Enhancing Assistive Communication: An EOG-Based Continuous Eye-Writing Recognition Deep Learning Model

J. Dai[1], Y. Zhao[1], R. Yi[1], and J. Su[2]

1. Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD, USA
2. Department of Data Science, The Johns Hopkins University, Baltimore, MD, USA
jdai15@jh.edu, yzhao155@jh.edu, ryi5@jhu.edu, jsu35@jhu.edu

1 Introduction

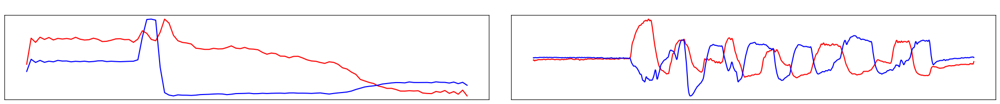
Modern human-computer interfaces (HCI) are crucial for individuals facing severe neurological diseases such as amyotrophic lateral sclerosis, brainstem stroke, and multiple sclerosis [1]. In many cases where voluntary movements are restricted to eye control, eye movements are critical for them to perform different tasks, like writing and communication[2][3]. Therefore, HCI systems that capture their gaze direction are pivotal in enhancing their communication and have a better quality of life.

Eye movement can be inferred through electrooculography (EOG) signals, which detects the potential difference created by the positive charge of the cornea and the negative charge of the retina. This potential difference can be recorded by electrodes placed around the eyes [4]. Numerous studies have developed methods that decode EOG signals for eye-writing. Lee et al.[5] employed waveform-matching techniques, while Fang et al.[6] used an HMM model with n-gram for Japanese Katakana character prediction. Multi-stage convolutional neural networks (CNNs) [7] and spiking neural networks (SNNs) [8] have also been applied for word classification based on EOG signals.

Existing eye-writing systems still encounter challenges in accurately translating EOG signals into human-readable languages. In this context, inspired by the deep transfer learning 2D-CNN model [7], we introduced a novel 1D CNN-LSTM deep-learning model with parallel branches. This model proves effective in decoding EOG signals for voice-free communication. Notably, with a few adjustments, this versatile model structure can function as both a word classifier, providing the probability of the word class, and a stroke labeler, offering the probability of the stroke sequence.

2 Dataset

The dataset is a Japanese eye-writing database of EOG signals from 6 health participants with a sample rate of 1.0 KHz[6]. It contains vertical and horizontal channels EOG data for 12 basic strokes, comprising 720 samples recorded 10 times by each participant, and EOG data for 150 words, with 4500 samples recorded 5 times by each participant, as illustrated in Figure 1(a) and Figure 1(b). Stroke EOG signals were downsampled to 100 sample length, removing the silent part at the beginning. Word EOG signals were downsampled to 1024 sample length.



(a) Stroke EOG Signal

(b) Word EOG Signal

Figure 1. EOG Signal
Notes: Color red - vertical channel; Color blue - horizontal channel

