2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC) | 979-8-3503-9938-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISACC56298.2023.10083548

Surface Electromyography based Hand Gesture Signal Classification using 1D CNN

Krishnapriya S
Department of ECE
National Institute of Technology
Rourkela, India
krishnapriyaskumar@gmail.com

Jaya Prakash Sahoo
Department of ECE
National Institute of Technology
Rourkela, India
sahoo.jprakash@gmail.com

Samit Ari
Department of ECE
National Institute of Technology
Rourkela, India
samit@nitrkl.ac.in

***Abstract*—Surface electromyography-based hand gesture recognition for human-machine interfacing has captivated the interest of researchers in recent years. Machine learning techniques are highly dependent on the features selected for classification. A poor choice of features can deteriorate the classification ability of the system. In order to overcome this issue, deep learning techniques which are capable of automatic feature extraction are employed these days. For any ideal classifying machine, it is required to have a short training period and an ability to produce classification results within a short interval of applying an input. This work proposes a 1D CNN model which is able to classify all the 52 gestures of the NinaPro DB1 database which can be applied in real time applications with fast response. The developed model achieves an accuracy of 78.95% with just 351,532 trainable parameters and an average inference time of 0.258 ms.**

***Index Terms*—sEMG, 1D CNN, NinaPro**

I. Introduction

Biomedical signal processing is extensively researched in the scientific society due to its wide range of applications in health care and human machine interactions. In recent years several techniques are being developed to recognize the different human movements and transform them into meaningful commands to control the operation of machines. The use of biomedical signals such as ECG, EEG, EMG, etc. for human machine interactions has become very popular. Electromyography (EMG) is the electrical signals that result from muscle contractions in human body. They reflects information about the movements of a subject. Surface electromyography (sEMG) which are signals collected from the surface of muscles using electrodes, is widely applied for hand gesture classification, robot control, development of prosthetic and in many more applications [1]–[3] because of its user friendliness and low cost. Several researchers in the recent years have proposed deep learning techniques to classify sEMG signals and have proved that it produces promising results. Although sEMG gesture recognition using hand crafted features achieve good results, they have several drawbacks. It performs poorly under varying conditions in the dataset and hence is impacted by external influences. These hand crafted machine learning techniques cannot deal with the complexity of sEMG signals and require a lot of manual technique for feature extraction. As an alternative, deep learning techniques which are automatic feature learning algorithms are employed. Most EMG gesture classifications convert the information into 2 dimensional data using short time fourier transform (STFT), Wavelet, etc. Oh *et al.* [4] have converted the signal into two dimensional (2D) images by taking the short time Fourier transform (STFT) and wavelet transform (WT). The transforms are performed channel wise to obtain the results. This image is then given to a simple CNN model for classification. Wang *et al.* [5] have converted the signal into the time- frequency domain by using STFT. The designed structure has five convolution layers followed by fully connected layers for classification. Wang *et al.* [6] have proposed a novel split spectrogram approach where the spectrogram is first split along the frequency for each channel and them merged along the channels. This image is then given to a CNN model. While 2D CNNs are mostly used for feature extraction in images, 1D CNNs are found to perform well on physiological signals [7]. Kim *et al.* [8] proposed a 1D CNN model that can recognize the respiratory patterns obtained from a ultra wide band radar and predict the health condition of a person. The CNN model consisted of 3 convolution layers, 4 dense layers and a dropout layer along with the dense layers to reduce over fitting. A novel 1D CNN architecture that can detect autism spectrum disorder from EEG signals is proposed by Mohi *et al.* [9]. The authors have developed a CNN model with 6 1D convolution layers and 4 fully connected layers that has a recognition accuracy of 92%. It was observed that as more layers are added, computation time as well as efficiency of the model increases. Cheikhrouhou *et al.* in [10] has proposed a 1D CNN model that analyses ECG signals and detect arrhythmia cardiovascular diseases with an accuracy of 99.46%. The proposed model consisted of two 1D convolution layers and two dense layers. Dropouts are applied during the training to avoid over fitting. From the literature survey, we can conclude that even though EMG signals are one dimensional signals, 1D CNN architecture has not been applied in EMG signal classification. Since 1D CNNs has been used in various other applications with several other biological signals and since it requires low computation due to its compact configuration, it is fit to apply in real time and low expense applications.

