Electrooculography-based Eye Movement Classification using Deep Learning Models

Thibhika Ravichandran[1]
[1]*Department of Electrical and Electronic Engineering Universiti Teknologi PETRONAS*
Seri Iskandar, Perak, Malaysia
thibhika2704@gmail.com

Nidal Kamel[1]
[1]*Centre for Intelligent Signal and Imaging Research (CISIR) Department of Electrical and Electronic Engineering Universiti Teknologi PETRONAS*
Seri Iskandar, Perak, Malaysia
nidalkamel2@hotmail.com

Abdulhakim A. Al-Ezzi[1]
[1]*Centre for Intelligent Signal and Imaging Research (CISIR) Department of Electrical and Electronic Engineering Universiti Teknologi PETRONAS*
Seri Iskandar, Perak, Malaysia
abdulhakim_17007021@utp.edu.my

Khaled Alsaih[2]
[2]*Laboratoire Hubert Curien Universit´e de Lyon* Saint- ´Etienne, France
khaledalsaih@gmail.com

Norashikin Yahya[1]
[1]*Centre for Intelligent Signal and Imaging Research (CISIR) Department of Electrical and Electronic Engineering Universiti Teknologi PETRONAS*
Seri Iskandar, Perak, Malaysia
norashikin_yahya@utp.edu.my

***Abstract*—Amyotrophic lateral sclerosis (ALS), also known as motor neuron disease (MND), is a specific disease that causes the death of neurons controlling voluntary muscles. Most ALS patients eventually lose the ability to walk, use their hands, speak, swallow, and breathe. In this paper, we use the electrooculogram (EOG) signals captured using four sensors placed on the controlling muscles of the eye movement in horizontal and vertical directions to classify four different eye movements. The classifier output is used to control a wheelchair, or any other device developed to help ALS patients in performing their daily needs. Contrary to the classical classification techniques where features are extracted first from the EOG signals, then used with a trained classifier, in this paper the EOG signals are fed directly into two deep neural networks using, respectively, the long-short term memory (LSTM) and the convolutional neural network (CNN). The results show an accuracy of 88.33% for the LSTM network and 90.3% for the CNN network in eye movement classification.**

***Index Terms*—EOG, LSTM, neural network, CNN, ALS**

I. Introduction

Technology is essential to ease our daily life activities and to aid people with disabilities. Nowadays, the interaction between human and assistive technologies in what is generally referred to as the human-machine interface (HMI) is getting better. However, patients diagnosed with amyotrophic lateral sclerosis (ALS) have no choice but to depend entirely on their family members or caregivers for their daily activities, movement, communication, and housekeeping.

Recently, many HMI systems for driving devices are made available. These devices are based on eye movements captured by a camera to drive the device. However, the requirement of having the eye facing the camera over lengthy durations is a significant constraint. In this paper, we replace the camera with few sensors placed over the moving muscles of the eye to capture the eye movement. These sensor captures the potential difference between the anterior cornea and back retina of the eyeball, which is known as the Electrooculography (EOG) signals [1]. For the normal eye, the steady electric dipole passed between the cornea and retina is approximately 0.4 mV to 1 mV. Progressively, the quiescent current flows from the retina side to the cornea side. This constant flow induces an electrical field with a positive pole at the cornea and a negative pole at the retina The resting current flows progressively from the retina side to the cornea side, such that an electrical field comes into play with a negative pole at the retina and a positive pole at the cornea [2]. Electrooculography is beneficial in rehabilitation applications, which is not bulky and does not restrict patients' movements [3].

In this study, a deep learning model is developed to classify the four eye movements using the captured EOG signals. Recently, deep learning models such as convolutional neural networks (CNN) and recurrent neural networks (RNN) became essential tools in solving the classification issue in various fields. CNN reduces the data dimension to extract meaningful features depending on the number of filters applied to each convolution layer. CNN is considered as an assembly of multi-layer perceptron like other neural networks [4]. Fundamentally, most CNNs consist of three necessary layer components, namely convolution, pooling, and fully connected layer. Besides, LSTM is a type of recurrent neural network (RNN) that distinguishes itself by having at least one feedback loop when compared to a feedforward neural network. For instance, an RNN may comprise a single layer of neurons, with each neuron feeding its output signal back to the inputs of all of the other neurons. There are no self-feedback loops in the network within the system. Self- feedback is triggered when the neuron's output is fed back into its input [5].

This paper is arranged as follows: Section,II highlights the recent works related to EOG applications. Section III describes the data collection as well as the deep network architectures. Section IV includes the results and their discussion and SectionV conclude the paper.

