coefficient are computed in time domain. The features are used to train a multilayer perceptron (MLP) neural network to classify the signals into three classes of autism severity (mild, moderate and severe). The training results in classification accuracy achieved up to 92.3% with MSE of 0.0362. Testing on the trained neural network shows that all samples used for testing is being classified correctly.

Keywords: discrete wavelet transforms (DWT), electroencephalogram (EEG), classification, feature extraction, sensory response.

INTRODUCTION

Electroencephalogram (EEG) is a non-evasive technique used on the human skull to acquire electrical impulse produced from neuron activation in the brain. EEG electrodes are attached to the specific region of the scalp according to the type of study to be conducted. EEG is able to measure electrical signal from the human brain in the range of 1 to 100 microvolt (µV) (Teplan, 2002). There have been numerous studies on EEG classification, looking for new possibilities in the field of Brain-Computer Interface (BCI), neurobiological analysis and automatic signal interpretation systems (Frédéric *et al*., 2006).

EEG signal can be categorized to bands of different frequency ranges. Delta wave lies below the frequency of 4Hz. Theta lies in the range of 4Hz to 8Hz while Alpha wave lies between 8Hz to 13Hz. The range of Beta wave lies in 14Hz to 32Hz where beyond 32Hz lies the Gamma wave. These frequency bands each corresponds to different activities carried out by the subject (Teplan, 2002).These different band of frequencies each contains certain information of the brain activity. However, the information hides within the EEG signal is not directly analytical by the human eyes. However, information on neural connectivity may be revealed with the analysis of signal complexity on multiple scale. The result of this analysis would be diagnostically useful (Varela *et al*., 2001).

Analyzing EEG signals basically involves few steps of signal processing; usually begin by data collection which require the subject to perform certain task. In this study, the selected channel of interest is first artefact-removed and filtered with a band pass filter with a pass band frequency of 0.4-60Hz to eliminate the power line frequency, noise and extremely low frequency.

Given the fact that EEG signals are non-stationary, time-varying computation is required to extract the features from the signal in order to be classified (Suleiman and Fatehi, 2007). Wavelet transform, being one of the non-stationary time-scale analysis methods, is used to decompose the signal for feature extraction. The transient features of EEG signals are able to be accurately captured (Jahankhani *et al.,* 2006). The extracted features are then used to train a neural network for classification purpose. All the processes are performed and encoded in MATLAB.

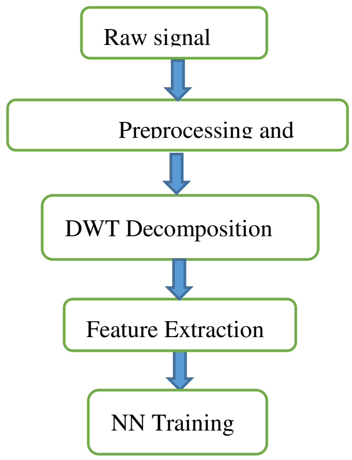


Figure-1. Processes involved in this study.

METHOD

Data acquisition and experimental setup

This study utilizes sensory data collected by Sudirman and Hussin, (2014) from 30 autism children

ARPN Journal of Engineering and Applied Sciences

aged between 3 to 10 years old. Among these children, 5 of them have mild autism, 11 have moderate autism and 14 have severe autism. All of them performed tasks on taste sensory, involving stimulation of three taste, which is sweet, sour and salty. Stimulation of the three tastes is done with sugar solution, vinegar solution and salt solution. While the data is being read, the subjects' eyes is blindfolded except during visual task to prevent visual artefact. In between different taste stimuli, the subjects are given plain water to rinse away the residual taste stimuli. The brain waves are recorded using Neurofax JE-921A EEG machine together with an electrode cap following the standard 10-20 international electrode placement system. The data was sampled with an interval of 2ms and was stored as ASCII files in the recording computer (Sudirman and Hussin, 2014). Out of the 30 samples, 26 are used for neural network training and 4 are reserved for testing (1 mild, 1 moderate and 2 severe) on the trained neural network.

Signal preprocessing

From the collected multichannel signal, only the parietal lobe channels, C_3, C_4 and C_Z which is related to the taste sensory is used for processing. The signal is first epoched and the epoch with artefact and corrupted signal are removed automatically using simple voltage threshold method. The threshold is set to the standard deviation of the whole signal of a particular channel. Flat lines are removed using blocking and flat line function. Both are performed using the source code of ERPLAB.

