sEMG-Based Tremor Severity Evaluation for Parkinson's Disease Using a Light-Weight CNN

Zengyi Qin, Zhenyu Jiang, Jiansheng Chen, *Senior Member, IEEE*, Chunhua Hu, and Yu Ma

***Abstract*—We propose a deep learning based approach for quantifying the tremor severity of Parkinson's disease (PD) based on surface electromyography (sEMG). We design the S-Net, a light weight and computational efficient convolutional neural network that learns the similarity between sEMG signals in terms of the tremor severity. Labeled sEMG samples are used for jointly voting for the final results. Experiments on 147 PD patients demonstrate that our approach outperforms traditional methods by a significant margin. In addition, our approach is simple and has potentials in real applications.**

***Index Terms*—Parkinson's Disease, tremor severity level, sEMG, UPDRS, CNN, similarity learning.**

I. Introduction

OCCURING on limbs and faces at frequencies of 4 to 6 Hz, rest tremor is one of the characteristic movement disorders of the Parkinson's Disease (PD) [1]. Severity of the rest tremor is a critical factor that neurosurgeons consider when deciding treatment and rehabilitation strategies for PD patients. In the widely used PD evaluation standard, Unified Parkinson's Disease Rating Scale (UPDRS) [2], tremor severity is classified in to 5 levels, from 0 (absent) to 4 (most severe). In practice, the tremor severity is usually evaluated by clinicians through subjective inspection. To perform the evaluation, patients are required to come to the hospital which may cause inconveniences considering their limited mobility. Furthermore, such a manual inspection approach may bring heavy burdens to the health care system in the aging societies. Therefore, it is worthwhile to explore automated approaches for subjective evaluation of PD tremor severity.

In this work, we study the PD tremor severity evaluation based on the surface electromyography (sEMG). As an electric signal generated during muscle contraction, sEMG directly reflects the activity patterns of the muscle [3]. It has already been extensively explored in the detection and classification of different types of tremor [4], [5]. However, little research has been done on the fine grained quantification of tremor severity using sEMG. This is probably due to the high variations in sEMG signals caused by factors such as muscle strength [6], fatty tissues [7], as well as the impedance between the skin and electrodes. Cole *et al.* [8] tried to apply a naive Bayesian maximum likelihood classifier to the high-pass energy features of sEMG to estimate PD severity levels. However, instead of using the 5 level UPDRS evaluation, they only classified the PD tremor severity into 3 levels, namely mild, moderate and severe. Later we will show that this method cannot be effectively scaled to the 5 level UPDRS evaluation task.

To effectively classify sEMG signals of PD patients to the 5 tremor severity levels defined in UPDRS, we propose a solution which combines a Convolutional Neural Network (CNN) with similarity learning [9]. We abandon the traditional way of directly applying a multi-class classifier to the sEMG signal. Instead, we propose the S-Net, a light-weight CNN that predicts whether two sEMG signals are of the same tremor severity level. As such, the tremor severity of a testing sEMG signal can be determined by comparing it to a set of sEMG signals of which the tremor severity levels are clinically labeled. Experimental results show that our method achieves a classification accuracy of 90.55% and outperforms traditional classifiers by a significant margin. Besides, S-Net is small in size, allowing fast computation and easy deployment.

II. Data Preparation

This study was approved by the IRB of Tsinghua University Yuquan Hospital with the reference number YQ2015-08-02. A total of 147 PD patients participated in the data collection. The UPDRS tremor severity level of each PD patient, which is an integer from 0 to 4, was labeled by an experienced neurosurgeon. From each participant, a 10-second long sEMG sequence was collected from his/her bicep at a sampling rate of 1000 Hz. Out of each sEMG sequence, around seventy different sub-sequences were cropped out. Each sub-sequence, or sEMG segment, contains 2048 data points. Altogether there are 10000 sEMG segments, among which 8000 are used for training and the remaining 2000 are used for testing. The segments were cropped out carefully so that there is no overlap between a training segment and a testing segment. However, there might be temporal overlap between two training segments or two testing segments. Each sEMG segment is a 2048 dimensional vector