II. Related work

Various electrode placement configurations have been studied in [6]. Lopez *et al.* investigated 4, 5, and 8 electrode placement configurations, and the study concluded that 5 and 8 electrodes configurations resulted in high accuracy and reliability signals in the horizontal plane compared to 4 electrode configuration. Moreover, blink signals are also best detected in the 5 and 8 electrodes configurations due to the appointing of one channel for the vertical axis. However, the person under the 5 and 8 electrodes configurations suffers some discomfort due to many electrode placements, which increases the number of blinks, resulting in heavy postprocessing steps to obtain a clearer signal. On the other hand, reducing the number of electrodes is preferred for application purposes, and 4 electrode configuration covers less area on the face.

Another work by [7] is manually triggered using polysomnographic (PSG) records. The CNN model is proposed for Electroencephalogram (EEG) and EOG signals classification. The architecture consists of two CNN branches with different filter sizes and stride values. Adam optimizer is used with a mini-batch size of 100 epochs. The overall accuracy of 84.13% is obtained.

Another paper used EOG signals on drowsiness detection compared to two approaches, using CNN and manual feature extraction. The manual method employed the discrete wavelet transform (DWT), which is used to extract frequency features from the EOG. A support vector machine (SVM) is trained. For CNN, two convolutional layers are used with 8 and 4 neurons for each layer where the acquired data is retrieved from 22 subjects. The CNN is pre-trained using stacked autoencoders. The results show that the CNN network performed better than the manual feature extraction algorithm with remarkably high accuracy. The feature extraction method is reported being redundant [8].

In another research related to EEG, LSTM is used to classify different types of emotions. The team used two LSTM layers, a dropout layer with a probability of 0.2, and a dense layer is used for classification. The first hidden layer (HL) consists of 64 neurons and the second hidden layer consists of 32 neurons. The activation function applied is ReLu having 4-fold cross-validation. RMSprop optimizer is used with a mini-batch size of 100. High accuracy of 85%-87% is achieved for every emotion classification. The same experiment is tested using feature extraction with the fast Fourier transform approach, and the accuracy of 55%-60% is achieved [9].

Finally, the last paper related to age and gender prediction from the EEG signal uses the LSTM network for classification. This model uses two LSTM hidden layers consisting of 128 and 64 neurons for each hidden layer. Batch normalization layer and a denser layer are applied, consisting of 6 neurons for age prediction. The alpha and the beta waves produced a high accuracy of 89.5% and 88.5% for age prediction. Furthermore, gender prediction accuracy achieved is 95.67%, and 96.35% for both alpha and beta waves [10].

III. Methodolgy

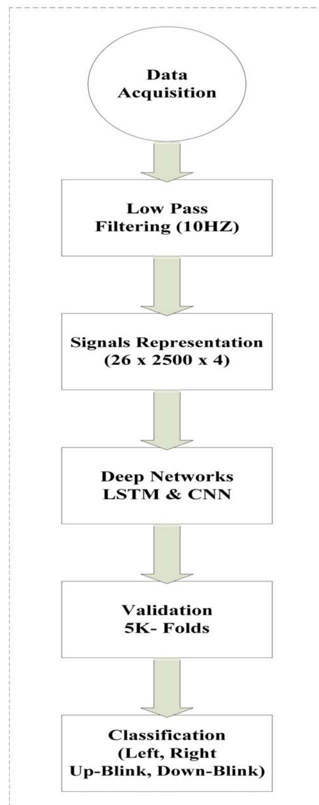


Fig. 1. Methodology pipeline.

This section outlines the steps used to conduct the experiments, data acquisition, and the results analysis.

A. Experimental Setup

In order to detect the eye movements, horizontal electrodes are positioned at the outer canthi for detecting horizontal movement, which is the left (Ch1) and right (Ch2). Eye movement in vertical direction is recorded via Ch3 and Ch

4 that are positioned, respectively up and down of the left eye. The position of the four channels are shown in Fig. 2. In addition to the four channels for eye movements a fifth channel is used as reference with electrode placed one the ear lobe.

B. Data Acquisition

The eye movements that are measured in this experiment are the up-blink, down-blink, left, and right eye movement. The subject is asked to look at a point positioned at the center of the field of view then to move his/her eyes to the maximum right and stay there for three seconds before coming back to the center. The subject is then instructed to rest for one minute before repeating the same experiment again. The same procedure is implemented for the left, up and down movements and 26 trials are collected for each type of eye movements. Eye blink is added to up/down movements in order to emphasize more on this type of eye movement. Fig. 3 shows the different EOG patterns of eye movement. The signals are recorded with a sampling frequency of 500 Hz.

The total size of the data is 120 eye blinks, in which data is imbalanced. the number of blinks for the left, right, up-blink, and down-blink classes is 31, 29, 34, and 26, respectively. Long short-term memory (LSTM) and convolution neural network (CNN) are developed for the classification of the four eye movements using the acquired EOG signals. The study is trained with 5 fold cross-validation. Fig. 1 depicts the study pipeline.