979-8-3503-4125-6/23/$31.00 ©2023 IEEE IEEE SPMB 2023 December 1, 2023

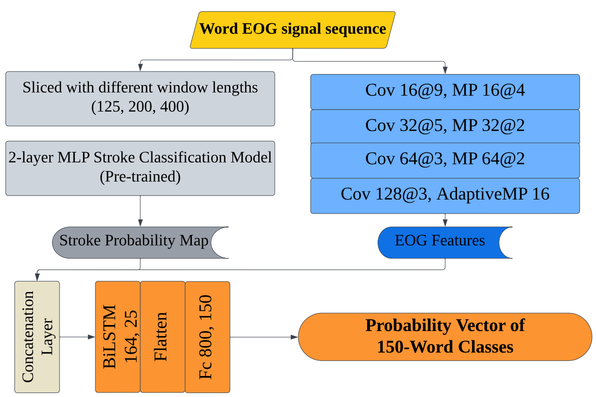
3 Methodology

3.1 Model Structure

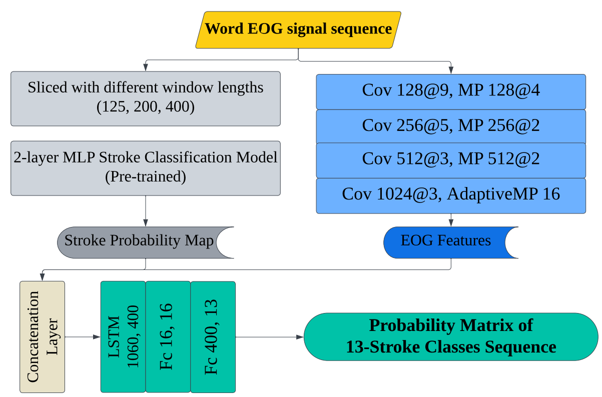
The proposed word classifier and stroke sequence labeler both utilize the same 1D CNN-LSTM deep-learning model structure, as depicted in figure 2. This model structure consists of 2 branches: the pre-trained branch and the EOG feature extraction branch.

The pre-trained branch calculates the stroke probability map. The input word EOG signal is initially segmented using a moving window with sizes of 128, 200, and 400. Each window moves equally to generate 16 segments, which are used to compute 12-stroke probability separately, as shown in figure 3. The pre-trained stroke classification model is a 2-layer multi-layer perceptron that has been trained with EOG stroke data.

The EOG feature extraction branch consists of 4 1D-convolution layers that take the original word EOG signal as input, maintaining the order of EOG features in the time domain. Then, the channels are maxpooled to 16 to match the size of the stroke probability map. The stroke probability map and EOG features are concatenated in the time domain. This is followed by a BiLSTM layer for the classification task and an LSTM layer for the labeling task, as shown in Figure 2(a) and Figure 2(b).



(a) Word Classifier



(b) Stroke Sequence Labeler

Figure 2. Model Structure
Notes: Cov - 1D-convolution layer; MP - maxpooling layer; Fc - fully connected layer.
Color grey - pre-trained stroke class probability generator; Color blue - EOG feature extractor;
Color orange - word classification part; Color green - stroke sequence labeling part.

After the stroke sequence labeler, a stroke-based n-gram model was applied. The model predicts items based on the context of the preceding n strokes, assigning probability scores to each class according to the language corpus distribution [9]. The probability $P(s_i|s_{i-1},\ldots,s_{i-n+1})$ is maximized with a greedy search.

3.2 Experimental Setup

We employed 5-fold cross-validation for the stroke classification model. The word classifier and stroke sequence labeler underwent evaluation using a leave-one-trial-out strategy repeated 3 times. We trained models for 500 epochs with an Adam optimizer, a 128 batch size, and a decreasing learning rate. We determined the models' parameters and some hyperparameters by the average performance in the validation set. Data augmentation included the addition of white noise.

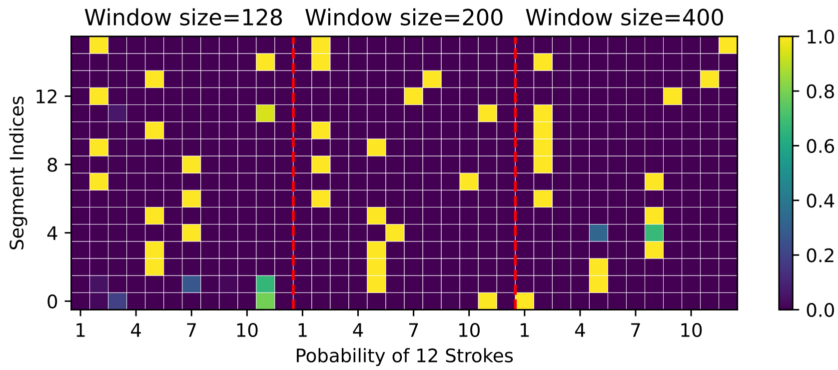


Figure 3. Stroke Probability Map

Table 1. Performance comparison with the 1D CNN-LSTM word classifier and previous models

