Proceedings Paper:

Bao, T, Zaidi, A, Xie, S et al. (1 more author) (2019) Surface-EMG based Wrist Kinematics Estimation using Convolutional Neural Network. In: 2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN). 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN 2019)), 19-22 May 2019, Chicago, IL, USA. IEEE . ISBN 978-1-5386-7477-2

https://doi.org/10.1109/BSN.2019.8771100

Surface-EMG based Wrist Kinematics Estimation using Convolutional Neural Network

Tianzhe Bao, Ali Zaidi, Shengquan Xie, Zhiqiang Zhang
School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, UK
Email address: {eltb, S.A.Zaidi, S.Q.Xie, Z.Zhang3}@leeds.ac.uk

***Abstract*—In the past decades, classical machine learning (ML) methods have been widely investigated in wrist kinematics estimation for the control of prosthetic hands. Currently deeper structures have shown great potential to further improve prediction accuracy. In this paper we present a single stream convolutional neural network (CNN) for mapping surface electromyography (sEMG) to wrist angles within three degrees-of-freedom (DOFs). Two types of two dimensional (2D) sEMG images are constructed in time domain and spectrum as CNN inputs, respectively. Six typical linear and nonlinear ML models are implemented for comparison, where four efficient time-spatial hand-crafted features are extracted to represent feature engineering. Experiment results with four able-bodied participants illustrate that CNN with 2D spectrum sEMG images can achieve highest accuracy in most testing sessions. In other sessions, it is still competitive to the most promising ML techniques. The core strength of deep learning (DL), i.e. feature learning via deep structures and efficient algorithms, is verified to be more powerful than classical feature engineering, particularly in smaller datasets.**

Keywords*—*sEMG, wrist kinematics estimation, machine learning, deep learning, convolutional neural networks

I. INTRODUCTION

It is reported that each year approximately 5000 new patients are suffering from upper limb amputation in UK [1]. To help amputees reconstruct lost functions of hands in activities of daily living (ADLs), active prostheses driven by surface electromyography (sEMG) are now widely applied. Currently the majority of commercial prosthetic hands are using simple but unintuitive on/off control strategy. For more functional motion estimations, classifier-based pattern recognition (PR) approaches have been centred on to discriminate hand and wrist gestures from multi-channel sEMG. However, classification methods can only allow for sequential and discrete control schemes [2]. Alternatively regression-based approaches are proposed for continuous joint angles to enable simultaneous and proportional control of all degrees of freedoms (DOFs) in a natural way.

In past two decades, numerous semi-unsupervised and supervised, linear and nonlinear classical regression methods have been investigated in both offline and online scenarios [2-4], and a variety of feature selection methods [5] have also been exploited to obtain more discriminative model inputs. Although plenty of promising results have been achieved in laboratory settings, few practical implementations have shown up yet, mainly due to the inadequate prediction accuracy particularly during complex muscle contractions [3]. Classical ML techniques rely deeply on the manual feature extraction, analysis and selection, i.e. feature engineering. This process requires expertise in sEMG properties and muscle functions, and remains challenging since some useful information may be easily buried in manual features [6]. Combination of different types of ML models and extracted features also requires excellent domain knowledge or experimental experience.

Recently deep learning (DL) techniques, especially deep neural networks, have shown more promising results not only in computer vision but also in bioinformatics [7]. The key advantage of DL can be summarized as feature learning which derives informative, representative and transferable data-dependent features from raw data automatically. This process, also known as feature learning, is considered to be more efficient than classical feature engineering. Therefore, explorations on DL, mainly the convolutional neural networks (CNN), are also becoming a trend in myoelectric control [6, 8-10]. However, compared with numerous literatures in PR approaches, studies of DL for regression-based schemes are quite inadequate.

In this paper, we exploit a single stream CNN as a representative of DL for wrist kinematics estimation in 3 DOFs. Sparse multi-channel sEMG signals are processed to construct two types of two dimensional (2D) images as model inputs. Small size of rectangular kernels are utilized in convolution layers to achieve a deeper model structure with Leaky Rectified Linear Units (leaky ReLU) for effective nonlinear projection. As baselines, six typical shallow ML models, i.e. linear regression (LR), multi-layer perceptron neural network (MLP), support vector regression (SVR) with linear kernel (LSVR) and radial basis function (RBF-SVR), random forest (RF) and Gaussian process regression (GPR) are implemented with four widely applied time-spatial manual features. The superiority of feature learning achieved by CNN over classical feature engineering can be verified clearly in experimental results.