978-1-6654-7350-7/22/$31.00 ©2022 IEEE

The contributions in this work are as follows:

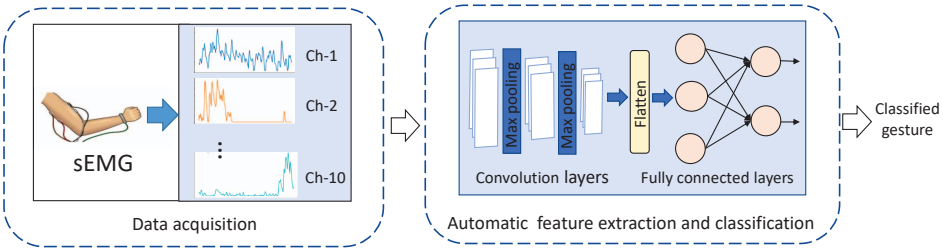

Fig. 1: Block diagram representation of the proposed s-EMG based hand gesture recognition system.

- A 1D CNN model architecture is proposed that can classify the 52 hand gesture signals of the Ninapro DB1 database accurately.
- The performance of the recognition system on a wide range of gestures is evaluated by using thre publicly available 52 gesture classes of NinaPro DB1 dataset
- A model with less number of trainable parameters and that requires a low average inference time making it suitable for real time applications is proposed.

The rest of the paper is organized as follows. The methodology of the proposed work along with the descrption of the database are discussed in Section II. The detailed experimental results and discussions are presented in the Section III. Section IV concludes the paper and provides future scope of the work.

II. METHODOLOGY

The main steps that are to be followed for gesture recognition using sEMG signals is shown in Fig. 1. The major steps are data acquisition, and automatic feature extraction and classification. In data acquisition, the sEMG signals are gathered using ten sensors of the MyoBock sensor. A 1D CNN structure is proposed for automatic feature extraction and classification of sEMG signals. The convolution layers perform the feature extraction while the fully connected layers perform the classification based on the features extracted.

A. Data acquisition and preprocessing

The NinaPro DB1 dataset is a collection of 52 hand gestures, which are repeated 10 times by 27 subjects. The gesture includes 3 exercises consisting of 12 basic finger movements and 8 isometric and isotonic hand configurations, 9 basic wrist movements and 23 functional and grasping movements as shown in Fig. 2. An amplified, bandpass-filtered and root mean square rectified version of the raw sEMG signal is expressed as the output of 10 double-differential OttoBock MyoBock 13E200 sEMG electrodes. The amplification factor is set to 14000 and the two filter cut off frequencies are at 90 Hz and 450 Hz. Data is acquired at a frequency of 100 Hz [2].

B. 1D CNN Model

The proposed 1D CNN model consists of five 1D convolution layers followed by four fully connected layers as shown in Fig. 3. The proposed architecture is arrived by

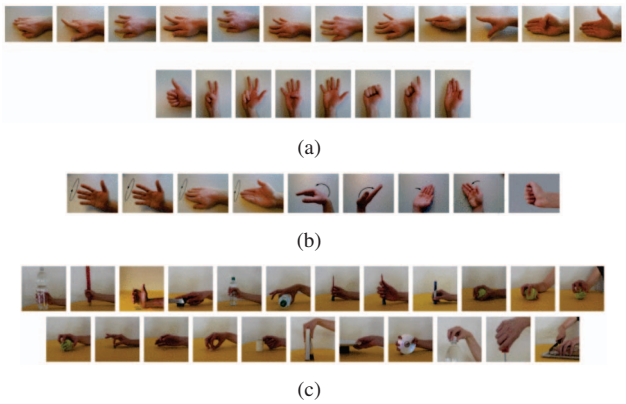

Fig. 2: 52 gesture classes of the NinaPro DB1 dataset [2]. (a) 12 basic finger movements and 8 isometric and isotonic hand configurations; (b) 9 basic wrist movements; (c) 23 functional and grasping movements.