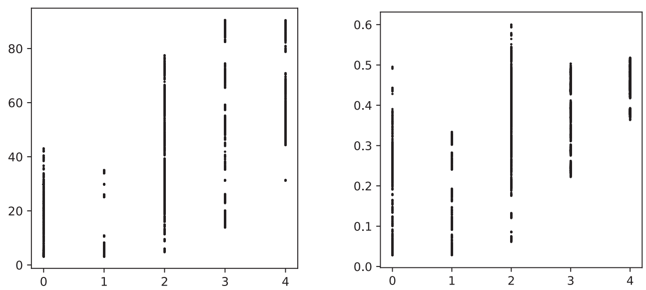
Manuscript received November 20, 2018; revised February 28, 2019; accepted February 28, 2019. Date of publication March 6, 2019; date of current version March 18, 2019. This work was supported in part by the National Natural Science Foundation of China (61673234) and in part by the National Key Research and Development Program of China (2016YFC0105904). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Monica F. Bugallo. *(Zengyi Qin and Zhenyu Jiang contributed equally to this work.) (Corresponding author: Jiansheng Chen.)*

Z. Qin, Z. Jiang, and J. Chen are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: qinzy16@mails.tsinghua.edu.cn; jiangzhe16@mails.tsinghua.edu.cn; jschenthu@tsinghua.edu.cn).

C. Hu is with the Man-Machine-Environment Engineering Institute, School of Aerospace Engineering, Tsinghua University, Beijing 100084, China (e-mail: huchunhua@tsinghua.edu.cn).

Y. Ma is with the Tsinghua University Yuquan Hospital, Beijing 100040, China (e-mail: mayu@tsinghua.edu.cn).

Digital Object Identifier 10.1109/LSP.2019.2903334



(a) Waveform Length (b) Slope Sign Change

Fig. 1. Feature distributions of all the training sEMG segments. The horizontal axis represents clinically observed tremor severity. The dynamic ranges of sEMG features are highly overlapped across different tremor severity levels.

denoted as $\vec{x} = [x_1, x_2, \ldots, x_{2048}]^T$. Considering that the training dataset size may still be too small for efficient feature learning, we use twelve hand-crafted features, including mean absolute value (MAV), mean square value (MSV), root mean square (RMS), variance (VAR), standard deviation (STD), waveform length (WL), Willison amplitude (WAMP), log detector (LOG), slope sign change (SSC), zero crossing (ZC), mean spectral frequency (MSF) and median frequency (MF), to represent $\vec{x}$. Therefore, for each $\vec{x}$ we can obtain a 12 dimensional feature vector $\vec{a} = [a_1, a_2, \ldots, a_{12}]^T$. These features have already been proved effective in sEMG processing [10]–[12]. We adopt feature names and definitions from [13]. For example, log detector is defined in (1).

$$LOG(\vec{x}) = e^{\frac{1}{2048}\sum_{k=1}^{2048} \log|x_k|} \tag{1}$$

III. Methodology

Clinically, the PD tremor severity is evaluated according to the typical symptoms such as the frequency and amplitude of the tremor. However, it is difficult to quantify tremor severity based on the sEMG signal which is usually of low discriminability due to various intrinsic and external interfering factors during capturing. As is shown in Fig. 1, dynamic ranges of sEMG features overlap seriously among different severity levels. This indicates that to effectively classify the tremor severity, the classification model should be flexible enough to both tolerate large intra-class variations and learn the fuzzy boundaries of different classes. However, considering the limited size of the training data, high model capability may lead to serious over-fitting. In our preliminary experiment, a non-linear SVM classifier with Gaussian kernel achieved a training accuracy of 98.20%, while the testing accuracy was only 57.00%.

An alternative approach to address the problem is the similarity learning based classification, of which key idea is to use a set of training samples to vote for the class label of a test sample [9]. There are two main advantages of this approach. Firstly, the multi-class classification task is simplified to a binary classification task, in which the model only estimates whether two sEMG segments belong to the same severity level or not by measuring their similarity. Secondly, since the model takes a pair of sEMG segments as input, the total number of available training samples is increased combinatorially. This facilitates the training of a classifier with high model capacity while reducing the risk of over-fitting.