Fig. 2. Electrode Placement.

C. Data evaluation

To evaluate the LSTM networks as well as CNN networks, five metrics are used namely, sensitivity, specificity, precision and accuracy.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

TP denotes the number of true-positive signals, FP is the false-positive signals, TN is the true-negative signals, and FN is the false-negative signals.

IV. Results and Discussion

A. Long short-term memory testing result

- 1 Hidden Layer: based on Table IV, the performance of the LSTM model is shown using one hidden layer only. Various numbers of epochs are examined, and 5 values of hidden units (HU) are imposed per layer. Compared to the overall testing accuracy, the network trained with 250 hidden units and 50 epochs achieved the highest accuracy of 86.7%. Also, it is noticed that the higher the number of hidden units per layer, the higher the accuracy. On the other hand, hidden layers depend on the data size fitted to the deep neural network, and increasing the number of hidden layers could lead to data over-fitting. Hence, increasing the number does not guarantee better performance of the network. Table I shows the confusion matrix of the best models from both CNN and LSTM architectures, in which the results are demonstrated using the average of the 5 cross validation method. Table II shows that the right direction movement is classified with the highest precision of 93.1%. Moreover, the highest sensitivity (recognition rate) is obtained for the down-blink class with 92%. The overall testing accuracy is 86.7%. LSTM 1 network in Table II is referred to the LSTM network with one hidden layer.
- 2 Hidden Layers: based on Table V, the performance of the LSTM model is depicted. Two hidden layers are applied to the network with 3 different epochs values. Besides, 3 different combinations of hidden unit values are also tested in the first layer (L1) and the second layer (L2). Out of all the trials, the LSTM network trained with 80 epochs having 150 hidden layers at the first layer, and 125 hidden units at the second layer achieved the highest overall testing accuracy as of 88.33%. LSTM 2 network in Table II is referred to the LSTM network with two hidden layers.

B. Convolutional Neural Network

Datasets are tested with different filter sizes and several filter numbers. The best parameters are stated in Table III. This network consists of a mini-batch size of 96, 3 convolution layers, 3 ReLu layers, 3 batch normalization layers, 2 max-pooling layers, and 1 fully connected layer. The network is trained with 500 epochs. From the confusion matrix in Table I, the output of the class right scored the highest precision of 96.55%, and the highest sensitivity (recognition rate) is achieved by the up-blink class with 92.31%. The overall testing accuracy is 90.8%. Table II shows the performance evaluation of the LSTM network and CNN network in all metrics.

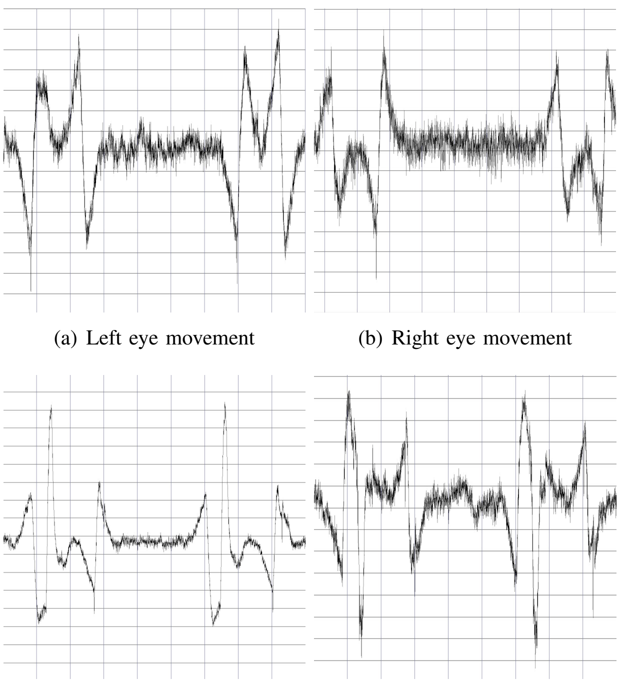
TABLE I
CONFUSION MATRIX FOR BEST PERFORMING NETWORKS.

Channels	LSTM with 1 Hidden Layer				LSTM with 2 Hidden Layers				CNN			
	Left	Right	Up-Blink	Down-Blink	Left	Right	Up-Blink	Down-Blink	Left	Right	Up-Blink	Down-Blink
Left	25	1	2	0	28	0	2	2	29	0	4	1
Right	5	27	2	0	1	27	2	2	2	28	1	1
Up-Blink	1	0	29	3	0	1	29	0	0	0	28	0
Down-Blink	0	1	1	23	2	1	1	22	0	1	1	24

TABLE II
PERFORMANCE EVALUATION.