experimenting on various combinations of layers and kernels which is explained in detail in section II. ReLU activation function is used in all of the convolution layers since it is easier to train and achieves better performance. In order to reduce the computational cost, the number of parameters are reduced by using max pooling operation after each convolution layer. The final dense layer uses soft max activation function while the other dense layers use ReLU activation function. Dropout layers are inserted after each max pooling layer to reduce over fitting. A dropout probability of 0.3 is used in between the convolution layers while a dropout probability of 0.15 is used in between the fully connected layers. Hyper parameters like the number of filters, kernel size, stride and dropout probability in each layer are specified in table I. The total number of trainable parameters in each layer is also presented in the table. The working of this model is just like any other CNN model. The only difference is that since 1D CNNs are used, the kernels in the CNN layers only traverse in one direction. Just like in any other CNN model, hidden features from the sEMG signal are extracted by this model.

C. Experimental Method

The entire 52 gestures of the NinaPro DB1 database is divided into three sections for training, validation and testing.

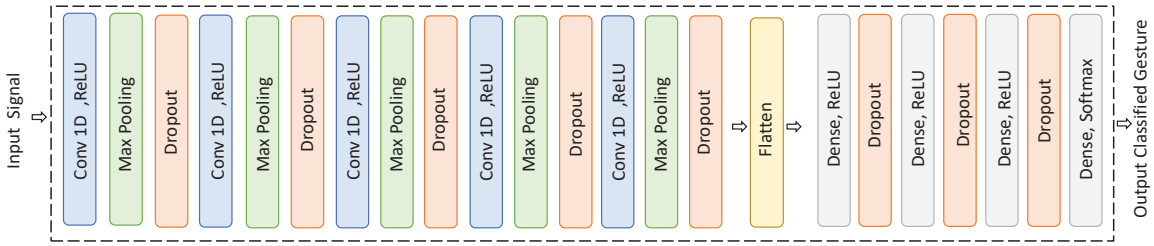

Fig. 3: Proposed 1D CNN architecture for recognition of sENG-based hand gesture signals.

TABLE I: Layer-wise details of the proposed 1D CNN

Layer	Type	Filters	Units	Kernel Size	Pool Size	Stride	Probability	Activation	No. of Parameters
conv1_1	1D Convolution	128		3				relu	3968
pool1	Max Pooling 1D				2	2			
	Dropout						0.3		
conv2_1	1D Convolution	64		3				relu	24640
pool2	Max Pooling 1D				2	2			
	Dropout						0.3		
conv3_1	1D Convolution	32		3				relu	6176
pool3	Max Pooling 1D				2	2			
	Dropout						0.3		
conv4_1	1D Convolution	16		3				relu	1552
pool4	Max Pooling 1D				2	2			
	Dropout						0.3		
conv5_1	1D Convolution	8		3				relu	392
pool5	Max Pooling 1D				2	2			
	Flatten								
dense_1	Fully-connected		512					relu	143872
	Dropout						0.15		
dense_2	Fully-connected		256					relu	131328
	Dropout						0.15		
dense_3	Fully-connected		128					relu	32896
	Dropout						0.15		
dense_4	Fully-connected		52					softmax	6708

The repetitions 1, 3, 4, 6, 8 and 10 are used for training the model while repetitions 5 and 9 are used in the validation and repetitions 2 and 7 for testing. Adam optimizer with a learning rate of 0.001 along with a categorical cross entropy loss is used to train the model in batches of size 64.

D. Evaluation metrics

1) F1 Score: It gives the weighted average of precision and recall. The closer its value is to 1, the better the ability of the model to classify accurately.

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{1}$$

2) Trainable parameters: Any parameter in the model that is learned during the training process is termed as trainable parameters (*TPs*). The time to train a model highly relies on the number of trainable parameters. Let the width and height of the filter be a and b respectively. If k and l are the number of filters in the previous and current layer, the trainable parameters in the convolution layer is calculated as in equation 2

$$TPs\ in\ convolution\ layer = ((a * b * k) + 1) * l \tag{2}$$

The input and pooling layer doesn't have any trainable parameters. For a fully connected layer it is computed as shown in equation 3. Here m neurons are present in the current layer and n neurons are present in the previous layer. The bias terms also need to be considered.