CNNs are believed to have high capacities for fitting complicated non-linear functions. It has already been successfully used to measure visual similarity between natural images [14], [15]. Inspired by these works, we propose to use CNN for learning the pair-wise similarity between sEMG segments in terms of the PD tremor severity level. Fig. 2 illustrates the architecture of the proposed S-Net, which is a highly light-weighted CNN containing only $0.44M$ parameters. The input to S-Net is a pair of 12-dimensional feature vectors extracted from two sEMG segments as is shown in (2). The 2×12 input is reshaped to a $12 \times 1 \times 2$ tensor, which can be considered as a 12×1 "image" with 2 channels.

$$\begin{bmatrix} \vec{a}^T \\ \vec{b}^T \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_{12} \\ b_1 & \cdots & b_{12} \end{bmatrix} \tag{2}$$

Different from most CNN architectures, the first layer of S-Net, namely "FN", is not trainable. It implements a feature-wise normalization on the input vectors as in (3) required by the great observed differences in the data ranges of the selected features. For example, as shown in Fig. 1, the value of the WL feature is generally two orders of magnitudes larger than that of the SSC feature. This may bring unnecessary numerical instability to the network and slow down the model convergence during training. Therefore, we empirically normalized the 12 pairs of feature elements in the FN layer so as to numerically balance their impacts on the following convolutional layers.

$$\vec{p}_{j,FN} = \begin{bmatrix} a_{j,FN} \\ b_{j,FN} \end{bmatrix} = \frac{1}{\sqrt{a_j^2 + b_j^2}} \begin{bmatrix} a_j \\ b_j \end{bmatrix}, j = 1, 2, \ldots, 12 \tag{3}$$

The FN layer is followed by four convolutional layers with kernel size 1×1 and stride 1. RELU activation and batch normalization [16] are used in these convolutional layers. As a whole, the four convolutional layers nonlinearly maps each 2-dimensional vector $\vec{p}_{j,FN}$ to a 64-dimensional vector $\vec{v}_j$. As such, the output of the last convolutional layer consists of twelve 64-dimensional vectors $\vec{v}_j, j = 1, 2, \ldots, 12$.

The last two layers of S-Net are both full connected layers that capture the global information and output two logits y_1 and y_2, which can be transformed into the classification probability using the softmax function in (4) and (5), where $sl()$ stands for the corresponding tremor severity level. During inference, the tremor severity levels of the sEMG segments corresponding to the two input vectors are considered equal if $P(sl(\vec{a}) = sl(\vec{b})) > 0.5$. During training, the cross entropy between the softmax output and the one-hot ground truth classification label is to be minimized.

$$P(sl(\vec{a}) = sl(\vec{b})) = e^{y_1}/(e^{y_1} + e^{y_2}) \tag{4}$$

$$P(sl(\vec{a}) \neq sl(\vec{b})) = e^{y_2}/(e^{y_1} + e^{y_2}) \tag{5}$$

Using S-Net, the tremor severity classification task can be performed by simple voting. Let S_i be a training subset in which all the sEMG segments are of the same tremor severity level $i, i = 0, 1, \ldots, 4$. In other words, $\forall \vec{a}_l \in S_i, sl(\vec{a}_l) = i$. Suppose the number of samples in S_i is n_i, then we define the likelihood that a test sample $\vec{b}$ is of the tremor severity level i in (6), in which $\mathbb{I}()$ is the indicator function. Consequently, the tremor severity level corresponding to the highest likelihood is regarded as the

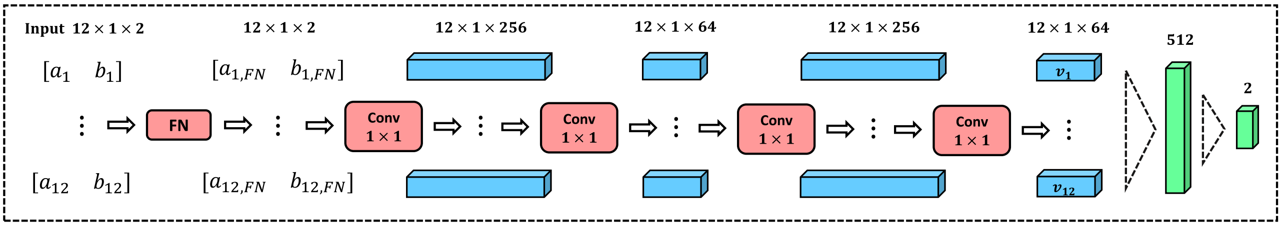


Fig. 2. The proposed S-Net consists of a feature-wise normalization layer, 4 convolutional layers and 2 fully connected layers, where the outputs are two logits y_1 and y_2. Its structure is adjustable. By reducing the input channels to 1 and increasing the number of output logits to 5, we can obtain S-Net_mod that takes only a single feature vector $\vec{a}$ as input, directly performs multi-class classification and outputs the tremor severity level of $\vec{a}$.