II. METHODS

A. Date Collection

This study was approved by the MaPS and Engineering joint Faculty Research Ethics Committee of University of Leeds, UK (reference MEEC 18-006). Four able-bodied subject (3 male and 1 female, age 24-28) took part in the data acquisition. A written informed consent was obtained from each subject before data collection. Recruited subjects were asked to perform a series of smooth wrist movements lasting around 3 minutes in a low speed (frequency $\approx$ 0.1Hz for sinusoidal contractions). In first three test sessions (S1-S3) motions in only one DOF, i.e. flexion/extension (#1DOF), wrist pronation/supination (#2DOF) or radial/ulnar deviation (#3DOF), were activated to complete single-DOF tasks. For multi-DOF tasks, all 3 DOFs were involved simultaneously and casually in the last session (S4). A detailed description of contractions is reported in Table 1.

TABLE I. LIST OF THE CONTRACTIONS PERFORMED

Session	Description	Active DOF
S1	Sinusoidal contractions	#1DOF
S2	Sinusoidal contractions	#2DOF
S3	Half sinusoidal contractions	#3DOF
S4	Casual contractions of unconstrained wrist movements	#1DOF + #2 DOF + #3DOF

During experiments, participants seated comfortably in front of a computer screen. The tested hand should be kept in a relaxing state to avoid muscle fatigue, with palm facing inside. As shown in the Fig. 1 (a), twelve bipolar electrodes were placed on the proximal portion of the left forearm to collect six channels of sEMG signals via Shimmer sensors. Reference electrodes were placed near the wrist. The inter-electrode distance in the proximal-distal direction was around 20 mm to reduce the crosstalk effect. For wrist angles as ground truth, 9-DOFs IMU signals were recorded to calculate the hand orientation. Sampling rates for IMU units and sEMG were set as 100Hz and 1024Hz respectively. The data streaming, synchronization and display were implemented online in a home-made software, as is illustrated in Fig. 1(b).

B. Pre-processing of sEMG

The acquired sEMG signals were processed using a 3rd order Butterworth high pass filter (20 Hz) and a low pass filter (450 Hz) to remove unusable noise. A Min-Max scaling [11] was applied to normalize raw sEMG in each channel. This process aimed at enhancing the deep network to learn optimal weights and bias for each input mode more efficiently. For data segmentation, the analysis window was set to be 100ms with 25ms increment to supply a comparatively larger dataset for DL models. Since subjects were asked to rotate the wrist in a comparatively low speed, angles in each data segment can be averaged as data labels.

C. CNN Image Construction

To apply CNN, the filtered multi-channel signals in one data segment should be shaped into an $N \times L$ sized rectangular sEMG image, where N is the number of channels and L is the length of data in one segment. Herein N=6 due to six channels in experiments and L=102 corresponding to the sampling rate and window length. According to previous research, two types of sEMG images have been most widely applied, i.e. the 2D time domain image [9] and 2D spectrum image [10]. The latter is commonly converted from the former via fast Fourier transform (FFT). To our best knowledge, no conclusion has yet been drawn on the superiority of these two types of images in either PR or regression scheme in myoelectric control.

D. CNN Architecture

Although more complicated structures might own greater potentials in feature learning, they tend to have more challenges in model optimization. So far there has yet to be any consensus on the optimal structure of CNN, particularly in wrist angles estimation. In this study, a single stream architecture is implemented for a fast training process. According to Fig. 2, the presented CNN consisted of four convolutional blocks (Conv Block) and three fully connected blocks (FC Block). A max pooling layer and a dropout layer were added between the 4th Conv Block and 1st FC Block.

(a) (b)

Fig. 1. Electrodes placement (a) and online data acquisition (b).

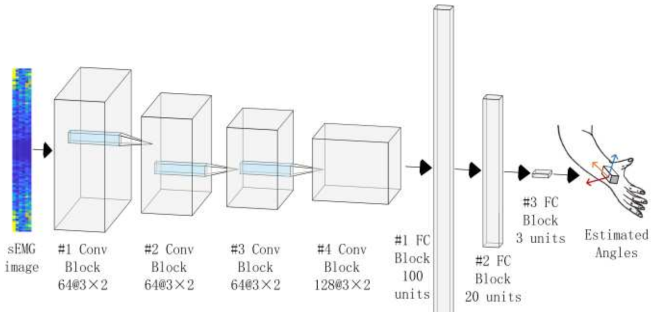

Fig. 2. The single stream CNN architecture for wrist angles estimation.