Then, the signal is filtered using a band pass filter with pass band frequency of 0.4Hz to 60Hz and filter order of 60 to remove the extremely low frequency components such as those caused by movement and breathing (less than 0.4Hz) (Suleiman and Fatehi, 2007), power line frequency (60Hz) and noise (more than 60Hz).

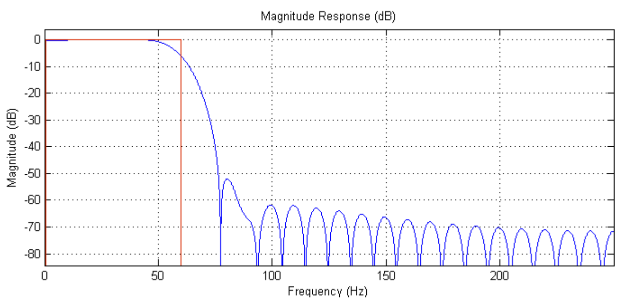


Figure-2. Bandpass filter used to filter raw signal.

Feature extraction in time domain using DWT

Wavelet transform is a non-stationary time-scale analysis method suitable to be used with EEG signals. It is a useful tool to separate and sort non-stationary signal into its various frequency elements in different time-scales (Hazarika *et al*., 1997).

Quantitatively, discrete wavelet transform can be applied to decompose a discrete time series, $f(n)$ where $f(n)$ is the discrete signal of $f(x)$ sampled at 500Hz in this study, to its sub-bands of wavelet coefficients that contains the feature (Hazarika et al., 1997). The wavelet coefficients can be computed by dilation and translation of the mother wavelet $\varphi_{s,\tau}(x)$ as shown in (1), where $s, \tau \in R, s > 0$, and R is the wavelet space, while s and τ are the scaling factor and shifting factor respectively (Murugappan et al., 2010).

$$\varphi_{s,\tau}(x) = \frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right) \quad (1)$$

The decomposition is computed by filtering the discrete signal $f(n)$ repeatedly up to a predetermined level N. The filter consist of a low pass filter to obtain the approximation coefficient (CA) and high pass filter to obtain the detailed coefficient (CD) (Murugappan et al., 2010). After each level of filter, the signal is down-sampled by half the sampling frequency in the previous level $N-1$ since the frequency element is reduced by half.

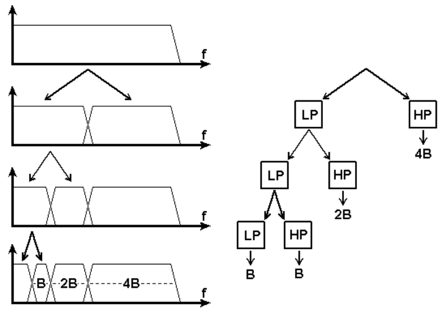


Figure-3. Level 3 decomposing the signal f(n).

Daubechies 4 (db4) wavelet is used as the mother wavelet in this study since that it is most suitable to process biomedical signals. The input signal $f(n)$ has a frequency band of 0-500Hz. With the interest area of 0-60Hz for EEG signal, the signal should be decomposed up to level 8 to be fully separated into the lowest frequency delta band but since the relevant frequency band lies in the alpha rhythm (8-16Hz), the filtered signal will be decomposed only up to level 6 to obtain the alpha band in CD6 as shown in Table-4. The detail coefficient of level 1, 2 and 3 is considered noise as their frequency did not lie within the EEG frequency of 0-60Hz.

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Table-1. Wavelet coefficient and its signal information.

Wavelet coefficient	Frequency (Hz)	Signal information
D1	250 – 500	Noise
D2	125 – 250	Noise
D3	63 – 125	Noise
D4	32 - 63	Gamma
D5	16 - 32	Beta
D6	8 - 16	Alpha
D7	4 - 8	Theta
D8	0 - 4	Delta

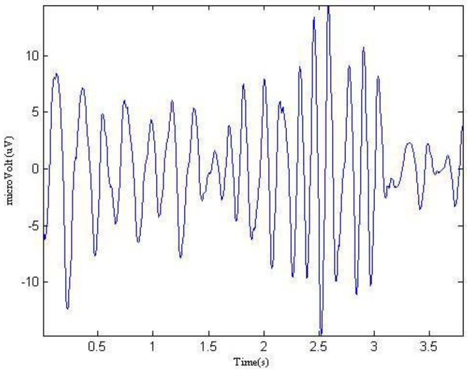


Figure-4. Reconstructed CD6 coefficient containing alpha band of the signal.

