2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

eyeSay: Make Eyes Speak for ALS Patients with Deep Transfer Learning-empowered Wearable

Jiadao Zou, Qingxue Zhang, *IEEE Senior Member*

Purdue School of Engineering and Technology
Indiana University-Purdue University at Indianapolis

***Abstract*—Eye dynamics, a typical expression of brain activities, is an emerging modality for emerging and promising smart health applications. Electrooculogram (EOG) – a natural bio-electric signal generated during eye movements, if decoded, is of great potential to reveal the user's mind and enable voice-free communication for patients with amyotrophic lateral sclerosis (ALS). ALS patients usually lose physical movement abilities including speech and handwriting but fortunately can move their eyes. In this study, we propose a novel deep transfer learning-empowered system, called "eyeSay", which leverages both deep learning and transfer learning for intelligent eye EOG-to-speech translation. More specifically, we have designed a multi-stage convolutional neural network (CNN) to analyze the eye-written words, named as CNN-word. Moreover, to reveal fundamental patterns of eye movements, we build a transferable feature extractor, CNN-stroke, upon eye strokes that are building components of an eye word. Then, we transfer the CNN-stroke model to the eye word learning task in an innovative way, that is, use CNN-stroke as an additional branch of CNN-word to generate a stroke probability map. The achieved boostCNN-word model, enhanced by the transferable feature extractor, has greatly improved the eye word decoding performance. This novel study will directly contribute to voice-free communications for ALS patients, and greatly advance the ubiquitous eye EOG-based smart health area.**

***Index Terms*—Smart Health, Deep Learning, Transfer Learning, Electrooculography, Amyotrophic Lateral Sclerosis.**

I. INTRODUCTION

Empowered by advancements in electronics, signal processing and artificial intelligence (AI), smart health is igniting next-generation healthcare platforms that are intelligent, light weighted, easily accessible, user friendly, cost effective and fully extensible [1-3]. Smart health wearable, as a booting paradigm, brings us many irreplaceable advantages. In this study, we take a special interest in Electrooculogram (EOG)-based eye decoding applications [4]. EOG is an electrical potential difference between the cornea and the ocular fundus. The around-eye electric field changes during different eyeballs' movements, and if we can decode these EOG dynamics, we will be able to "hear the sound" of the users if they intentionally speak, without needing any other traditional communication means

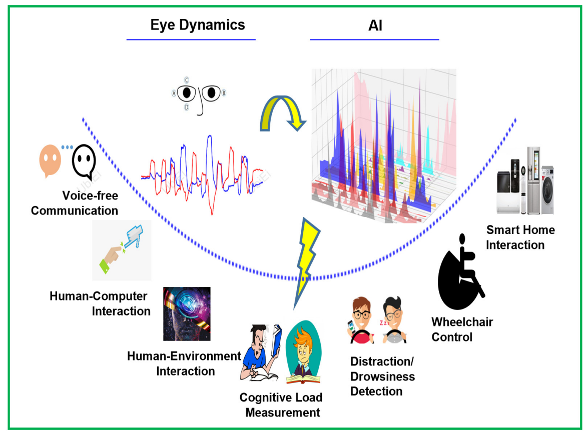

Fig. 1 A big picture of how eye dynamics decoding can advance many smart health and human-X interactions. In this study, we take special interest in leveraging AI to analyze, learn, and decode the eye dynamics for voice-free communications for ALS patients. This study will also greatly advance various applications illustrated here.

Specifically, we investigate how to decode eye EOG signals in the context of voice-free communication for amyotrophic lateral sclerosis (ALS) patients [5, 6]. Compared with video-based methods, eye EOG decoding is insensitive to environmental light changes and has better privacy protection. Besides that, the proposed approach can be generalized to other smart health applications like human-computer interaction, attention tracking, wheelchair control, cognitive load measurement, driver distraction and drowsiness detection, virtual reality, and augmented reality [7-9] (Fig. 1).

