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

Optimizing Input for Gesture Recognition using Convolutional Networks on HD-sEMG Instantaneous Images

Michael Houston, Albon Wu, and Yingchun Zhang, *Senior Member, IEEE*

***Abstract*— Hand gesture recognition using high-density surface electromyography (HD-sEMG) has gained increasing attention recently due its advantages of high spatio-temporal resolution. Convolutional neural networks (CNN) have also recently been implemented to learn the spatio-temporal features from the instantaneous samples of HD-sEMG signals. While the CNN itself learns the features from the input signal it has not been considered whether certain pre-processing techniques can further improve the classification accuracies established by previous studies. Therefore, common pre-processing techniques were applied to a benchmark HD-sEMG dataset (CapgMyo DB-a) and their validation accuracies were compared. Monopolar, bipolar, rectified, common-average referenced, and Laplacian spatial filtered configurations of the HD-sEMG signals were evaluated. Results showed that the baseline monopolar HD-sEMG signals maintained higher prediction accuracies versus the other signal configurations. The results of this study discourage the use of extra pre-processing steps when using convolutional networks to classify the instantaneous samples of HD-sEMG for gesture recognition.**

***Keywords*: High-density Surface EMG; Classification; Gesture; Convolutional Network**

I. Introduction

Gesture recognition using features from electromyography (EMG) is not a new concept, but the methods and implementations have advanced over the years. It has been shown that gesture recognition using features extracted from EMG can be more efficient [1] and cost-effective [2] versus traditional rehabilitation methods. Informing the gesture recognition with EMG patterns has clinical relevance in helping regain some level of motor control for stroke patients, among other motor-impaired individuals. For example, features extracted from EMG at forearm muscles have been used in muscle-computer interfaces to assist motor-impaired individuals such as chronic stroke patients in performing hand gestures with training in a case study [3]. Lu et al. demonstrated in this study that robot-assisted rehabilitation using features extracted in real-time from EMG an improvement in the Fugl-Meyer (Part C) clinical score from 0 to 7. However, the feasibility of this kind of work should be assessed on a patient-by-patient basis [4]. EMG-driven gesture recognition has also been successful in some applications with cervical spinal-cord injury (SCI) patients [5]. High-density surface electromyography (HD-sEMG) has also been applied by Zhang and Zhou [6] to classify 20 different arm, hand, and finger movements using 89 electrodes with accuracies of 96.1% ± 4.3%, indicating that there is substantial amount of information to be implemented from HD-sEMG with the inherent benefits of high spatio-temporal resolution. HD-sEMG surface maps were shown by Rojas-Martínez et al. to be quite useful in classifying the motor intent [7, 8]. Leveraging the high spatio-temporal resolution of HD-sEMG signals allows for the high-density mapping of the muscles, providing more information than single-site electrodes per muscle.

Deep learning methods such as convolutional neural networks (CNN) and have also recently been explored for the purpose of decoding fine motor control from HD-sEMG. Geng et al. first proposed the idea of using a CNN to classify so-called HD-sEMG instantaneous images, produced by re-arranging individual samples of HD-sEMG into 2D arrays with above state-of-the-art classification accuracies of 89.3% within-subjects on just 1 image and 99.0% after 40 images with 1000 Hz sampling rate [9]. Other studies have followed suit in expanding upon the original neural network architecture proposed by Geng et al. by including a deep-learning-based domain adaptation framework [10], using hybrid convolutional recurrent neural networks [11], multi-stream convolutional networks [12], 3D convolutional networks versus the usual 2D convolutional networks [13], and multi-label classification [14]. Deep learning is extremely useful for EMG classification, identifying spatial and temporal patterns in HD-sEMG signals without the need to manually generate any features from the signals before inputting into networks.

Several deep learning methods have been studied to improve upon the existing classification accuracies, but different pre-processing techniques have yet to be explored. In this study, common and simple pre-processing techniques were used prior to HD-sEMG image generation to identify what the most ideal pre-processing pipeline would be to increase the classification accuracies of hand gesture recognition. It is hypothesized that re-referencing the HD-sEMG signals prior to image generation can provide higher classification accuracies while not adding too much added latency to the process.