The wavelet coefficient of the decomposed signal is still too large and not suitable to be directly used for pattern recognition with neural network. Therefore, feature extraction is done to reduce the signal to its representation set of features vector by simplifying the description of a large set of data (Nandish *et al.*, 2012).

The feature can be extracted into time domain feature and frequency domain feature. The most simple and commonly used feature to represent the large set of data is by statistical approach of the time domain feature. Statistical feature such as mean, median, mode, standard deviation, maximum and minimum can be used. In this study, standard deviation of the wavelet coefficient discrete-time series is computed using (2), where n represents the discrete signal length while x represents the signal level of the particular n.

$$\sigma^2 = \frac{1}{n-1}\sum_{i=0}^{n-1}(x_i - \mu)^2 \qquad (2)$$

Other methods such as those in frequency domain can also be used for feature extraction. For example, previous study by Suleiman and Fatehi, (2007) uses STFT and FFT to extract feature in the frequency domain. The 2 different methods yields different result of classification accuracy (Suleiman and Fatehi, 2007).

Classification

Neural network are composed of interconnecting artificial neurons, modelling in the way of how human brain works. Various neural network architecture have been developed over the years for different functions, where one of the most popular architecture is the feed forward network. Feed forward network is commonly known for its ability to recognize pattern, predict and fit nonlinear function (Nandish *et al.*, 2012).

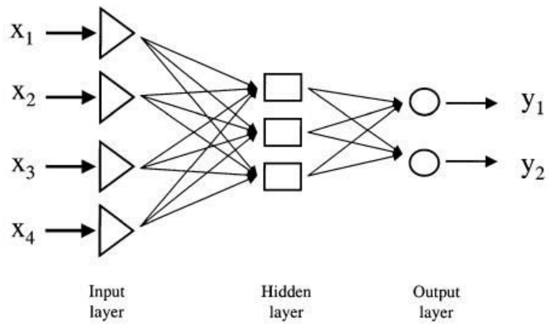


Figure-5. Feed forward neural network.

This work involve the use of multilayer perceptron (MLP) feed forward neural network as the signal classifier. It doesn't require a large training set to learn and hence reducing the operation overhead (Jahankhani *et al.*, 2006). Training the neural network require two sets of data, which is the input data that represents the information of the signal and the target data that defines desired output of the neural network.

In this study, features of the discrete-time wavelet coefficient CD6 is presented to the neural network for training with scaled-conjugate backpropagation algorithm. The accuracy of the neural network is measured by the percentage of correct classification shown in (3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + TN} \times 100\% \qquad (3)$$

The computation of the accuracy takes in account of the true positive (TP), true negative (TN), false positive (FP) and false negative (FN):

- TP = Number of correctly classified positive samples
- TN = Correctly classified negative samples while
- FP = Negative sample being classified as positive
- FP = Positive sample classified as negative.

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Neural network training parameters used in this study is shown in Table-2. Training stops when any of the parameter is fulfilled. Default data division setup (75% training, 15% validation and 15% testing) and 10 hidden layer is used to obtain the best cross entropy and percent error in the neural network training GUI. A script is then generated and performance is further improved by using command line approach until a desirable accuracy and MSE is obtained.

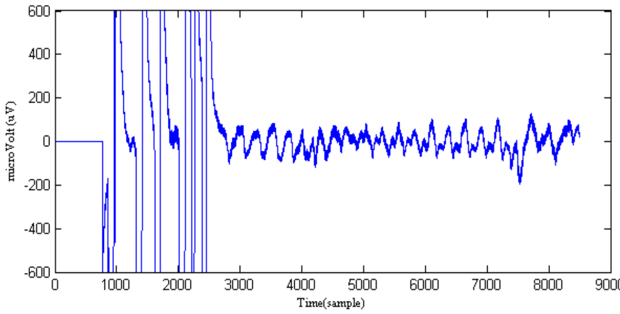
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - \propto_i)^2 \tag{4}$$

Table-2. Training parameters of the neural network.