ALS is an idiopathic, fatal disease that makes the patients suffer from the degeneration of motor neurons and loss of physical capability. Worldwide, about 450,000 people have ALS and every 90 minutes, there will be one diagnosed as ALS patient [10]. Lots of severely-ill patients become incapacitated and lose the ability to move, grasp, and speak. Research shows ALS patients can still hear, smell as normal and their senses of taste and sight remain unaffected. Especially, the fact that ALS would not hurt muscle of the eyes makes the eye controlling system extremely helpful to those people. Therefore, researchers have investigated approaches to allow the ALS patients to communicate with eye movements, yielding the so-called eye-writing systems.

Jiadao Zou (Ph.D. Student) and Qingxue Zhang (Assistant Professor, qxzhang@purdue.edu) are with Department of Electrical and Computer Engineering, Purdue School of Engineering and Technology, on the joint Campus of Purdue University and Indiana University, Indianapolis, Indiana, USA. This study is supported by USA National Science Foundation Career Award Grant.

Previous studies on eye-writing can be categorized into two groups: keyboard typing and handwritten-style writing. In the former one, Kate et al. developed a screen-based eye-typing system with a virtual keyboard [11], and Ward et al. developed a system called "Dasher' [12]. To deal with the limitation of the input speed of these methods, the second category, i.e., handwritten-style writing has gained more and more attention, which is moving eye along the trajectory of the word like writing a word using the hand. Though early works in this category usually require "dwell time" to separate multiple characters in a word that induce significant time waste [13, 14], most recent works now use methods like Hidden Markov Model to continuously analyze EOG for handwritten-style writing. However, the detection performance is still a big challenge, mainly because of the high diversity of the signal characteristics. For instance, even when a patient performs the same eye-writing operation, the eye trajectories are usually inconsistent. This is natural since the human body is a complex dynamic system and the eyeballs have a high freedom in movements. Another fact that worsens the situation is that a word usually owns multiple characters, and a character usually owns more than one strokes, which are the most fundamental elements of eye writing like moving leftward or upward. And the stroke-character-word conditions can be highly diverse. For example, the duration of each stroke/character, the character-to-character transition time, and the total time of the word, may all change from time to time. Therefore, we propose deep learning algorithms to tackle this dilemma [13, 14].

In this study, we propose a novel deep transfer learning framework that can robustly decode EOG eye-writing dynamics, thereby enabling voice-free communication for ALS patients. We have not only designed a deep convolutional neural network (CNN) for eye word learning, but also designed a transferable feature extractor that can learn more fundamental characteristics at the eye stroke level. The innovative framework, through combining deep learning and transfer learning, can effectively reveal the eye stroke-character-word structures under high intra-word variabilities. More specifically, our major contributions include:

1. Design a deep CNN model, called CNN-word, for automatic EOG-based word decoding, which can continuously recognize the patterns within an EOG word by learning stroke-to-character-to-word constructions.
2. Propose a transferable deep feature extractor, called CNN-stroke, which is pre-trained on basic EOG strokes and then concatenated with the CNN-word model, aiming to provide a stroke probability map for the given eye word. The achieved framework is named as boostCNN-word.
3. Evaluate and validate the proposed novel framework and demonstrate that boostCNN-word can provide an eye-writing detection accuracy much superior to previous studies, by maximizing pattern mining efficiency from scarce learning inputs using deep transfer learning. The framework, evaluated on one language now, can be generalized to other languages too.

II. Methods

A. System Overview

The proposed novel deep transfer learning framework (Fig. 2) has two parallel branches for pattern mining, followed by a fully connected neural network for eye word prediction. One branch (CNN-word) extracts the spatial patterns directly from eye EOG word signals. The other branch (CNN-stroke) is a pre-trained transferable feature extractor, which is trained on fundamental EOG stroke data and then transferred to EOG word decoding, thereby generating stroke probability maps for sliced stroke images. After fusing above learned patterns and feeding them to a fully connect neural network (FNN), the achieved boostCNN-word model is expected to greatly boost the eye word decoding performance. Major components of this system are detailed below.