II. Methods

A. Data Description

The CapgMyo DB-a dataset is a publicly available dataset online originally provided by Geng et al. [9]. It consists of 128-

Manusccript received May 3, 2021.

Michael Houston is with the Department of Biomedical Engineering at the University of Houston, Houston, TX 77204 USA.

Albon Wu is with the William P. Clements High School, Sugar Land, TX 77479 USA, and the Department of Biomedical Engineering at the University of Houston, TX 77204.

Yingchun Zhang is with the Department of Biomedical Engineering at the University of Houston, Houston, TX 77204 USA (e-mail: yzhang94@uh.edu).

channel monopolar HD-sEMG sampled at 1000 Hz from eighteen participants, eight gestures, and ten repetitions per gesture. Each trial contains 1 second of steady state contraction for a specified gesture. The pre-processed version of the dataset was used in this study. The previous pre-processing included a bandpass filter between 20-380 Hz, and a notch filter at 50 Hz for powerline artifacts.

The DB-a dataset consists of the following hand gestures: thumb up, extension of middle and ring finger with flexion of all other fingers, flexion of ring and little fingers with extension of all other fingers, thumb opposing base of little finger, abduction of all fingers, all fingers flexed into a fist, index finger pointed, and adduction of extended fingers.

B. HD-sEMG Pre-Processing Techniques

HD-sEMG signals were subjected to one pre-processing technique before converting the signals into 2D grayscale images. The original HD-sEMG signals were either kept monopolar, converted into a bipolar configuration, fully rectified, re-referenced to the common average, or subjected to a Laplacian spatial filter such that the new HD-sEMG image, *LAP*, is formed by convolving the original image, *IMG*, with the Laplacian kernel (1). Bipolar HD-sEMG images were of size 7x16 whereas all other HD-sEMG image configurations were of size 8x16. Huang et al. have shown that Laplacian spatial filters can improve the accuracy of EMG classification on patients after targeted muscle reinnervation procedures [15].

$$LAP_{8x16} = IMG_{8x16} \otimes \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (1)$$

C. Neural Network Architecture, Training, and Validation

All data analysis, including image creation, network creation, training, and validation, was performed in the MATLAB environment (MATLAB R2020b, The MathWorks, Inc.). Grayscale images served as inputs into network, followed by two 2D convolutional layers (64 filters, 3x3, stride of 1), three fully connected layers (512, 512, and 128 units respectively), and a G-way fully connected layer followed by a softmax function. The image input was followed by a batch normalization layer. Each convolutional layer and fully connected layer were followed by batch normalization and ReLu activation. Dropout layers with a probability of 0.5 were placed after each fully connected layer.

Odd-numbered trials were used for model training and even-numbered trials were used for model validation per-subject. The models were trained with 20 epochs on the validation dataset with a validation frequency of 40. The mini-batch size was set to 1000. Stochastic gradient descent with momentum (SGDM) was used as the optimizer initialized at 0.1 and decreased by a factor of 10 after the 10^{th} epoch to 0.01 with a momentum of 0.9.

III. RESULTS

A. Comparing Prediction Accuracy

Prediction accuracies were assessed by counting the number of correctly predicted labels in the even-numbered trials used as the validation set. Table I reports the prediction accuracies for all five data configurations tested while Table II reports the subject-specific prediction accuracies achieved with the convolutional networks along with the subject identifiers.

TABLE I. PREDICTION ACCURACY COMPARISON

CapgMyo DB-a Dataset		
Data Type	**Mean**	**Deviation**
Baseline	83.31%	± 7.24%
Bipolar	81.25%	± 6.91%
Rectified	82.34%	± 6.05%
Common Average	83.30%	± 7.03%
Laplacian	82.60%	± 6.83%