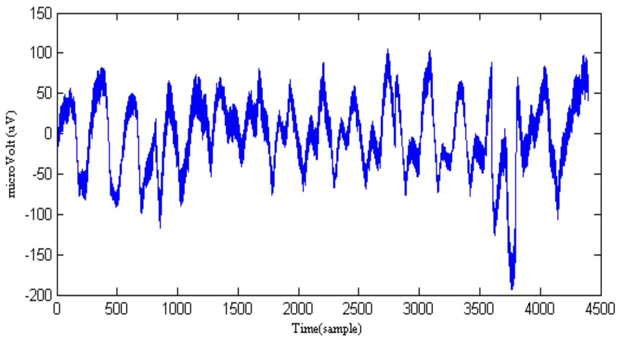
Maximum number of epochs	1000
Minimum performance gradient	0.000001
Performance goal	0
Maximum validation failures	5

RESULTS AND DISCUSSIONS

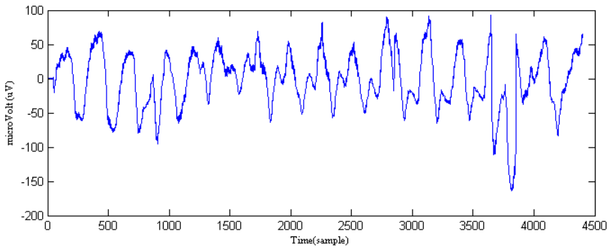
Figure 6(a) shows one of the raw signal acquired. The signal after artefact removal, rejection of corrupted epochs and removal of flat line is shown in Figure-6(b) while filtering gives a clean signal as in Figure-6(c).



(a)



(b)



(c)

Figure-6. (a) Raw EEG signal, (b) Removed artefact, corrupted signal and flat line, (c) Clean signal after filtering.

DWT decomposition is performed on C_3, C_4 and C_Z channel of the clean signal to obtain the alpha band which contains information that reflects the sensory responsiveness during a relaxed state. The level 6 decomposition yields 6 detailed coefficients containing different band of frequencies as shown in Figure-7. Alpha band signal as shown in Figure-4 lies in the detailed coefficient at the 6th level decomposition (CD6).

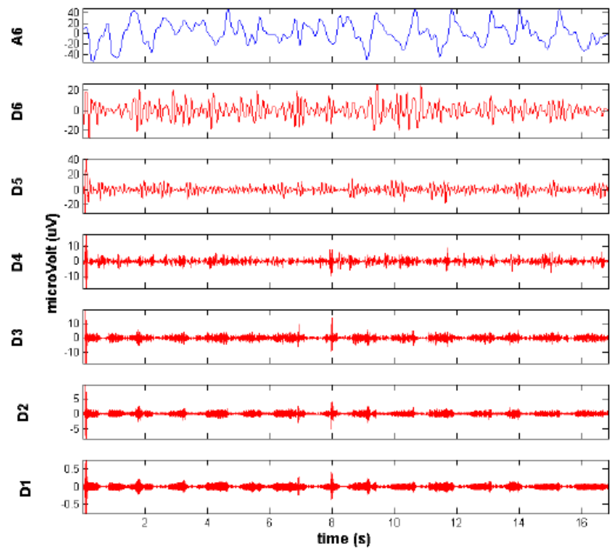


Figure-7. Level 6 decomposition of the signal.

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Table-3. Features extracted from 26 subjects and their corresponding autism classes.