TABLE I
CLASSIFICATION RESULTS OF DIFFERENT METHODS

	Recall (%)					Precision (%)					Overall
	Class 0	Class 1	Class 2	Class 3	Class 4	Class 0	Class 1	Class 2	Class 3	Class 4	Accuracy (%)
Bayes [8]	72.25	47.50	55.25	22.00	52.50	46.69	63.33	56.09	29.73	53.71	49.90
MLP	70.25	25.50	64.50	71.00	53.50	50.27	46.58	67.89	54.62	66.46	56.95
KSVM	72.50	40.75	46.00	75.00	50.75	63.46	45.15	48.04	60.12	67.67	57.00
KNN	85.49	16.75	70.00	68.75	50.75	45.97	65.05	49.12	77.46	89.04	58.35
MLP+Similarity Learning	67.00	68.50	26.75	73.50	**99.50**	62.04	56.61	84.92	81.00	66.90	67.05
S-Net_mod	**89.50**	63.25	**90.00**	70.00	93.75	68.19	82.95	73.62	91.80	**99.73**	83.14
Proposed	81.00	**93.25**	85.00	**94.00**	**99.50**	**85.49**	**87.55**	**90.43**	**96.90**	92.34	**90.55**

classification result of $\vec{b}$ as in (7).

$$\mathrm{lkh}(sl(\vec{b}) = i) = \frac{1}{n_i} \sum_{l=1}^{n_i} \mathbb{I}(P(sl(\vec{b}) = sl(\vec{a}_l)) > 0.5) \tag{6}$$

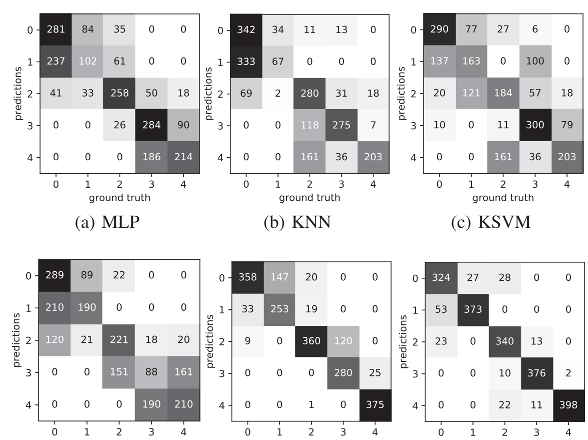
$$sl(\vec{b}) = \arg\max_i \ \mathrm{lkh}(sl(\vec{b}) = i), \ i = 0, 1, \cdots, 4 \tag{7}$$

IV. EXPERIMENTS

The proposed S-Net is implemented using Tensorflow [17]. The loss function consists of the softmax cross-entropy loss mentioned above and an L_2 regularization loss. The penalty coefficient of the L_2 regularization loss is set to 10^{-5} for the convolution layers and 10^{-7} for the fully connected layers. During training, each batch contains 1000 pairs of sEMG features randomly chosen from the training set. The loss function is minimized using the Momentum [18] optimizer at a constant learning rate of 10^{-4} for 680 K iterations. The training takes 8 hours on a single Tesla K80 GPU. In testing, a single forward pass with batch size 1000 takes 8.26 ms.

A. Comparison With Existing Methods

To verify the effectiveness of our method, we implemented several multi-class classifiers to directly classify sEMG signals into 5 tremor severity levels, including a multiple layer perceptron (MLP), a K-nearest neighbor classifier (KNN), a SVM with Gaussian kernel (KSVM), the naive Bayes model in [8] (Bayes), and a modified version of S-Net (S-Net_mod, see the caption of Fig. 2) in which the input and output sizes were changed to enable multi-class classification. The classification accuracies and confusion matrices on the test set are shown in Table I and Fig. 3.



(a) MLP (b) KNN (c) KSVM
(d) Bayes (e) S-Net_mod (f) Proposed

Fig. 3. Classification confusion matrices. There are 2000 test samples evenly distributed among the 5 tremor severity level classes.