B. EOG-word Deep Translator

Fig. 3 illustrates the EOG-word deep translator, which is composed of the CNN-word model for spatial EOG pattern learning and the FNN model for EOG word prediction [15]. These two models are also used in boostCNN-word model, where they are both trainable to order to coordinate with the non-trainable CNN-stroke model. CNN-word model includes

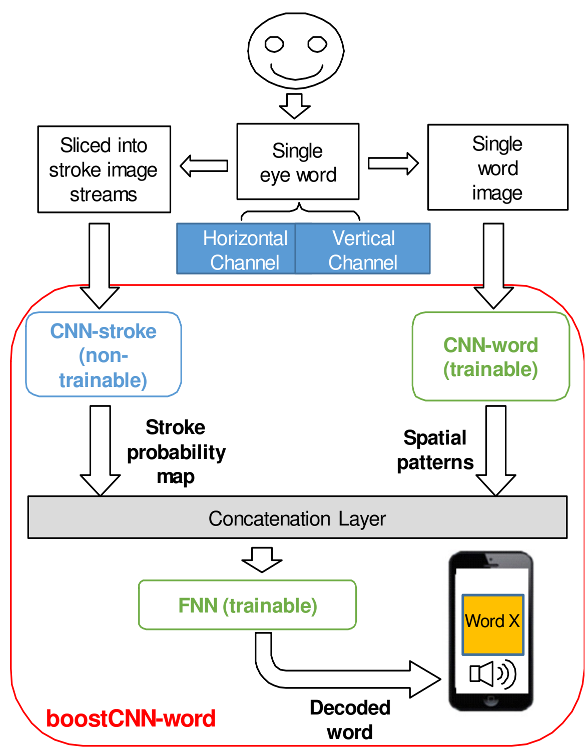

Fig. 2 Proposed deep transfer learning-empowered eye EOG decoding system, which owns two parallel feature abstraction branches: a CNN-word branch for word feature extraction, and a CNN-stroke branch for stroke-level probability map generation. Finally, a fully connected neural network fuses the output of these two branches and generates the final word prediction. Note: horizontal channel corresponds to bio-potential between location A & B in Fig. 1, and vertical channel corresponds to C & D.

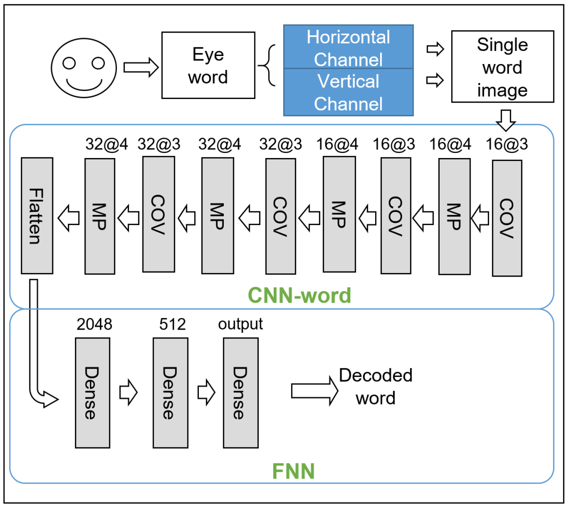

Fig. 3 CNN-based EOG translator for EOG word decoding: CNN-word for feature learning and FNN (fully connected neural network) for word prediction. Notes: COV – convolution.