$$TPs\ in\ fully\ connected\ layer = ((m * n) + 1 * m) \tag{3}$$

3) Average inference time: The time taken by the model to produce the classification results for some test input data is called average inference time (*AIT*). In order to apply the model in real time applications, the average inference time requires to be low. The calculation of *AIT* by evaluating the time to classify the entire test data is shown in equation 4

$$AIT = \frac{Time\ to\ evaluate\ entire\ test\ data\ set}{Number\ of\ samples\ in\ test\ set} \tag{4}$$

III. RESULTS AND DISCUSSIONS

In this section, the performance of the proposed 1D CNN are studied for variation in several parameters such as kernel size of the convolution layers, number of convolution layers and number of dense layers. Fig. 4(a) depicts the variation

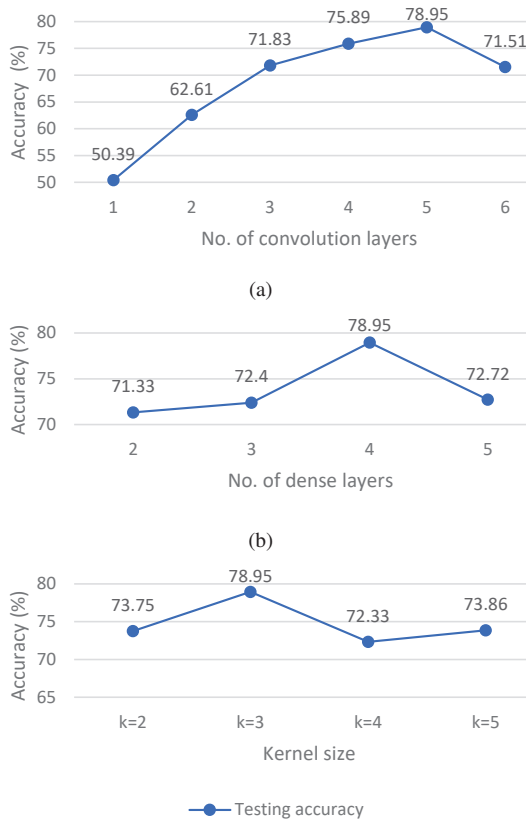

Fig. 4: Variation in testing accuracy on varying (a) the number of convolution layers (b) the number of dense layers (c) the kernel size in the convolution layers

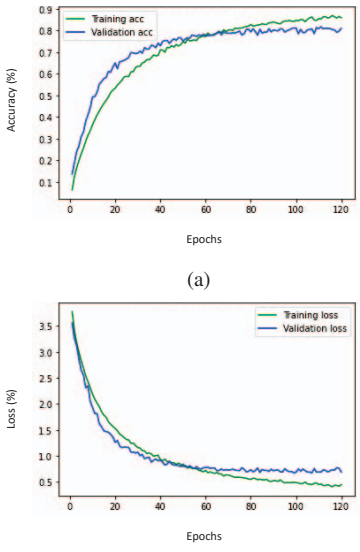

Fig. 5: Training plots (a)Training and validation accuracy plots, (b)Training and Validation loss plots

TABLE II: Performance measure of the 1D CNN model on the NinaPro DB1 database.

Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)	F1 score	Trainable parameters	Average inference time (msec)
85.90	80.95	78.95	0.78949	351,532	0.2584