In first six blocks, a batch normalization layer worked to minimize the influence of weights initialization. As wrist angles are distributed in both positive and negative fields, the leaky ReLU layers were applied for nonlinear projection instead of classical ReLU layers which are easy to suffer from a zero gradient for negative inputs [12]. The leaky ReLU is described mathematically in Eq. (1). In four convolutional layers, 64, 64, 64 and 128 rectangular filters in size 2×3 were utilized with the padding strategy. The small filter size is chosen to construct a deeper model structure for a more potential feature learning capability. There were 100 hidden units in the 1st fully connected layer, 20 in the 2nd and 3 in the 3rd for regressions in 3 DOFs.

$$h_i = \begin{cases} \mathbf{w}_i \mathbf{x}_i & \mathbf{w}_i \mathbf{x}_i \geq 0 \\ a_i \mathbf{w}_i \mathbf{x}_i & \mathbf{w}_i \mathbf{x}_i < 0 \end{cases} \tag{1}$$

where $\mathbf{x}_i$ is the input of ith leaky ReLU layer, $\mathbf{w}_i$ is the weight vector, h_i is the output and a_i is a scale parameter in range (0,1) to assign a noon-zero slope.

E. CNN Training

In our experiments hyper-parameters of CNN were firstly identified referring to studies in PR scheme [8-10] and then determined via empirical manual tuning. The final settings were fixed to experiments in all subjects. The network was trained in a 128 sized mini batch for 50 epochs by stochastic gradient descent. The dynamic learning rate was 0.001 at first and dropped 90% after every 10 epochs. The slope scale was set as 0.1 in all leaky ReLU layers. The dropout rate was 30% for regularization. Other training strategies followed the default settings in deep learning tool box of Matlab 2018a. The data size for training was around 5000 in most sessions. For sake of computational efficiency, the data augmentation by adding white Gaussian noise to the new copied training set [9] was only applied in several sessions with insufficient training samples.

F. Model Evaluation

In this study a 10-fold cross-validation was applied for offline evaluation. Nine out of ten repetitions were used for training with the rest for testing. The R^2 value [13], which is robust to the numerical range of labels, was used as the metric for estimation accuracy in each DOF separately. According to Eq. (2), a higher value corresponds to a better regression performance of the model. The R^2 value at perfect estimation is equal to one, whilst a negative value means that estimation errors are larger than the variance of targets.

$$R^2 = 1 - \frac{Var(\mathbf{Y}^d - \hat{\mathbf{Y}}^d)}{Var(\mathbf{Y}^d)} \tag{2}$$

where $\mathbf{Y}^d$ are angles in dth DOF and $\hat{\mathbf{Y}}^d$ are model estimations.

III. Results and Discussion

A. Wrist Angles Estimation

Referring to protocols in Table 1, S1-S3 were designed for single-DOF regression tasks while S4 for the multi-DOFs one requiring wrist motions activated in all three DOFs simultaneously. Although the latter is naturally more difficult to perform, it is highly emphasized to speed up experiments and simulate real-time scenarios. For a brief demonstration, Fig. 3(a) shows wrist angles of S4 captured via sensors in Subject 4. The motion estimated by CNN with spectrum images in one testing fold is shown in Fig. 3(b), where estimated angles are capable of matching the real data in 3 DOFs effectively. The R^2 values of CNN in #1DOF, #2DOF and #3DOF are 0.75, 0.70 and 0.66 respectively, which are consistent with related results achieved by other techniques in [3-4, 13]. However estimation errors are still noticeable in complex muscle contractions, particularly for pronation and supination. It is because some sEMG gets missed due to the forearm anatomy. Besides, muscle contractions in this DOF can lead to electrode shifts more easily.

B. Comparison between ML Techniques and CNN

In our experiments, we also implemented traditional ML methods for comparison purpose. The data pre-processing of sEMG in ML is similar to that in DL. A slight difference in ML is the full-wave rectification so that all energy of signals is retained for feature extraction. Referring to studies on feature evaluations, a combined time domain (TD) and time-serial domain (TSD) feature set consisting of mean absolute value (MAV), root mean square (RMS), variance (VAR) and 4^{th} order autoregressive coefficients (4^{th} AR) were used after the Min-Max scaling in each channel. To achieve promising projections between sEMG features and joint angles, six widely applied linear and nonlinear models, i.e. LR [2, 13], LSVR, MLP [13], RBF-SVR [14], RF [15] and GPR [16], were selected to work as baselines. In MLP, the Levenberg–Marquardt backpropagation algorithm was used to train the network. The number of hidden neurons was 3 and the learning rate was 0.001. To project the input data into to a higher dimensional feature space efficiently, a linear and RBF function were selected separately as kernels of SVR. For GPR, Squared-Exponential (SE), a universal and clear kernel, was taken as the covariance function for joint distributions. Other hyper parameters of ML followed the default setting of regression toolbox in Matlab 2018.