	Word Classification(%)			Parameter Size
	Accuracy	Precision	Recall	
Boost CNN-word	92.04[7]	88.51[7]	91.98[7]	1.6M[7]
SNN	91.6[8]			158K[8]
1D CNN-LSTM	95.52	96.09	95.52	192K

3.3 Metrics

The classification task is evaluated by accuracy, precision, and recall[10]. The labeling task is evaluated by the Levenshtein ratio. The ratio ranges from 0 to 1, with a higher value indicating greater similarity between two sequences.

4 Results and Discussion

Table 1 displays the results of our 1D CNN-LSTM word classification model, averaged over each trial. This model performs exceptionally well despite its small parameter size. Our approach treats EOG data as a 1D sequence, which is more suitable than using a 2D CNN or SNN. EOG data is rich in time-domain features, and the 1D model excels in extracting these signal features. Additionally, the pre-trained stroke classification model achieves an impressive accuracy of 86% on the stroke EOG signal.

Table 2 shows the results of the stroke sequence labeler. With the help of the 3-gram model, the Levenshtein ratio improves to 0.973. Out of the 70 characters that appear in the 150 words, 91.43% of the characters consist of no more than 3 strokes. Therefore, the best 3-gram model considers more than one previous character, which helps improve the performance the most.

Compared to the two models, the stroke sequence labeler incorporates a larger number of filters. This suggests that labeling a sequence is a more complex task, requiring more features to represent EOG.

Table 2. Performance summary of the 1D CNN-LSTM stroke labeler

	Levenshtein ratio			
	1-gram	2-gram	3-gram	4-gram
1D CNN-LSTM	0.9719	0.9719	0.973	0.9659

5 Conclusions and Future work

We introduced a parallel-branch 1D CNN-LSTM structure that excels in both word classification and stroke sequence labeling tasks while maintaining a small model parameter size. Its impressive performance in stroke sequence labeling suggests potential for online communication without the limitations of language corpora, which could be used for individuals facing some diseases or in a VR environment.

However, this study has limitations due to the restricted data size and input speed. Future work may include expanding the corpus to different languages using basic EOG strokes and enhancing writing speed. Furthermore, the development of user-friendly EOG hardware is an exciting avenue to reduce eye burden.

Acknowledgements

This research would not have been possible without the assistance of Dr. Laureano Moro-Velazquez, Ph.D., Associate Professor in the Department of Electrical and Computer Engineering at Johns Hopkins University. He has provided invaluable guidance and instruction throughout the duration of this research project.

References

[1] C-T. Lin, W-L. Jiang, S-F. Chen, K-C. Huang, and L-D. Liao, "Design of a Wearable Eye-Movement Detection System Based on Electrooculography Signals and Its Experimental Validation," *Biosensors*, vol. 11, no. 9, p. 343, Sept. 2021.

[2] R. Barea, L. Boquete, S. Ortega, E. López, and J.M. Rodríguez-Ascariz, "EOG-based eye movements codification for human computer interaction," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2677–2683, Feb. 2012.

[3] R. Jacob, "Eye movement-based human-computer interaction techniques: Toward non-command interfaces," *Advances in human-computer interaction*, vol. 4, pp. 151–190, 1993.

[4] W-D. Chang, "Electrooculograms for Human-Computer Interaction: A Review," *Sensors*, vol. 19, no. 12, p. 2690, June 2019.

[5] K-R. Lee, W-D. Chang, S. Kim, and C-H. Im, "Real-Time "Eye-Writing" Recognition Using Electrooculogram," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 1, pp. 37–48, Jan. 2017.