TABLE II. INDIVIDUAL RESULTS

Subject	Baseline	Bipolar	Rectified	Common Average	Laplacian
1	92.15%	89.54%	92.19%	91.98%	90.74%
2	85.29%	82.55%	83.37%	85.24%	83.99%
3	90.34%	88.01%	87.84%	90.24%	89.57%
4	88.94%	87.33%	84.94%	88.77%	87.96%
5	78.25%	75.82%	75.96%	77.81%	78.45%
6	91.06%	88.68%	87.67%	90.91%	89.99%
7	85.89%	81.41%	86.52%	85.41%	83.42%
8	88.86%	88.29%	86.09%	89.27%	88.77%
9	64.32%	64.95%	84.23%	65.78%	65.78%
10	80.71%	78.95%	79.96%	81.00%	80.54%
11	77.77%	76.31%	75.64%	77.81%	77.98%
12	80.01%	78.74%	76.39%	79.79%	79.08%
13	85.46%	82.40%	84.09%	85.59%	85.33%
14	78.91%	73.94%	74.92%	78.72%	76.67%
15	80.28%	78.55%	76.47%	80.06%	79.17%
16	74.35%	73.17%	71.70%	74.30%	73.70%
17	92.22%	90.82%	91.93%	92.35%	91.53%
18	84.79%	83.04%	82.19%	84.43%	84.07%

The baseline monopolar HD-sEMG signals had the best classification results with a prediction accuracy of 83.31% whereas the bipolar HD-sEMG signals had the worst prediction accuracy at 81.25%. The common average re-referenced data performed nearly like the baseline data with a prediction accuracy of 83.30%. Individual results were also compared with paired boxplots and are displayed in Fig. 1 below. Fig. 1 demonstrates the consequences of the added pre-processing on individual subject datasets. Subject 9 was the only subject to receive a notable increase in prediction accuracy, most pronounced with the rectified data. The prediction accuracy increased about 20% from 64% to 84% approximately.