in testing accuracy of the model on increasing the number of convolution layers from 1 to 6. It can be observed that the accuracy increases on increasing the number of convolution layers and the highest accuracy is obtained with 5 convolution layers. Further increase in the number of layers decreases the accuracy. Fig. 4(b) displays the variation in testing accuracy with increase in the number of dense layers from 2 to 5. We can conclude from the graph that the accuracy increases on increasing the number of dense layers from 2 to 4 and then decreases on further increase in layers. So the highest accuracy is obtained at 4 dense layers. Fig. 4(c) shows the variation in testing accuracy on varying the kernel size in all the convolution layers. It can be observed that a kernel size of 3 performs the best. From the hyper parameter studies shown in Fig. 4 we can conclude that the best accuracy is obtained with 5 convolution layers, 4 dense layers and a kernel size of 3. The weighted average F1 score which is the average of F1 scores obtained for each class, number of trainable parameters which is the sum of parameters in each layer and the average inference time which is the time taken by the model to classify one gesture is also calculated. The performance measures of the proposed 1D CNN are presented in Table II. The accuracy and loss curves for the training and validation of the model are displayed in Fig. 5 for 120 epochs. Fig. 6 shows the confusion matrix obtained for the proposed 1D CNN. The confusion matrix results show a high value along the diagonals which is expected for a good classifier. We can also see that there are several gestures that show miss classification, especially the ones in exercise 3 (corresponding to functional and grasping movements) which is one of the drawbacks of this model. Its inability to learn some hidden features which can help in distinguishing between very similar hand gestures. Real time applications mostly make use of hand configurations and wrist movements (gestures in exercise 1 and 2). So, by choosing a subset of gestures, one can get a much better accuracy and make use of them for real time applications. Table III shows the comparison of the performance of the proposed 1D CNN with existing hand crafted machine learning techniques and neural networks. It is evident from the table that the 1D CNN architecture has a better accuracy when compared to hand crafted machine learning techniques and some neural networks.

IV. Conclusions

This work has introduced a 1D CNN architecture that is able to distinguish the 52 hand gesture classes of sEMG signals. The experimental results show that a recognition accuracy of

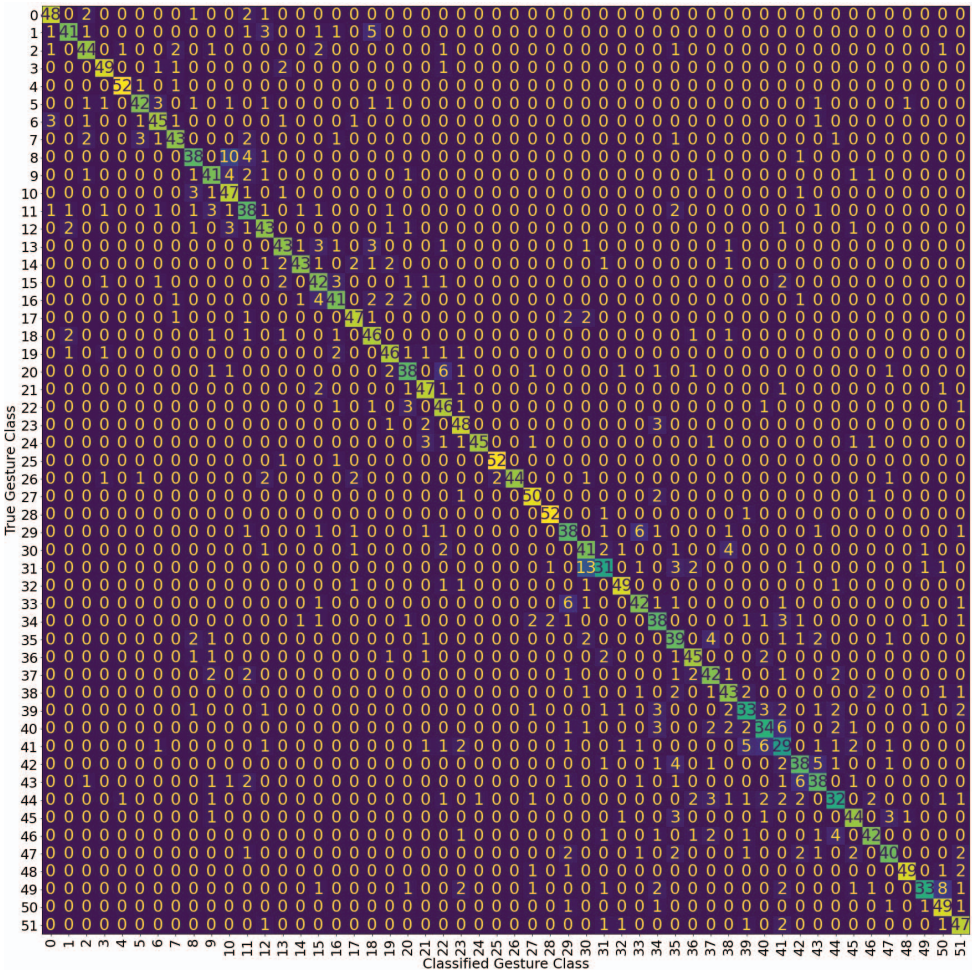

Fig. 6: Confusion matrix of the proposed 1D CNN model on NinaPro DB1 dataset

TABLE III: Performance comparison the proposed 1D CNN with existing methods on NinaPro DB1 Database