Expected class	Salty, σ^2 (μV)			Sour, σ^2 (μV)			Sweet, σ^2 (μV)		
	C_3	C_Z	C_4	C_3	C_Z	C_4	C_3	C_Z	C_4
Severe	117.14	119.44	113.38	130.90	189.93	142.32	134.51	140.91	144.04
Moderate	84.71	89.74	68.81	88.52	100.06	102.00	60.73	68.42	58.84
Moderate	98.54	146.20	131.41	46.69	33.24	32.57	93.21	100.74	118.98
Severe	179.89	178.90	214.50	198.89	196.99	195.94	122.93	137.09	130.12
Moderate	154.25	129.66	162.06	67.55	55.05	62.23	101.77	102.74	98.93
Moderate	149.32	126.15	108.94	47.18	47.80	51.07	47.03	55.41	50.17
Moderate	71.49	75.75	77.86	92.15	85.48	89.25	82.04	81.59	79.86
Severe	124.61	222.02	261.51	81.44	44.88	88.70	96.11	91.68	101.73
Mild	52.10	48.23	45.08	46.41	48.96	40.36	36.69	36.29	43.41
Severe	172.26	183.89	222.51	133.83	146.68	134.89	157.43	210.36	151.54
Moderate	100.96	98.86	130.14	78.63	80.87	82.06	93.58	78.41	90.30
Mild	80.59	58.48	64.34	67.34	76.55	93.04	75.12	72.15	74.34
Severe	314.29	296.12	337.38	165.43	153.63	147.12	82.04	81.59	79.86
Severe	244.48	305.36	209.36	272.81	370.34	283.65	247.92	283.59	254.78
Moderate	37.31	38.55	45.93	216.21	177.18	81.10	147.16	38.17	29.99
Moderate	146.46	136.62	138.21	71.86	82.07	79.64	59.90	52.18	46.55
Severe	96.97	106.01	109.01	163.21	149.82	157.79	98.48	97.13	103.48
Severe	122.88	130.59	125.47	103.28	107.98	99.93	124.46	148.08	143.33
Mild	88.40	77.24	52.96	37.02	52.20	47.46	115.17	42.12	153.41
Severe	57.84	51.92	60.38	166.46	181.28	191.80	99.05	102.99	116.65
Severe	210.84	183.59	202.89	72.74	72.81	70.84	52.60	52.11	49.91
Severe	166.66	176.92	162.85	94.77	98.34	97.65	116.03	120.38	151.40
Mild	52.31	51.65	61.11	49.03	45.71	57.76	42.30	47.65	54.93
Moderate	113.31	111.26	111.83	99.77	112.47	114.09	85.74	94.33	85.71
Moderate	55.22	68.88	55.58	59.98	67.53	66.51	114.88	103.73	101.37
Severe	202.06	200.13	258.68	192.18	179.16	199.35	233.97	222.31	202.41
Mean	**126.73**	**131.24**	**135.85**	**109.40**	**113.73**	**108.04**	**104.65**	**102.39**	**104.46**
SD	**66.32**	**71.52**	**78.25**	**61.81**	**73.64**	**59.38**	**50.96**	**60.02**	**52.50**

Feature extraction is performed in time domain by computing the standard deviation of the discrete signal level of the alpha band (D6) in microvolt (μV) using equation (2) for all 3 taste sensory with 3 channels each. The extracted features of the 3 taste sensory are shown in Table-3 with mean and standard deviation of the features in each channel.

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

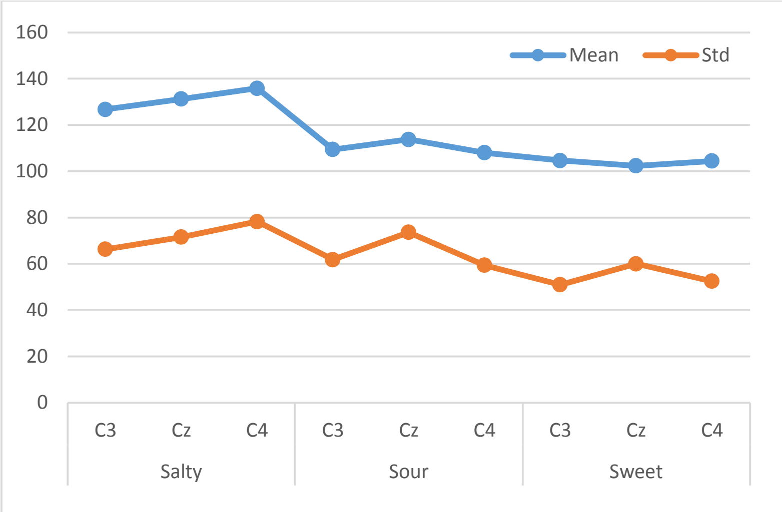


Figure-8. Mean and standard deviation of features across channels and taste.

From Figure-8, it was observed that the mean of the 3 features of salty taste is slightly higher (126.73 μV, 131.24 μV, 135.85 μV) compared to that of sweet taste (104.65 μV, 102.39 μV, 104.46 μV), which indicates that the feature value acquired by salty taste is higher. This is potentially due to the children being not comfortable with the taste of salt (Sudirman and Hussin, 2014).

Generally, it can be seen that feature of subjects with mild autism generally have a lower value, which also has higher coherence across different type of taste sensory. Subjects with severe autism has higher standard deviation, where the coherence of the standard deviation across different type of taste sensory is lower. Standard deviation is the lowest at the C_4 channel of subject 3 (32.57 μV) with sour taste and highest in C_4 channel of subject 19 (337.38 μV) with salty taste. Finding of the highest feature value on salty taste is similar to the study by Sudirman and Hussin, (2014), where the highest standard deviation obtained is 336.83 μV from salty taste.