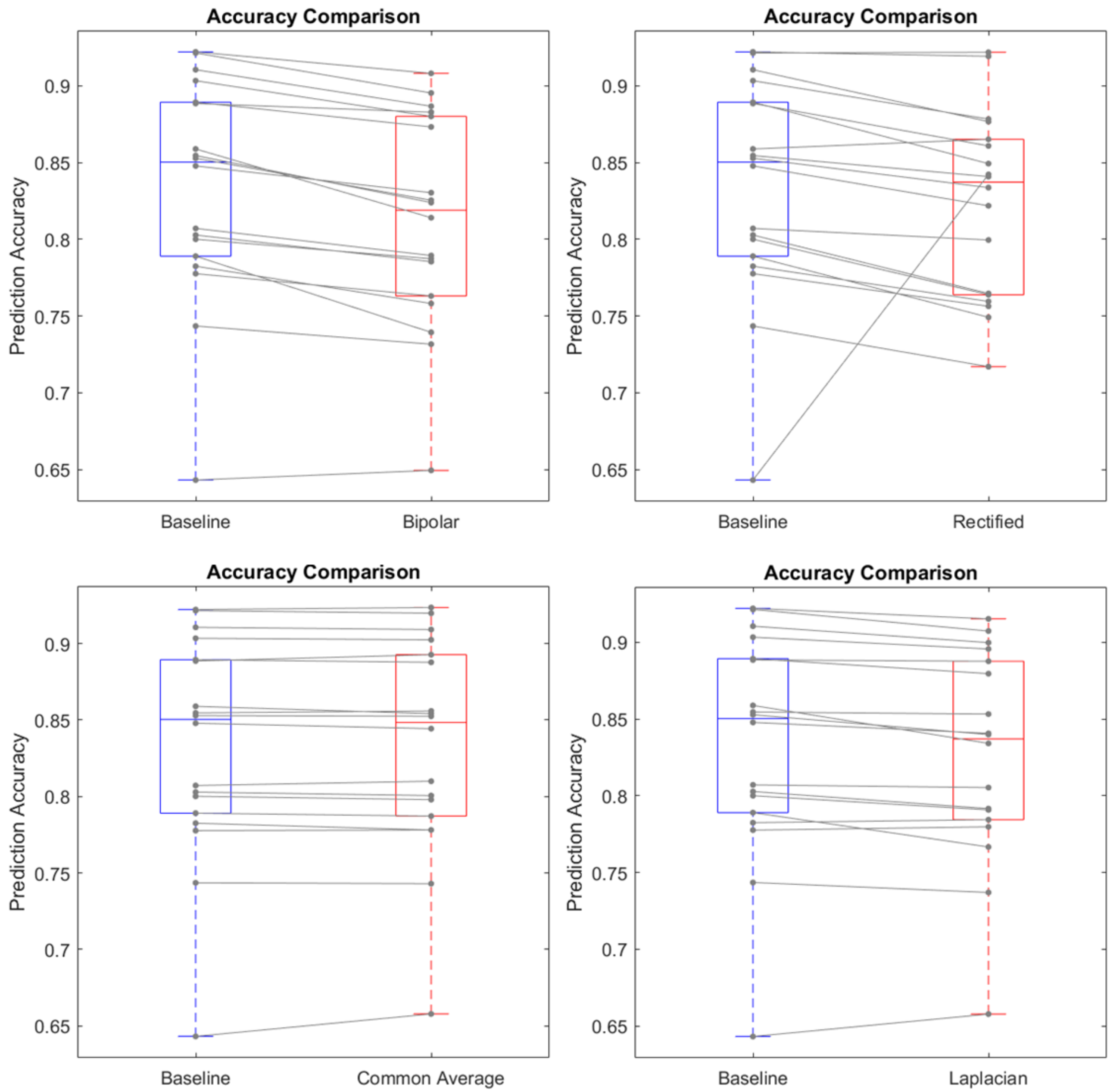

Figure 1. Individual Results on DB-a Dataset. All four pre-processing steps were compared to the baseline monopolar signals.

IV. DISCUSSION

HD-sEMG signals provide a high spatio-temporal resolution which can be exploited for nearly instantaneous recognition of hand gestures in real-time applications. Such spatio-temporal patterns of HD-sEMG images can be identified using deep learning methods involving convolutional networks. Reducing the amount of pre-processing necessary for real-time applications such as prosthesis-control and motor rehabilitation can improve the quality of these applications for use with motor-impaired individuals.

The results of this study did not meet the expectations of the proposed hypothesis. Monopolar HD-sEMG signals were found to produce better classification results versus adding any other common EMG pre-processing steps prior to image generation and classification. On a subject-by-subject basis, converting the HD-sEMG to bipolar signals was only beneficial to one subject. Rectifying the HD-sEMG signals increased the classification accuracies only for three subjects. Re-referencing the HD-sEMG to the common average improved classification results for six subjects. Convolving the HD-sEMG images with the Laplacian spatial filter helped only for three subjects. While the common average re-referencing did improve the results for one-third of the sample-size it was not beneficial enough to be considered a significant improvement from the baseline signals across all subjects. Future studies may consider re-referencing their signals to the common average on a subject-specific basis to observe any potential benefits from removing the common noise component.

However, the results of this study do support the idea that minimal pre-processing is necessary for accurate hand gesture recognition of HD-sEMG signals, which is helpful due to the limited time constraints of real-time control applications. Additionally, Yao et al. identified that active shielding in the recording system improved the classification of EMG when greater than 12 channels were used by removing the common component of signal [16]. Such active shielding could benefit new studies using deep-learning methods to classify HD-sEMG and its usage should be explored.

The results of this study are comparable to that of Geng et al. [9] when it comes to the decrease in prediction accuracy with added pre-processing steps. Geng et al. also explored the consequences of rectifying the HD-sEMG as well as removing crosstalk from the signals. They identified a reduction in the prediction accuracy for both of those methods while a similar reduction was found when rectifying the data here once again as well as removing the common-average from the signals. While the effects of added pre-processing were observed and quantified for HD-sEMG signals it remains to be seen whether the same effects are visible in sparser sEMG datasets that are not HD-sEMG such as the NinaPro dataset.

This study demonstrates the advantages and disadvantages of certain HD-sEMG signal configurations for hand gesture recognition. The results of this study can inform future clinical studies intending to use the CNN HD-sEMG for prosthesis control and or motor rehabilitation for neurologically impaired individuals such as chronic stroke patients or amputees.