The overall accuracies of the MLP, KNN, KSVM and Bayes models are very low and the predictions made by these models can be highly confusing. For example on the test set, when the predicted tremor severity of a sEMG segment is 2, the corresponding ground truth can be of any value. By combining S-Net and the similarity-based classification strategy, the proposed method outperforms these methods by significant margins. Such performance improvement mainly comes from two aspects. Firstly, CNN has a high expressiveness and is suitable for modeling complex nonlinear functions. Directly using S-Net_mod for multi-class classification already achieves a noticeable accuracy improvement. Secondly, the similarity learning strategy further enhances the generalization capability of the classifier. It can be observed from Fig. 3(e) that the S-Net_mod tends to underestimate the tremor severity. For a ground truth severity level i, over 88% of the false predictions are under i.

TABLE II
CLASSIFICATION RESULTS FOR S-NETS OF DIFFERENT KERNEL SIZES

Kernel	Avg. Recall (%)	Avg. Precision (%)	Params	FLOPs
1×1	**90.55**	**90.54**	0.44M	0.99M
3×1	85.35	85.91	0.54M	2.18M
5×1	83.90	85.88	0.64M	3.37M
7×1	83.85	84.62	0.74M	4.57M
9×1	84.40	85.65	0.84M	5.76M

With the similarity learning and voting strategy, prediction errors are much less biased as is shown in Fig. 3(f). In addition, by combining with similarity learning, the overall accuracy of MLP also increased by 10.10%, as is shown in Table I.

B. Architecture Design Analysis

In S-Net, 1×1 convolution kernels are used in all the convolutional layers. To justify such a design, we empirically analyzed the impact of the kernel size on the classification performance. While keeping the overall architecture of S-Net unchanged, we modified the size of the convolution kernels to 3×1, 5×1, 7×1 and 9×1 respectively, so that four different versions of S-Net were generated. Considering that the width of the input tensor is 1, it is unnecessary to use wider kernels.

Classification experiments were performed using these four models respectively. Table II lists the average precisions and recalls for S-Net models with different kernel sizes. For each model, the total number of parameters (Params) and the number of floating point operations (FLOPs) of one inference pass are also shown. Obviously, enlarging the convolution kernel brings about increases in the space and time complexity. However, the classification performances degrade when using large kernel sizes. In computer vision, convolution kernels larger than 1×1 have been successfully applied in image classification for exploring correlations among neighboring pixels in natural images. Nevertheless, the twelve selected input features are not necessarily correlated with each other. For example, the RMS feature and ZC feature can be totally uncorrelated by definition. Therefore, it is expected that enlarging the kernel size brings no performance gain. The performance degradation, however, is probably caused by the increased risk of model over-fitting due to larger parameter sizes. As such 1×1 kernels are suitable for building a light-weight model while achieving high performances in this task.

C. Implications of sEMG Features

The twelve input features have their own physical significances and reflect different characteristics of the muscle movement. By investigating the network's attention on different features , it is possible to estimate influences of different sEMG feature on the classification result. In S-Net, each input feature pair (a_j, b_j) is normalized to a unit vector $(a_{j,FN}, b_{j,FN})$, which can also be represented by its its included angle to the horizontal axis $\theta_j, j = 1, 2, \ldots, 12$. We calculated the gradient of the classification result w.r.t. θ_j. The normalized magnitudes of the gradients for different features are shown in Fig. 4. Numerically, the MF (Median Frequency) feature plays the most important role in determining the tremor severity level prediction. This complies well with recent clinical findings on the significance of sEMG frequency on the PD evaluation [19]–[21]. Meanwhile, features such as ZC, VAR and MSF are also very informative. To verify the above implications, we chose only part of the features that are statistically more important than others as network input and evaluate the classification accuracy. As is shown in Table III, a classification accuracy as high as 89.15% can be achieved by ignoring the five less significant features (WAMP, STD, SSC, MSV, MAV). This indicates that for real-time applications, it is possible to obtain comparable results with less computation cost by choosing more informative features. We hope these results can provide insights for future studies in sEMG feature extraction and analysis.

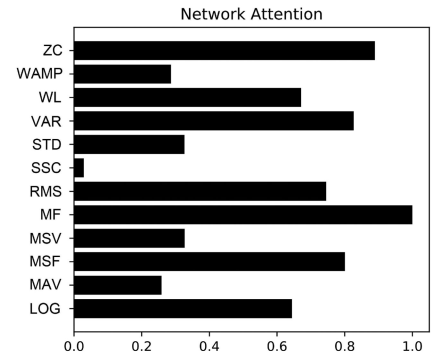


Fig. 4. The calculated gradients of the 12 pairs of input features to the classification score output by the last FC layer of S-Net. Results are normalized for better visualization.