This dataset is used as an input data consisting of 26 samples with 9 elements and is fed into the neural network for training. Trial and error is performed to obtain the suitable data division ratio and number of hidden neurons. The settings that gave the best performance in cross entropy and percent error is shown in Table-4. The neural network is designed to have 9 input neurons for the 9 features, 8 hidden neurons, and 3 output neurons for the 3 output classes, which is mild, moderate and severe autism.

Table-4. Network setup that gives best performance.

Data division setup	
Training percentage	65 %
Validation percentage	25 %
Testing percentage	10 %
Hidden layer setting	
Hidden neurons	8

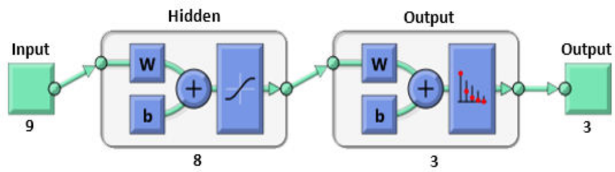


Figure-9. Architecture of the neural network.

Training of the neural network with settings shown in Table-4 yields accuracy of 92.3%. Despite the high accuracy, the mean squared error (MSE) is quite high at 0.0362 with the cross entropy at 0.15822. This is probably due to the large number of features and the limited amount of samples for the neural network to generalize the data.

The confusion matrix shown in Figure-10 shows that only 1 sample from moderate autism and 1 from severe autism is wrongly classified during training and testing. The best performance is obtained after 18 iterations with the best validation performance obtained at epoch 12 and gradient of 0.0729 as shown in the performance plot in Figure-11. The constantly decreasing cross-entropy indicates that the cross-entropy performance is decreased as the training proceeds.

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

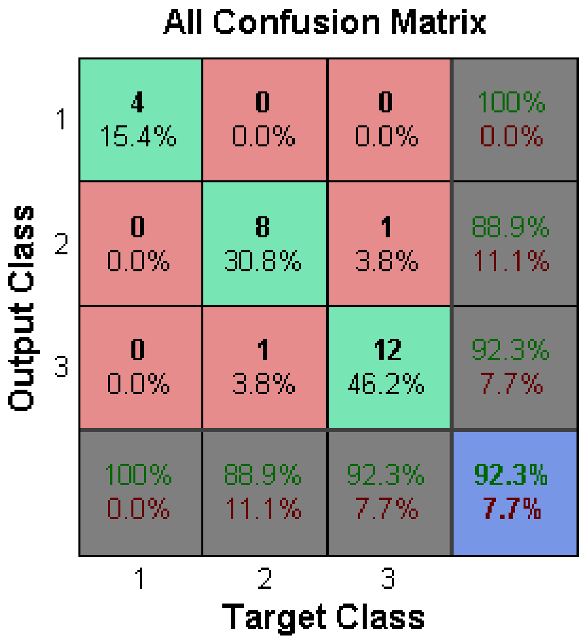


Figure-10. Confusion matrix showing output of training.

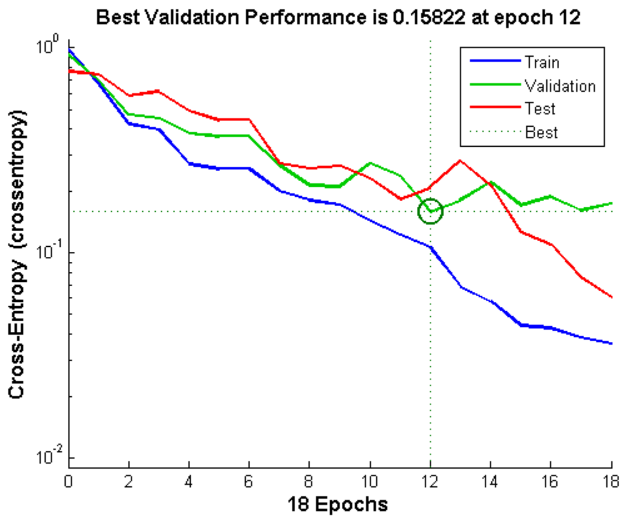


Figure-11. Performance plot of the training.