Channels	Specificity(%)			Sensitivity(%)			Precision(%)			Accuracy(%)		
	CNN	LSTM 1	LSTM 2	CNN	LSTM 1	LSTM 2	CNN	LSTM 1	LSTM 2	CNN	LSTM 1	LSTM 2
Left	97.56	92.94	96.3	85.29	89.29	87.5	93.55	80.65	90.32	93.97	92.04	93.81
Right	98.78	97.47	97.53	87.50	79.41	84.4	96.55	93.10	93.10	95.61	92.04	93.81
Up-Blink	93.10	93.75	93.9	100	87.88	96.67	82.35	85.29	85.29	94.78	92.04	94.64
Down-Blink	97.70	96.43	95.45	92.31	92.00	84.62	92.31	88.46	84.62	96.46	95.41	92.98

htp]



(a) Left eye movement (b) Right eye movement

(c) Up-blink eye movement (d) Down-blink eye movement

Fig. 3. Signal Patterns for every Eye movement.

TABLE III
CNN FILTER SIZE AND NUMBER.

	Filter size	Filter Number
Convolution Layer 1	7x7	8
Convolution Layer 2	5x5	32
Convolution Layer 3	3x3	64

V. CONCLUSION

In this paper, two DNN models based on CNN and LSTM architectures are developed for the accurate classification of four eye movements using EOG signals. The results clearly show the capability of the proposed DNN models in solving the issue of eye movement classification. Furthermore, the results demonstrate the relatively better accuracy shown by the CNN model compared to the LSTM model. Besides, the results show that the LSTM model with two hidden layers is performing better, in terms of accuracy, than a shallow neural network of one hidden layer. The results also show that a CNN network that consists of 3 convolution layers with larger filter sizes and filter numbers produces better performance.

TABLE IV
PERFORMANCE OF THE LSTM MODEL USING 1 HIDDEN LAYER.

No. HL	Epochs	HU per layer	Overall Accuracy(%)
1	50	100	69
1	50	200	70
1	50	250	86.7
1	80	150	80
1	80	200	83.3
1	80	250	71
1	100	150	76
1	100	200	76
1	100	250	73

ACKNOWLEDGMENT

This research is supported by two research grants:(1) Ministry of Education Malaysia under Higher Institutional Centre of Excellence (HICoE) Scheme awarded to CISIR under grant number 015MA0-050(6) and (2) Yayasan Universiti Teknologi PETRONAS under grant number YUTP-FRG 015LC0-031

REFERENCES

[1] M. Hashimoto, K. Takahashi, and M. Shimada, "Wheelchair control using an EOG-and EMG-based gesture interface," in *2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 2009, pp. 1212–1217.

TABLE V
PERFORMANCE OF THE LSTM MODEL USING 2 HIDDEN LAYERS.

No. HL	Epochs	HU per layer		Overall Accuracy(%)
		L1	L2	
2	50	100	80	85
2	50	125	80	85
2	50	150	125	84
2	80	125	80	87.5
2	80	150	125	88.33
2	80	150	60	85.83
2	100	125	80	87
2	100	125	60	85
2	100	150	80	85

[2] "Extraocular muscle actions: Eye movements, rectus muscles, oblique muscles, (https://emedicine.medscape.com/article/1189759-overview)," accessed on 16 July 2019.

[3] M. S. Hossain, K. Huda, S. S. Rahman, and M. Ahmad, "Implementation of an EOG based security system by analyzing eye movement patterns," in *2015 International Conference on Advances in Electrical Engineering (ICAEE)*. IEEE, 2015, pp. 149–152.

[4] K. Yamagishi, J. Hori, and M. Miyakawa, "Development of EOG-based communication system controlled by eight-directional eye movements," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 2574–2577.

[5] S. Haykin, *Neural Networks and Learning Machines, 3/E*. Pearson Education India, 2010.

[6] A. Lopez, F. J. Ferrero, M. Valledor, J. C. Campo, and O. Postolache, "A study on electrode placement in EOG systems for medical applications," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2016, pp. 1–5.

[7] S. I. Niroshana, X. Zhu, Y. Chen, and W. Chen, "Automatic sleep stage classification based on convolutional neural networks," in *2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE, 2019, pp. 275–276.

[8] X. Zhu, W.-L. Zheng, B.-L. Lu, X. Chen, S. Chen, and C. Wang, "EOG-based drowsiness detection using convolutional neural networks," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 128–134.

[9] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.

[10] P. Kaushik, A. Gupta, P. P. Roy, and D. P. Dogra, "EEG-based age and gender prediction using deep BLSTM-LSTM network model," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2634–2641, 2018.