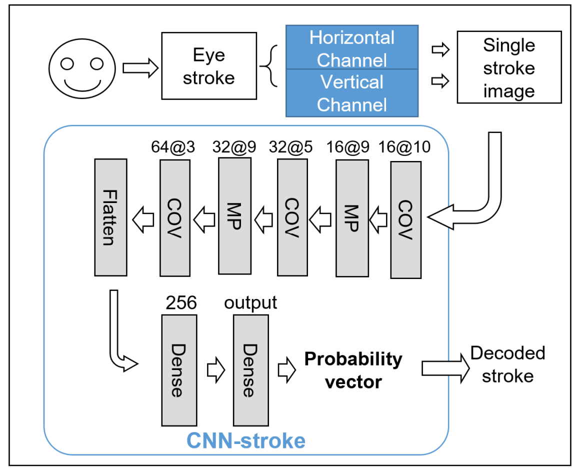

Fig. 4 Proposed CNN-based for EOG stroke decoding, CNN-stroke, which will later be concatenated with CNN-word as a pre-trained transferable feature extractor. Notes: COV – convolution.

multiple stages of convolutional layers (COV) and max pooling layers (MP), for spatial motif learning and dimension reduction, respectively [16]. The 2D-EOG image (horizontal channel and version channel) is fed into the deep translator in Fig. 3 for spatial pattern mining and eye word prediction. While we expect it can effectively predict the eye words, but further effort of leveraging transfer learning onto current classifier will be beneficial considering that the EOG data is scarce and the data collection effort on the user needs to be minimized.

C. EOG-stroke Transferable Deep Feature Extractor

A word-level eye-writing movement can be sliced into multiple continuous stroke-level eye-writing movements. Actually, a general stroke-to-character-to-word model can be applied to many languages. As the basic unit of constructing every single word, the number of strokes is dramatically less than the number of words. Therefore, one motivation is to leverage a small dataset of strokes to learn basic EOG patterns, and transfer what have been learned to more complex word recognition tasks as shown in Fig. 2.

Therefore, we propose to learn a transferable EOG feature extractor as shown in Fig. 4. The proposed CNN-stroke model includes multiple convolutional layers, max pooling operators, and fully connected layers, to learn the fundamental dynamics within the eye EOG strokes. Then output for a stroke slice is a probability vector with a dimension of 1-by-13, where 13 corresponds to 12 different strokes plus 1 null stroke (specific dataset information will be given in the results section).

This CNN-stroke model, after pre-trained on the small stroke dataset that will be detailed later, will have the capability to reveal stroke-level patterns of complex eye EOG signals. It will then be transferred to build the boostCNN-word model as given in Fig. 2, where the CNN-stroke model is a non-trainable feature extractor in parallel with the CNN-word branch.

D. Deep Transfer Learning-empowered eyeSpeak System

The transferrable CNN-stroke model in Fig. 2 helps minimize the training effort of the eye word learning model through revealing the fundamental patterns evolved in the eye writing movements. As detailed in the results section, for the Japanese language, if collecting 12 fundamental strokes for 10 times, that will be 120 recordings. But if collecting 150 words for 5 time, there will be 750 recordings, not mention if we consider like thousands of words. Evidently, collecting and learning fundamental eye strokes will effectively lower the data collection, model training, and user efforts.

The CNN-stroke model can analyze each eye word slice (segmented by a sliding window method) and give a probability vector corresponding to different eye stroke types. When combining vectors from all slices of an eye word, a probility map (as visualzied in Fig. 6 in the results section) will be generated and treated as a new data representation

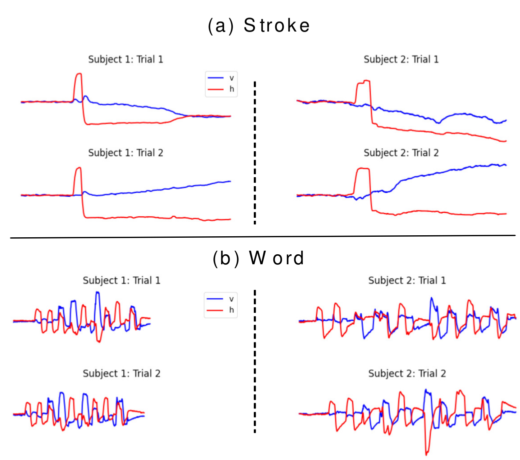

Fig. 5 For the same stroke (a) or word (b), the intra-subject inter-trial comparison, and inter-subject comparison both indicate high variabilities, which pose challenges to eye EOG decoding and necessitates advanced pattern learning algorithms. Notes on axis: horizontal is time and vertical is voltage.