Author	Method	Accuracy (%)
Atzori.*et.al.* [2]	*WL*+ kNN	73
Atzori.*et.al.* [2]	*WL*+ SVM	75
Atzori.*et.al.* [3]	*RMS, mDWT*+ kNN	65
Atzori.*et.al.* [3]	*MS + TD + HIST + mDWT*+ RF	75
Pizzolato.*et.al.* [11]	*RMS + TD + HIST + mDWT*+ SVM	60
Pizzolato.*et.al.* [11]	*RMS + TD + HIST + mDWT*+ RF	65
Cene.*et.al.* [12]	*RMS + VAR + MAV + SD*+ R-RELM	75.03
Y.He.*et.al.* [13]	LSTM + MLP	75.45
Atzori.*et.al.* [3]	2D CNN	66.59
Proposed work	1D CNN	78.95

78.95% is achieved using the 1D CNN on 52 gesture classes of NinaPro DB1 dataset. By making use of a subset of gestures as per the requirement, one can achieve recognition accuracies suitable for real time applications. The model has very few trainable parameters and a small average inference time when compared to the state of the art techniques which makes it suitable to operate in real time applications that require easy and fast recognition systems. The accuracy of the model can be further improved to reduce misclassification by increasing the dataset size.

REFERENCES

[1] L. Guo, Z. Lu, and L. Yao, "Human-machine interaction sensing technology based on hand gesture recognition: A review," *IEEE Transactions on Human-Machine Systems*, 2021.

[2] M. Atzori, A. Gijsberts, I. Kuzborskij, S. Elsig, A.-G. M. Hager, O. Deriaz, C. Castellini, H. Müller, and B. Caputo, "Characterization of a benchmark database for myoelectric movement classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 1, pp. 73–83, 2014.

[3] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific data*, vol. 1, no. 1, pp. 1–13, 2014.

[4] D. C. Oh and Y. U. Jo, "Emg-based hand gesture classification by scale average wavelet transform and cnn," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, 2019, pp. 533–538.

[5] Q. Wang and X. Wang, "Emg-based hand gesture recognition by deep time-frequency learning for assisted living and rehabilitation," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2020, pp. 0558–0561.

[6] S. Wang and B. Chen, "Split-stack 2d-cnn for hand gestures recognition based on surface emg decoding," in *2020 Chinese Automation Congress (CAC)*, 2020, pp. 7084–7088.
[7] S. Bianco and P. Napoletano, "Biometric recognition using multimodal physiological signals," *IEEE Access*, vol. 7, pp. 83 581–83 588, 2019.
[8] S.-H. Kim and G.-T. Han, "1d cnn based human respiration pattern recognition using ultra wideband radar," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 411–414.
[9] Q. Mohi–ud–Din and A. K. Jayanthy, "Autism spectrum disorder classification using eeg and 1d-cnn," in *2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)*, 2021, pp. 01–05.
[10] O. Cheikhrouhou, R. Mahmud, R. Zouari, M. Ibrahim, A. Zaguia, and T. N. Gia, "One-dimensional cnn approach for ecg arrhythmia analysis in fog-cloud environments," *IEEE Access*, vol. 9, pp. 103 513–103 523, 2021.
[11] S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori, "Comparison of six electromyography acquisition setups on hand movement classification tasks," *PloS one*, vol. 12, no. 10, p. e0186132, 2017.
[12] V. H. Cene, M. Tosin, J. Machado, and A. Balbinot, "Open database for accurate upper-limb intent detection using electromyography and reliable extreme learning machines," *Sensors*, vol. 19, no. 8, p. 1864, 2019.
[13] Y. He, O. Fukuda, N. Bu, H. Okumura, and N. Yamaguchi, "Surface emg pattern recognition using long short-term memory combined with multilayer perceptron," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 5636–5639.