[6] F-M. Fang and T. Shinozaki, "Electrooculography-based continuous eye-writing recognition system for efficient assistive communication systems," *PloS one*, vol. 13, no. 2, p. e0192684, Feb. 2018.

[7] J-D. Zou and Q-X. Zhang, "eyeSay: Make Eyes Speak for ALS Patients with Deep Transfer Learning-empowered Wearable," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, Nov. 2021, pp. 377–381.

[8] J. Stauffer and Q-X. Zhang, "SpikeBASE: Spiking Neural Learning Algorithm With Backward Adaptation of Synaptic Efflux," *IEEE Transactions on Computers*, vol. 71, no. 11, pp. 2707–2716, Nov. 2022.

[9] P. Brown, V. Della Pietra, P. Desouza, J. Lai, and R. Mercer, "Class-based n-gram models of natural language," *Computational linguistics*, vol. 18, no. 4, pp. 467–480, Dec. 1992.

[10] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, no. 8, pp. 861–874, June 2006.

Enhancing Assistive Communication: An EOG-Based Continuous Eye-Writing Recognition Deep Learning Model

J. Dai[1], Y. Zhao[1], R. Yi[1], and J. Su[2]

[1]Department of Biomedical Engineering, The Johns Hopkins University [2]Deparment of Data Science, The Johns Hopkins University

Abstract

- The dataset used is recorded Electrooculography(EOG) signal when eye-writing Japanese strokes and words.
- A paralleled-branch 1D CNN-LSTM structure is presented to translate the EOG signal.
- The model structure can work both as a word classifier and a stroke sequence labeler with a few parameter modifications.
- Superior performance can be attained in comparison to prior works on this dataset with a smaller parameter size.

Introduction

- Human-computer interface (HCI) has the potential to revolutionize communication accessibility for people with limited mobility to communicate effectively.
- Eye movement can be inferred through electrooculography (EOG) signals, which detects the potential difference created by the positive charge of the cornea and the negative charge of the retina.
- Eye writing system uses EOG as the input signals facilitate the recognition purposes for patients who are familiar with handwriting
- Fang et al.[1] proposed an HMM model with n-gram to predict 70 Japanese Katakana characters with EOG.
- 2D CNN and SNN were utilized for word classification bases on EOG signal[2][3].
- We proposed a paralleled-branch 1D CNN-LSTM structure that archives an 95.52% accuracy on the word classification task and a 0.973 Levenshtein ratio on stroke sequence labeling task.

Dataset

- The dataset is a Japanese eye-writing database of EOG signals from 6 health participants with a sample rate of 1.0 KHz [1].
- Japanese words are made of characters, and characters are made of strokes.

ノイ → イ → イジル

- The dataset contains vertical and horizontal channels EOG data for 12 basic strokes, comprising 720 samples recorded 10 times by each participant, and EOG data for 150 words, with 4500 samples recorded 5 times by each participant. The EOG signal exmaple is shown in figure 1.
- The continuous data includes 150 Japanese words from the Corpus of Spontaneous Japanese. The words are averagely 2.8-Katakana-character long, each repeatedly recorded 3 to 8 times by each participant.

Data Prepocessing

- For stroke classification task, we removed silent segments at the beginning of the stroke EOG signal, resampled data to 100 data points, and normalized the data.
- For the word classification task and stroke prediction task, we resampled the word EOG signal to 1024 data points and normalized the data.

Methodology

Overall of the Model

- The proposed word classifier and stroke sequence labeler both utilize the same 1D CNN-LSTM deep-learning model structure, as depicted in figure 2.
- The model structure consists of 2 branches: the pre-trained branch and the EOG feature extraction branch.