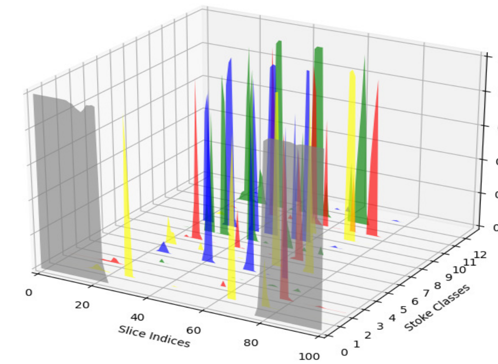

Fig. 6 The probability map is generated, where each slice corresponds to a signal segment in the eye EOG word and results in a 13-dimension probability vector (12 stroke classes plus 1 null class). This probability map is treated as an additional feature representation (Fig. 2), which, together with the features extracted by CNN-word, are fused by FNN to yield robust eye word predictions.

branch, thereby achieving the boostCNN-word model and boosting the eye word detection performance.

III. Results

A. Experimental Setup and Data Used

As mentioned above, to boost the eye word recognition performance, we propose to leverage the most fundamental eye movements – eye strokes to train a transferable feature extractor. Therefore, we have used a database that includes an EOG word dataset, and a small EOG stroke dataset. It is a Japanese eye writing database [17], recorded from six participants, each of which eye-wrote 150 words for 5 trials, and 12 strokes for 10 trials.

Here, for each participant, 120 stroke recording are used for CNN-stroke model training and testing, and 750 words for CNN-word and boostCNN-word training and testing. The leave-one-trial-out strategy is used to split the training and testing data for thorough evaluation purpose. Considering the randomness induced by deep learning training, we have repeated the evaluation by three times and reported the average performance.

B. Diversity of EOG Strokes and EOG Words

EOG is highly different among users and even among different trials of the same user. As shown in Fig. 5a, two trials of a same type of eye-writing task performed by two subjects are given, which indicates that the two-channel EOG-stroke signals are highly different even for the same subject. The inter-subject difference is much more obvious.

In Fig. 5b, the subjects have written a word that includes multiple strokes. We can find that firstly, two trials of an eye movement from the same subject are also very different. The inter-subject EOG word variability is much higher. The durations of the EOG words are different due to the different speed of eye movement of different users. Besides, the morphologies are also user-specific, meaning that the biological processes behind the eye dynamics also differ from each other. These all necessities advanced detection algorithms.

C. Transferable Feature Extractor Learning

To illustrate how the transferable feature extractor can facilitate the eye word decoding process, we have visualized the probability maps generated by the CNN-stroke model when feeding the sliced EOG word into it. As shown in Fig. 6, for a given slice which is selected from the eye EOG word using a sliding window method, a vector with a dimension of 13 are generated (12 stroke classes plus 1 null class). For all the slices of an EOG word, the corresponding probability vectors together make a probability map, which indicates the estimation of occurrence of each stroke in each moment (slice).

This probability map generation process is treated as another pattern abstraction branch in the boostCNN-word model. Together with CNN-word, the new framework can, not only directly learn eye EOG word patterns, but also reveal fundamental stroke patterns. Therefore, these parallel learning strategy can boost the eye writing decoding performance.

D. boostCNN-word vs. CNN-word

In Fig. 7, the boostCNN-word model is compared with the CNN-word model in terms of eye EOG word detection accuracy, precision and recall. Participant-wise comparison is given subfigures (a), (b) and (c), where the maximum improvements of accuracy, precision, and recall are 3.43%, 4.49% and 3.42%, respectively. This clearly indicates the transferable CNN-stroke feature extractor has boosted the eye EOG word recognition performance.