Fig. 4 illustrates the results of eight techniques in S4 of Subject 1. The R^2 values (mean ± standard deviation of 10 test folds) via the 10-fold random permutation validation in 3 DOFs are shown in three groups separately. Herein several interesting observations can be listed as follows: (1) CNN with spectrum images (CNN_spectrum) achieved the highest accuracy in #1 DOF as 0.77 and in #2 DOF as 0.68, while GPR ranked the first in #3 DOF with 0.79. Both of them outperformed other four techniques dramatically in each DOF. (2) Nonlinear ML models performed much better than linear ones, indicating that intensities of presented manual features along with wrist motions are highly nonlinear. (3) With same structure and hyper parameters, CNN_spectrum showed great superiority to CNN_time (CNN with time domain images) in all DOFs. A possible explanation is that sEMG is measured as superimposition of EMG signals from multiple muscles [9], thus spectrum images should be more representative and distinguishable than time domain ones.

Table 2 summarizes the averaged values of four most prosing regression techniques (RBF-SVR, RF, GPR and CNN_spectrum) in all test sessions of four subjects. The highest value in one session is labelled in bold and italic format. As we can see from the table, CNN_spectrum ranked the first in nearly three quarters of all sessions whilst GPR won the evaluation in the rest. In more than half sessions, the performance of these two techniques was dramatically better than those of MLP, RBF-SVR and RF. Although RF outperformed GPR slightly sometimes (such as the results in S3 of Subject 1, S1-S2 of Subject 3), GPR was much more robust and performed more impressively in estimations during combined muscle contractions such as #2 DOF in S4.

C. Strength of CNN feature learning

The key magic of CNN can be summarized as layer-by-layer data processing and in-model feature transformation contributed by convolutional layers, i.e. feature learning which substitutes feature engineering in ML. To verify its strength with our dataset, raw data in eight of nine training folds in each validation was used to train the CNN_spectrum model first, after which the final outputs of the 2^{nd} fully connected layer worked as CNN features to train six ML models with the rest one training fold. Evaluations were taken in the testing fold. It is noticeable that the data size for training ML models was only 1/9 of that in conventional validations. It approximated a fast fine-tuning process in daily applications of DL. To match the input dimension, the number of hidden units in the #2 FC Block of CNN was set to be 42 with respect to the length of manual feature sets.

Fig. 5 illustrates the results in S4 of Subject 1 where CNN features substitute hand-crafted ones. In this scenario performances of most ML models, particularly the linear ones, were improved sharply in all DOFs. In contrast to the dramatic variation when taking hand-crafted features, six models performed quite equivalently in each DOF. More impressively, even with much less training data, the best mean values, i.e. 0.74 in #1 DOF, 0.64 in #2 DOF and 0.73 in #3 DOF respectively, were very close to the highest ones shown in Fig. 4. Apparently CNN features are somehow linearly distinguishable and more effective for ML models.

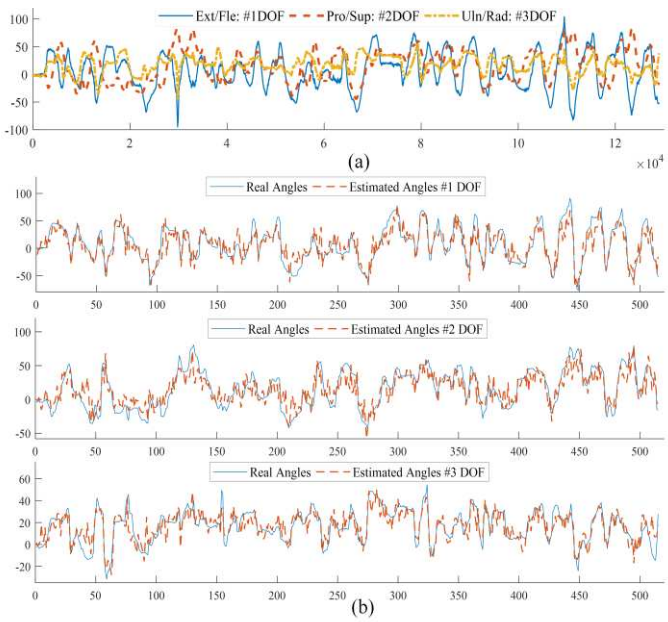

Fig. 3. Wrist motions and estimations. (a) Wrist angles in 3 DOFs in S4 of Sunject 4. (b) Captured angles and CNN estimations in one testing fold.