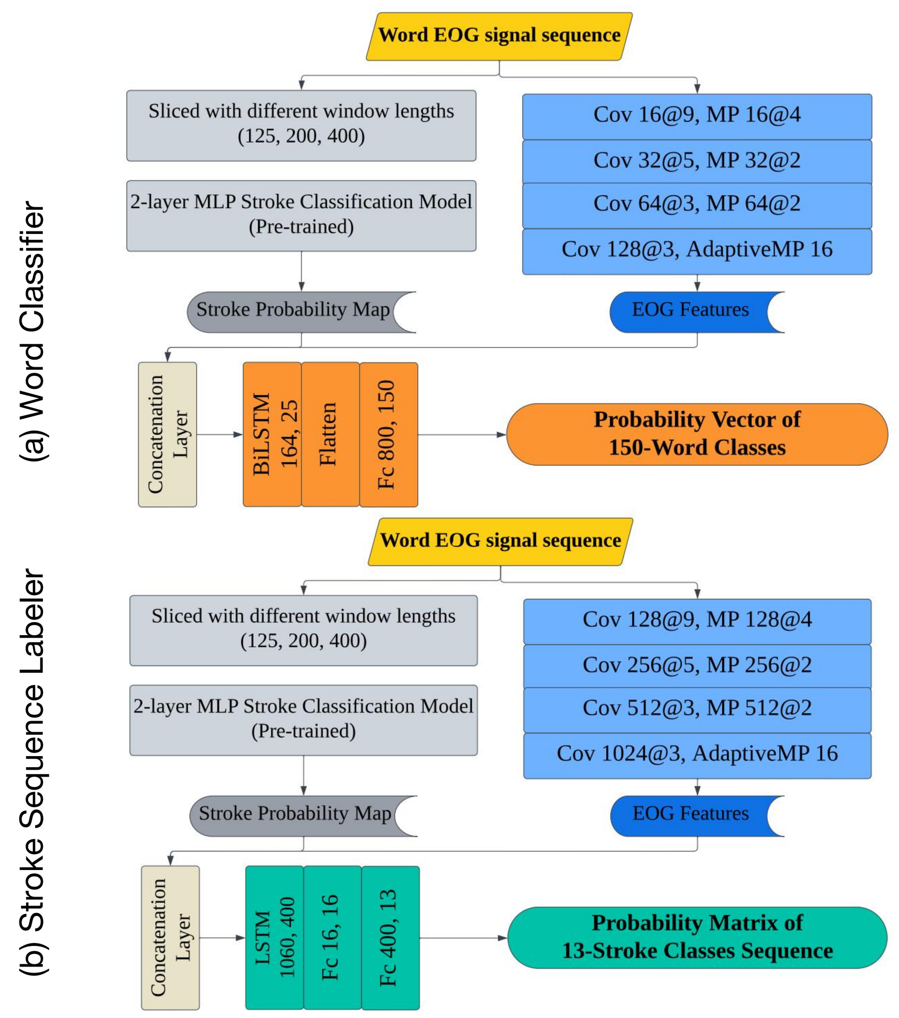


Figure 2. Model Structure
Notes: Color grey - pre-trained stroke class probability generator;
Color blue - EOG feature extractor;
Color orange - word classification part;
Color green - stroke sequence labeling part.
Cov - 1D-convolution layer; MP - maxpooling layer; Fc - fully connected layer.

The Pre-trained Branch

- This branch calculates the stroke probability map.
- The input word EOG signal is initially segmented using a moving window with sizes of 128, 200, and 400. Each window moves equally to generate 16 segments, which are used to compute 12-stroke probability separately, as shown in figure 3.
- The pre-trained stroke classification model is a 2-layer multi-layer perceptron that has been trained with EOG stroke data.