Another finding is that, for participants that have low CNN-word performance, the transfer learning approach can bring more performance improvement. The reason lies in the fact that the additional fundamental patterns extracted by CNN-stroke can greatly enrich the data representation, thereby allowing boostCNN-word to more robustly translate eye EOG words.

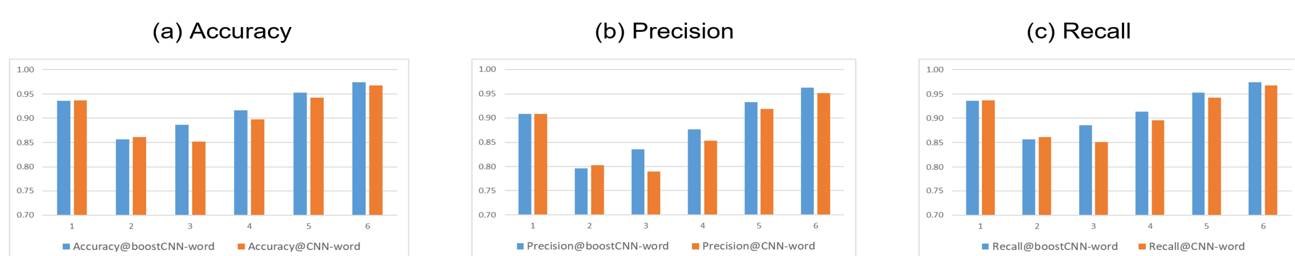

Fig. 7 The boostCNN-word model outperforms the CNN-word model in terms of eye EOG word detection accuracy, precision and recall. Participant-wise comparison is given subfigures (a), (b) and (c), where the maximum improvement of accuracy, precision, and recall are 3.43%, 4.49% and 3.42%, respectively. This clearly indicates the transferable CNN-stroke feature extractor has boosted the eye EOG word recognition performance.

D. Further Comparison

In addition to further summarize the performance of different deep models, we have also compared our study with previously reported Hidden Markov Model (HMM) [17] and Dynamic Time Warping (DTW) methods [17], given in Table I. The previous methods leveraged basic EOG features and expanded feature set, and reported the eye stroke recognition precision and recall up to 87.2% and 85.9%. Our CNN-stroke model improves them to be 89.88% and 92.64%, respectively.

Further, from Fig. 7, we have summarized the average performance across all the participants for word recognition, and the accuracy, precision, and recall are 92.04%, 88.51%, and 91.98%, respectively. The maximum improvements of criterion (3.43%, 4.49% and 3.42%) are also given in Table I. Overall, the proposed boostCNN-word deep learning framework, enhanced by transfer learning, can robustly recognize eye EOG words and will greatly advance the voice-free communication practices for ALS patients.

Table I: Performance summary of proposed deep models, and further comparison with previously reported HMM & DTW methods.

Approaches	Stroke Recognition (%)			Word Recognition (%)		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Max boosting				3.43	4.49	3.42
boostCNN-word				92.04	88.51	91.98
CNN-word				90.97	87.08	90.91
CNN-stroke	92.64	89.98	92.64			
HMM (basic features) [a]		85.00	84.20			
DTW (basic features) [a]		78.20	76.80			
HMM (expanded features) [a]		87.20	85.90			
DTW (expanded features) [a]		78.10	77.10			

Note: [a] – reference [17].

E. Future Work

We plan to further investigate the feature extractor for further performance enhancement, including analyzing the relationship between the incorrect word prediction and the probability vectors generated by the transferable feature extractor. We also plan to extend the study to other languages. These future works will further advance our understanding the potential of deep and transfer learning on intelligent EOG decoding.

And we are also interested in implementing the system on the edge devices like the smart phone for real-time eye decoding.