V. Conclusion

Different common pre-processing techniques were assessed in HD-sEMG signals to determine the most optimal input images to a CNN for hand gesture recognition in a benchmark dataset. It was found that the baseline monopolar signals provided the best prediction accuracy compared to all other configurations studied. The results of this study suggest that only filtering the HD-sEMG signals prior to creating instantaneous images is sufficient for classifying hand gestures with high prediction accuracy.

References

[1] P. S. Lum, C. G. Burgar, P. C. Shor, M. Majmundar, and M. Van der Loos, "Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke," *Archives of Physical Medicine and Rehabilitation*, vol. 83, no. 7, pp. 952–959, 2002.

[2] T. H. Wagner, A. C. Lo, P. Peduzzi, D. M. Bravata, G. D. Huang, H. I. Krebs, R. J. Ringer, D. G. Federman, L. G. Richards, J. K. Haselkorn, G. F. Wittenberg, B. T. Volpe, C. T. Bever, P. W. Duncan, A. Siroka, and P. D. Guarino, "An Economic Analysis of Robot-Assisted Therapy for Long-Term Upper-Limb Impairment After Stroke," *Stroke*, vol. 42, no. 9, pp. 2630–2632, 2011.

[3] Z. Lu, X. Chen, X. Zhang, K.-Y. Tong, and P. Zhou, "Real-Time Control of an Exoskeleton Hand Robot with Myoelectric Pattern Recognition," *International Journal of Neural Systems*, vol. 27, no. 05, p. 1750009, 2017.

[4] Z. Lu, K. Tong, X. Zhang, S. Li, and P. Zhou, "Myoelectric Pattern Recognition for Controlling a Robotic Hand: A Feasibility Study in Stroke," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 365–372, 2019.

[5] Z. Lu, X. Chen, X. Zhang, K.-Y. Tong, and P. Zhou, "Real-Time Control of an Exoskeleton Hand Robot with Myoelectric Pattern Recognition," *International Journal of Neural Systems*, vol. 27, no. 05, p. 1750009, 2017.

[6] X. Zhang and P. Zhou, "High-Density Myoelectric Pattern Recognition Toward Improved Stroke Rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1649–1657, Jun. 2012.

[7] M. Rojas-Martínez, M. A. Mañanas, and J. F. Alonso, "High-density surface EMG maps from upper-arm and forearm muscles," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, 2012.

[8] M. Rojas-Martínez, M. A. Mañanas, J. F. Alonso, and R. Merletti, "Identification of isometric contractions based on High Density EMG maps," *Journal of Electromyography and Kinesiology*, vol. 23, no. 1, pp. 33–42, 2013.

[9] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Scientific Reports*, vol. 6, no. 1, 2016.

[10] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, "Surface EMG-Based Inter-Session Gesture Recognition Enhanced by Deep Domain Adaptation," *Sensors*, vol. 17, no. 3, p. 458, 2017.

[11] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *Plos One*, vol. 13, no. 10, 2018.

[12] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, and W. Geng, "A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface," *Pattern Recognition Letters*, vol. 119, pp. 131–138, 2019.

[13] J. Chen, S. Bi, G. Zhang, and G. Cao, "High-Density Surface EMG-Based Gesture Recognition Using a 3D Convolutional Neural Network," *Sensors*, vol. 20, no. 4, p. 1201, 2020.

[14] A. E. Olsson, P. Sager, E. Andersson, A. Björkman, N. Malešević, and C. Antfolk, "Extraction of Multi-Labelled Movement Information from the Raw HD-sEMG Image with Time-Domain Depth," *Scientific Reports*, vol. 9, no. 1, 2019.

[15] H. Huang, P. Zhou, G. Li, and T. Kuiken, "Spatial Filtering Improves EMG Classification Accuracy Following Targeted Muscle Reinnervation," *Annals of Biomedical Engineering*, vol. 37, no. 9, pp. 1849–1857, 2009.

[16] B. Yao, Y. Peng, X. Zhang, Y. Zhang, P. Zhou, and J. Pu, "The influence of common component on myoelectric pattern recognition," *Journal of International Medical Research*, vol. 48, no. 3, p. 030006052090361, 2020.