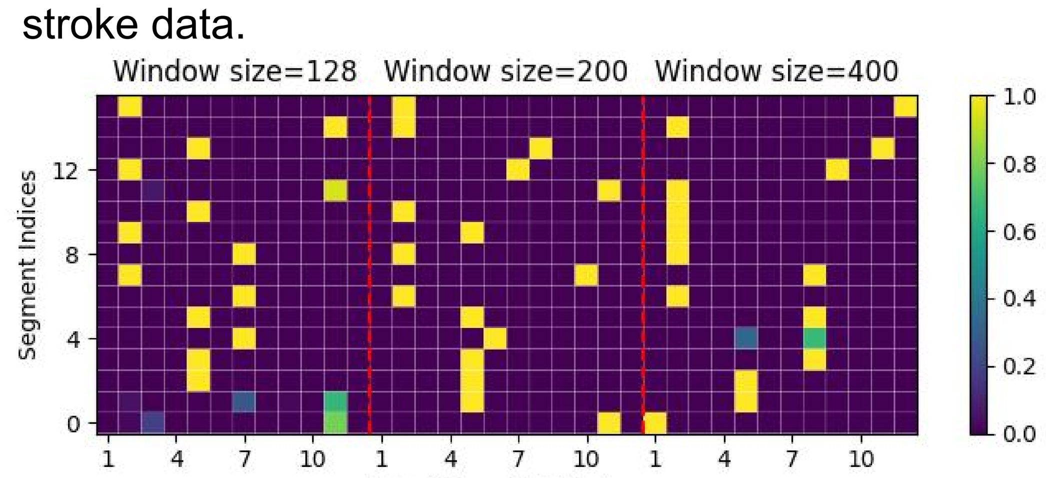
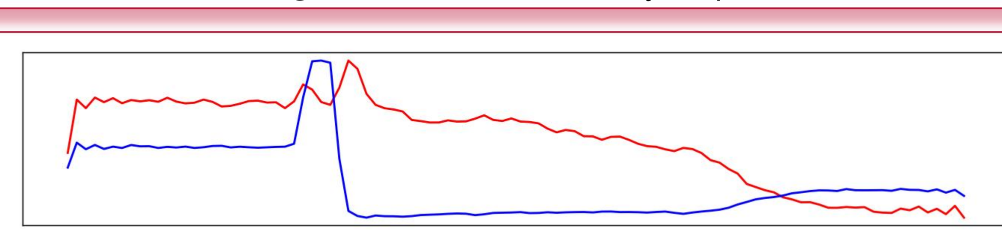
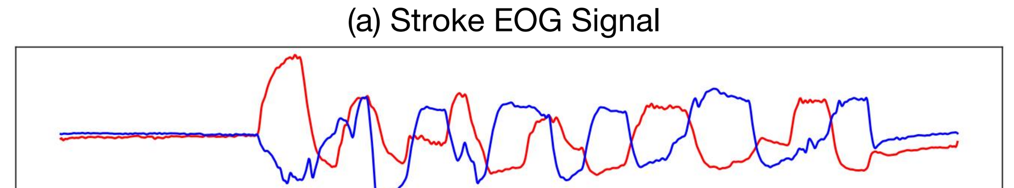


Figure 3. Stroke Prabability Map



(a) Stroke EOG Signal



(b) Word EOG Signal

Figure 1. EOG Signal
Notes: Color red - vertical channel; Color blue - horizontal channel

The EOG Features Extractor Branch

- The EOG feature extraction branch consists of 4 1D-convolution layers that take the original word EOG signal as input, maintaining the order of EOG features in the time domain.
- The channels are maxpooled to 16 to match the size of the stroke probability map.
- The stroke probability map and EOG features are concatenated in the time domain.
- The concatenation layer is followed by a BiLSTM layer for the classification task and an LSTM layer for the labeling task, as shown in Figure 2(a) and Figure 2(b).

The Stroke Based n-gram model

- The stroke-based n-gram model was applied after the stroke sequence labeler.
- The model predicts items based on the context of the preceding n strokes, assigning probability scores to each class according to the language corpus distribution.
- The probability $P(s_i|s_{i-1}, \dots, s_{i-n+1})$ is maximized with a greedy search.