IV. Conclusion

In this study, we have proposed a deep transfer learning framework for intelligent EOG-based eye-writing recognition, which is crucial for the voice-free communicating system designed for ALS patients. It can also greatly advance other EOG-based smart health applications such as human-computer interaction, attention tracking, wheelchair control, cognitive load measurement, driver distraction and drowsiness detection, virtual reality, and augmented reality.

The proposed framework includes two parallel feature abstraction branches: a CNN-word branch for word-level feature extraction, and a CNN-stroke branch for stroke-level probability map generation. The CNN-stroke feature extractor is transferred from another learning task, i.e., eye stroke recognition, which can reveal fundamental patterns of eye movements at the stroke-level. The achieved boostCNN-word framework has greatly enhanced the eye word recognition performance compared with the CNN-word model. This novel study will directly contribute to voice-free communications for ALS patients, and greatly advance the ubiquitous eye EOG-based smart health area.

References

[1] M. M. Baig and H. Gholamhosseini, "Smart health monitoring systems: an overview of design and modeling," *Journal of medical systems,* vol. 37, no. 2, pp. 1-14, 2013.

[2] M. I. Pramanik, R. Y. Lau, H. Demirkan, and M. A. K. J. E. S. w. A. Azad, "Smart health: Big data enabled health paradigm within smart cities," vol. 87, pp. 370-383, 2017.

[3] A. Solanas *et al.*, "Smart health: a context-aware health paradigm within smart cities," *IEEE Communications Magazine,* vol. 52, no. 8, pp. 74-81, 2014.

[4] M. Merino, O. Rivera, I. Gómez, A. Molina, and E. Dorronzoro, "A method of EOG signal processing to detect the direction of eye movements," in *2010 First International Conference on Sensor Device Technologies and Applications*, 2010, pp. 100-105: IEEE.

[5] R. Del Bo *et al.*, "Absence of angiogenic genes modification in Italian ALS patients," *Neurobiology of aging,* vol. 29, no. 2, pp. 314-316, 2008.

[6] S. Garbuzova-Davis *et al.*, "Impaired blood–brain/spinal cord barrier in ALS patients," *Brain research,* vol. 1469, pp. 114-128, 2012.

[7] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human–computer interaction," *IET Computer Vision,* vol. 12, no. 1, pp. 3-15, 2017.

[8] L. Oliveira, J. S. Cardoso, A. Lourenço, and C. Ahlström, "Driver drowsiness detection: a comparison between intrusive and non-intrusive signal acquisition methods," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*, 2018, pp. 1-6: IEEE.

[9] J. Xiao, J. Qu, and Y. Li, "An Electrooculogram-based interaction method and its music-on-demand application in a virtual reality environment," *IEEE Access,* vol. 7, pp. 22059-22070, 2019.

[10] V. Neuro. (2020). *ALS Incidence*. Available: https://alstreatment.com/amyotrophic-lateral-sclerosis-incidence/

[11] J. Ten Kate, E. Frietman, F. Stoel, and W. Willems, "Eye-controlled communication aids," *Medical Progress through Technology,* vol. 8, no. 1, pp. 1-21, 1980.

[12] D. J. Ward and D. J. MacKay, "Fast hands-free writing by gaze direction," *Nature,* vol. 418, no. 6900, pp. 838-838, 2002.

[13] J. Tsai, C. Lee, C. Wu, J. Wu, and K. Kao, "A feasibility study of an eye-writing system based on electro-oculography," *Journal of Medical and Biological Engineering,* vol. 28, no. 1, p. 39, 2008.

[14] 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, 2016.

[15] J. Zou and Q. Zhang, "eyeSay: Eye Electrooculography Decoding with Deep Learning," in *The 39th International Conference on Consumer Electronics (ICCE 2021)*, 2020.

[16] Q. Zhang, "Deep Learning of Biomechanical Dynamics in Mobile Daily Activity and Fall Risk Monitoring," in *The 6th annual IEEE EMB Strategic Conference on Healthcare Innovations and Point-of-care Technologies (IEEE HI-POCT 2019)*, DC, USA, 2019.

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