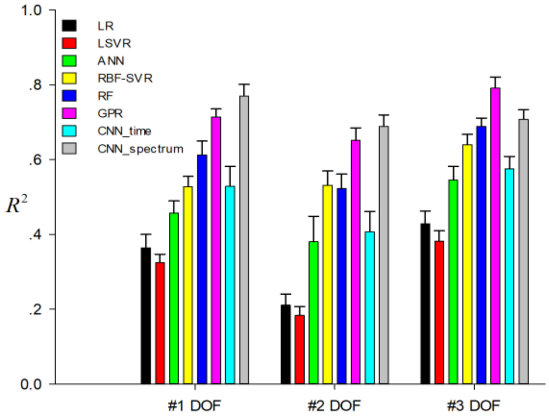

Fig. 4. Evaluations of ML and DL techniques in S4 of Subject 1.

TABLE II. EVALUATIONS OF ML AND CNN IN ALL SESSIONS

Subjects	Sessions		SVR	RF	GPR	CNN_spectrum
1	S1 (#1DOF)		0.783	0.805	0.810	***0.904***
	S2 (#2DOF)		0.557	0.486	0.614	***0.687***
	S3 (#3DOF)		0.772	0.794	0.776	***0.839***
	S4	#1DOF	0.522	0.618	0.720	***0.763***
		#2DOF	0.541	0.496	0.628	***0.686***
		#3DOF	0.668	0.682	0.761	***0.769***
2	S1 (#1DOF)		0.812	0.797	0.818	***0.856***
	S2 (#2DOF)		0.482	0.428	***0.655***	0.595
	S3 (#3DOF)		0.540	0.448	0.630	***0.633***
	S4	#1DOF	0.322	0.374	0.618	***0.664***
		#2DOF	0.456	0.420	***0.643***	0.627
		#3DOF	0.377	0.351	0.578	***0.590***
3	S1 (#1DOF)		0.660	0.731	0.674	***0.832***
	S2 (#2DOF)		0.662	0.738	0.667	***0.855***
	S3 (#3DOF)		0.702	0.616	0.767	***0.826***
	S4	#1DOF	0.603	0.616	***0.771***	0.733
		#2DOF	0.530	0.595	0.614	***0.717***
		#3DOF	0.562	0.581	***0.759***	0.732
4	S1 (#1DOF)		0.742	0.727	0.779	***0.861***
	S2 (#2DOF)		0.392	0.380	***0.582***	0.565
	S3 (#3DOF)		0.766	0.721	0.795	***0.857***
	S4	#1DOF	0.550	0.502	0.718	***0.756***
		#2DOF	0.463	0.419	0.655	***0.676***
		#3DOF	0.570	0.543	***0.752***	0.741

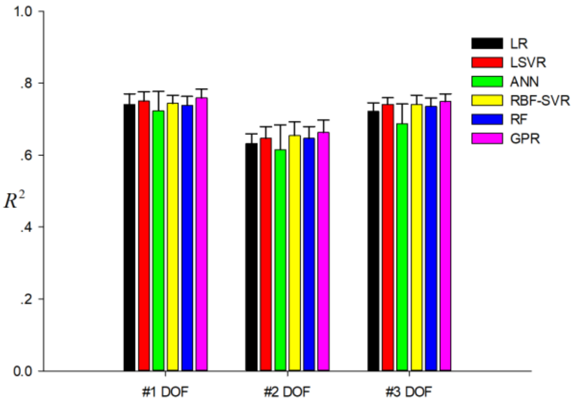

Fig. 5. Performance of ML models with CNN features in S4 of Subject 1

CONCLUSION AND FUTURE WORK

This paper presents a CNN based DL techniques for wrist kinematics estimation in multi-DOFs. We find that spectrum sEMG images are more distinguishable than time domain ones, and that utilization of the former helps CNN to be more competitive than traditional ML techniques. The superiority of feature learning in DL over feature engineering is clearly demonstrated via promising results achieved by ML models in very small datasets with representative CNN features. It also indicates the potential of fast fine-tuning in practice.