TABLE III
CLASSIFICATION USING ONLY A PART OF FEATURES

Features							Overall
ZC	MF	VAR	MSF	RMS	WL	LOG	Accuracy (%)
✓	✓	✓	✓				82.85
✓	✓	✓	✓	✓			83.15
✓	✓	✓	✓	✓	✓		86.45
✓	✓	✓	✓	✓	✓	✓	**89.15**

V. CONCLUSIONS

In this letter, we study the quantification of PD tremor severity using sMEG signals only. We adopt the widely accepted MDS-UPDRS [2] evaluation metrics so as to facilitate potential application in the automatic assessment of PD patients. To overcome the problem of low discriminative power of sEMG signals, we propose S-Net, a light-weight and computational efficient CNN that is capable of learning complicated class decision boundaries. The pairwise similarity learning is also utilized to enhance the model generalization capability by simplifying the classification objective and augmenting the training data.

Our future work will lie in the automatic sEMG feature extraction techniques towards tremor evaluation. Although the handcrafted features used in this research have been proved effective, we believe that there is still room of performance improvement by automatically learning features suitable for tremor severity evaluation.

ACKNOWLEDGMENT

The authors would also like to thank Suhua Miao, Rongsong Zhou, and Xumin Yu.

REFERENCES

[1] J. Jankovic, "Parkinsons disease: Clinical features and diagnosis," *J. Neurology, Neurosurgery Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[2] C. G. Goetz *et al.*, "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.

[3] J. Ryu, B. H. Lee, and D. H. Kim, "sEMG signal-based lower limb human motion detection using a top and slope feature extraction algorithm," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 929–932, Jul. 2017.

[4] A. I. Meigal *et al.*, "Novel parameters of surface EMG in patients with Parkinson's disease and healthy young and old controls," *J. Electromyography Kinesiology*, vol. 19, no. 3, pp. E206–E213, 2009.

[5] R. M. G. Tello, T. Bastos-Filho, A. Frizera-Neto, S. Arjunan, and D. K. Kumar, "Feature extraction and classification of sEMG signals applied to a virtual hand prosthesis," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 1911–1914.

[6] A. Jha and M. Sen, "A study on the effect of age, gender and paralysis on sEMG signals," Oct. 1, 2015, arXiv:1510.04113.

[7] P. Bartuzi, T. Tokarski, and D. Roman-Liu, "The effect of the fatty tissue on EMG signal in young women," *Acta Bioengineering Biomechanics*, vol. 12, no. 2, pp. 87–92, 2010.

[8] B. T. Cole, S. H. Roy, C. J. De Luca, and S. H. Nawab, "Dynamical learning and tracking of tremor and dyskinesia from wearable sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 982–991, Sep. 2014.

[9] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," *J. Mach. Learn. Res.*, vol. 11, pp. 1109–1135, Mar. 2010.

[10] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.

[11] K. Englehart, B. Hugdins, and P. Parker, "Multifunction control of prostheses using the myoelectric signal," in *Intelligent Systems and Technologies in Rehabilitation Engineering*, New York, NY, USA: CRC Press, 2000, pp. 153–208.

[12] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003.

[13] R. H. Chowdhury, M. B. Reaz, M. A. B. M. Ali, A. A. Bakar, K. Chellappan, and T. G. Chang, "Surface electromyography signal processing and classification techniques," *Sensors*, vol. 13, no. 9, pp. 12431–12466, 2013.

[14] S. Appalaraju and V. Chaoji, "Image similarity using deep CNN and curriculum learning," 2017, arXiv:1709.08761.

[15] S. Bell and K. Bala, "Learning visual similarity for product design with convolutional neural networks," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 98:1–98:10, 2015.

[16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[17] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org/

[18] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.

[19] M. M. Mehndiratta, M. Kumar, S. Pandey, and D. Sharma, "Clinical and surface electromyography (sEMG) correlation in tremors," *Neurology*, vol. 82, no. 10 Suppl., pp. P4–068, 2014.

[20] L. V. Metman and K. V. Slavin, "Advances in functional neurosurgery for Parkinson's disease," *Movement Disorders*, vol. 30, no. 11, pp. 1461–1470, 2015.

[21] R. Fekete and J. Jankovic, "Revisiting the relationship between essential tremor and Parkinson's disease," *Movement Disorders*, vol. 26, no. 3, pp. 391–398, 2011.