Experimental Setup

- 5-fold cross-validation was employed for the stroke classification model.
- The word classifier and stroke sequence labeler underwent evaluation using a leave-one-trial-out strategy repeated 3 times.
- We trained models for 500 epochs with an Adam optimizer, a 128 batch size, and a decreasing learning rate.
- We determined the models' parameters and some hyperparameters by the average performance in the validation set.
- Data augmentation included the addition of white noise.

Metrics

- Classification Task:

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad Precision = \frac{TP}{TP+FP} \quad Recall = \frac{TP}{TP+FN}$$

- Stroke Sequence Prediction task:

$$Levenshtein\ ratio = 1 - normalized\ Levenshtein\ distance$$

- Levenshtein distance is the minimum amount of insertions, deletions, and substitutions to change one sequence into the other.

Results and Discussion

Stroke Classification Model

- The 2-layer MLP stroke classification model, which is used to generate the stroke probability map, obtains an accuracy of 86% on the stroke EOG signal.

Word Classification Model

- Table 1 shows the result of our 1D CNN-LSTM word classification model on average for each trial.
- Our model performs exceptionally well despite its small parameter size.
- Our approach treats EOG data as a 1D sequence, which is more suitable than using a 2D CNN or SNN. EOG data is rich in time-domain features, and the 1D model excels in extracting these signal features.

	Word Classification(%)			
	Accuracy	Precision	Recall	Parameter Size
Boost CNN-word	92.04 [2]	88.51 [2]	91.98 [2]	1.6M [2]
SNN	91.6 [3]			158K [3]
1D CNN-LSTM	95.52	96.09	95.52	192K

Table 1. Performance comparison with the 1D CNN-LSTM word classifier and previous models

Stroke Sequence Labeling Model

- Table 2 shows the results of the stroke sequence labeler.
- The labeler achieves 0.9719 Levenshtein ratio without considering previous strokes.
- With the help of the 3-gram model, the Levenshtein ratio improves to 0.973.
- Out of the 70 characters that appear in the 150 words, 91.43% of the characters consist of no more than 3 strokes.
- The best 3-gram model considers more than one previous character, which helps improve the performance the most.

	Stroke Sequence Prediction(Levenshtein ratio)			
	1-gram	2-gram	3-gram	4-gram
1D CNN-LSTM	0.9719	0.9719	0.973	0.9659

Table 2. Performance summary of the 1D CNN-LSTM stroke labeler

Conclusions and Future Work

- We introduced a parallel-branch 1D CNN-LSTM structure that excels in both word classification and stroke sequence labeling tasks while maintaining a small model parameter size.
- The impressive performance in stroke sequence labeling suggests potential for online communication without the limitations of language corpora, which could be used for individuals facing some diseases or in a VR environment.
- This study also has limitations due to the restricted data size and input speed.
- Future work may include expanding the corpus to different languages using basic EOG strokes and enhancing writing speed.
- The development of user-friendly EOG hardware is also an exciting avenue to reduce eye burden.

Acknowledgements

This research would not have been possible without the assistance of Dr. Laureano Moro-Velazquez, Ph.D., Associate Professor in the Department of Electrical and Computer Engineering at Johns Hopkins University. He has provided invaluable guidance and instruction throughout the duration of this research project.

Reference

[1] F-M. Fang and T. Shinozaki, "Electrooculography-based continuous eye-writing recognition system for efficient assistive communication systems," PloS one, vol. 13, no. 2, p. e0192684, Feb. 2018.

[2] J-D. Zou and Q-X. Zhang, "eyeSay: Make Eyes Speak for ALS Patients with Deep Transfer Learning-empowered Wearable," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Nov. 2021, pp. 377–381.

[3] J. Stauffer and Q-X. Zhang, "SpikeBASE: Spiking Neural Learning Algorithm With Backward Adaptation of Synaptic Efflux," IEEE Transactions on Computers, vol. 71, no. 11, pp. 2707–2716, Nov. 2022.