Besides, GPR is firstly discovered to be superior to many shallow models in wrist angles estimation, particularly during challenging combined muscle contractions. In contrast to sophisticated optimization in CNN, this nonparametric model can obtain optimal hyper-parameters straightforwardly through maximum likelihood estimation. In fact, there has yet to be a consensus on the superiority of ML and DL in myoelectric control. For deeper insights, a systematic study among ML and DL techniques is expected in future work, where amputees are planned to be recruited and the real-time analysis in dynamic scenarios is also considered.

REFERENCES

[1] F. Cordella, A.L. Ciancio, R. Sacchetti, A. Davalli, A.G. Cutti, E. Guglielmelli and L. Zollo. "Literature review on needs of upper limb prosthesis users." Frontiers in neuroscience, 10, 2016, p.209.

[2] J.M. Hahne, M.A. Schweisfurth, M. Koppe and D. Farina. "Simultaneous control of multiple functions of bionic hand prostheses: Performance and robustness in end users. " Science Robotics, 3(19), 2018, p.eaat3630.

[3] Bakshi, Koushik, M. Manjunatha and C. S. Kumar. "Estimation of continuous and constraint-free 3 DoF wrist movements from surface electromyogram signal using kernel recursive least square tracker." Biomedical Signal Processing and Control 46, 2018, pp.104-115.

[4] C. Lin, B. Wang, N. Jiang and Farina, D. "Robust extraction of basis functions for simultaneous and proportional myoelectric control via sparse non-negative matrix factorization." Journal of neural engineering 15, no. 2, 2018, p.026017.

[5] Z. Li, X. Zhao, J. Han and G. Liu. "Comparisons on different sEMG-features with dimension-reduction methods in hand motion recognition." In Advanced Robotics and Mechatronics (ICARM), International Conference on, pp. 631-634. IEEE, 2016.

[6] M. Zia ur Rehman, A. Waris, S. Gilani, M. Jochumsen, I. Niazi, M. Jamil. "Multiday EMG-based classification of hand motions with deep learning techniques." Sensors 18, no. 8, 2018, p.2497.

[7] S. Min, B. Lee and S. Yoon. "Deep learning in bioinformatics." Briefings in bioinformatics 18, no. 5, 2017, pp.851-869.

[8] W. Wang, B. Chen, P. Xia, J. Hu and Y. Peng. "Sensor Fusion for Myoelectric Control Based on Deep Learning With Recurrent Convolutional Neural Networks." Artificial Organs, 42(9), 2018, pp.E272-E282.

[9] Z. Ding, C. Yang, Z. Tian, C. Yi, Y. Fu and F. Jiang. "sEMG-Based Gesture Recognition with Convolution Neural Networks." Sustainability 10, no. 6, 2018, p.1865.

[10] Y. Yamanoi and R. Kato. "Control method for myoelectric hand using convolutional neural network to simplify learning of EMG signals." In Cyborg and Bionic Systems (CBS), 2017 IEEE International Conference on, pp. 114-118. IEEE, 2017.

[11] L. Al Shalabi, Z. Shaaban and B. Kasasbeh. "Data mining: A preprocessing engine." Journal of Computer Science 2, no. 9, 2006, pp.735-739.

[12] K. Chitturi and P. Onyisi. "How easily can neural networks learn relativity?" In Journal of Physics: Conference Series, vol. 1085, no. 4, p. 042020. IOP Publishing, 2018.

[13] J.M. Hahne, F. Biessmann, N. Jiang, H. Rehbaum, D. Farina. "Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control." IEEE Transactions on Neural Systems and Rehabilitation Engineering 22, 22(2), 2014, pp.269-279.

[14] A. Ameri, E.N. Kamavuako, E.J. Scheme, K.B. Englehart and P.A. Parker. "Support vector regression for improved real-time, simultaneous myoelectric control." IEEE Transactions on Neural Systems and Rehabilitation Engineering 22, no. 6, 2014, 1198-1209.

[15] A. Ameri, E.J. Scheme, K.B. Englehart and P.A. Parker. "Bagged regression trees for simultaneous myoelectric force estimation." In Electrical Engineering (ICEE), 2014 22nd Iranian Conference on, pp. 2000-2003. IEEE, 2014.

[16] J.G. Ngeo, T. Tamei and T. Shibata. "Continuous and simultaneous estimation of finger kinematics using inputs from an EMG-to-muscle activation model." Journal of neuroengineering and rehabilitation 11, no. 1, 2014, p.122.