The trained neural network is tested with the 4 samples reserved earlier. These samples perform the similar preprocessing and feature extraction steps. Then, they were classified with the trained neural network. Classification shows that all 4 samples is correctly classified as shown in Table-5.

Table-5. Output of classification testing.

Subject number	Expected severity	Classification output (%)		Output class
6	Mild	Mild	65.20	Mild
		Moderate	34.76	
		Severe	0.04	
10	Moderate	Mild	6.77	Moderate
		Moderate	92.36	
		Severe	0.86	
26	Severe	Mild	0.06	Severe
		Moderate	6.27	
		Severe	93.66	
36	Severe	Mild	0.30	Severe
		Moderate	11.06	
		Severe	88.64	

Previous study by Suleiman and Fatehi, (2007) who performed feature extraction with STFT to perform classification with MLP for BCI purpose achieve average classification accuracy of 85.99% for all channels which is slightly lower than by using DWT. While wavelet transform is a time-scale analysis method, this simple comparison of feature extraction with frequency analysis might suggest that time domain features provides a slightly clearer class boundary than frequency domain features. However, the difference might also due to the difference in training parameters being used during neural network training and different linearity of dataset.

CONCLUSIONS

As EEG signal analysis is gaining popularity in the field of neuroscience, brain-computer interface and physiological evaluation, a robust method of feature extraction must present to increase the reliability of the method in providing a representation of the data.

DWT's ability to decompose a signal down to its frequency components shows that it is a simple and direct method to analyze EEG signals in different frequency band representing different activities in the brain. Results shows that features extracted with DWT is able to display various correlations between standard deviation of the alpha band and the feature characteristics of different taste sensory and also the severity of autism. This makes DWT a suitable tool to analyze EEG signal of autism patients. Training of the neural network with features extracted with DWT shows that the network is able to achieve classification accuracy at 92.3% despite having high MSE of 0.0362. The trained network is able to classify all testing data correctly.

In future, researchers are suggested to find the best combination of feature extraction method and classifier that give the best accuracy and performance. This can maximize the potential of using EEG classification as a reliable method to diagnose autism.

ACKNOWLEDGEMENT

This work is supported by the Faculty of Electrical Engineering, Universiti Teknologi Malaysia with funding from the Ministry of Science, Technology and Innovation of Malaysia (MOSTI) under Vot. 4S094. The author would like to express gratitude to those who provided support, guidance and technical knowledge during the course of this work.

REFERENCES

Frédéric, A., Nizar, K., Khalifa, B. and Hedi, B. 2006. Supervised Neuronal Approaches For EEG Signal Classification: Experimental Studies. The 10th IASTED International Conference on Artificial Intelligence and Soft Computing.

Hazarika, N., Chen, J. Z., Tsoi, A. C. and Sergejew, A. 1997. Classification of EEG signals using the wavelet transform. Digital Signal Processing Proceedings, 1997. DSP 97. 1997 13th International Conference. 1, pp. 89-92.

Jahankhani, P., Kodogianni, V. and Revett, K. 2006. EEG Signal Classification Using Wavelet Feature Extraction and Neural Networks. Modern Computing. pp. 120-124.

Murugappan, M., Ramachandran, N. and Sazali, Y. 2010. Classification of human emotion from EEG using discrete wavelet transform. Journal of Biomedical Science and Engineering. 3, pp. 390-396.

Nandish, M., Michahial, S., P, H. K. and Ahmed, F. 2012. Feature Extraction and Classification of EEG Signal Using Neural Network Based Techniques. International Journal of Engineering and Innovative Technology (IJEIT). 2.

Sudirman, R. and Hussin, S. S. 2014. Sensory Responses of Autism via Electroencephalography for Sensory Profile. Control System, Computing and Engineering (ICCSCE), 2014 IEEE International Conference. pp. 626-631.

Suleiman, A. B. R. and Fatehi, T. A. H. 2007. Features Extraction Techniques of EEG Signal for BCI Applications. Faculty of Computer and Information Engineering, Department College of Electronics Engineering, University of Mosul, Iraq.

Teplan, M. 2002. Fundamentals of EEG Measurement. Measurement Science Review. 2(2).

Varela, F., Lachaux, J., Rodriguez, E. and Martinerie, J. 2001. The brainweb: phase synchronization and large-scale integration. Nature Reviews Neuroscience. 2, pp. 229-239.