On the Selection of Neural Network Architecture for Supervised Motor Unit Identification from High-Density Surface EMG*

Filip Urh, *Student Member, IEEE,* Damjan Strnad, Alex Clarke, *Student Member, IEEE,* Dario Farina, *Fellow, IEEE,* and Aleš Holobar, *Member, IEEE*

Abstract— **In the last decade, accurate identification of motor unit (MU) firings received a lot of research interest. Different decomposition methods have been developed, each with its advantages and disadvantages. In this study, we evaluated the capability of three different types of neural networks (NNs), namely dense NN, long short-term memory (LSTM) NN and convolutional NN, to identify MU firings from high-density surface electromyograms (HDsEMG). Each type of NN was evaluated on simulated HDsEMG signals with a known MU firing pattern and high variety of MU characteristics. Compared to dense NN, LSTM and convolutional NN yielded significantly higher precision and significantly lower miss rate of MU identification. LSTM NN demonstrated higher sensitivity to noise than convolutional NN.**

Clinical Relevance— **MU identification from HDsEMG signals offers valuable insight into neurophysiology of motor system but requires relatively high level of expert knowledge. This study assesses the capability of self-learning artificial neural networks to cope with this problem.**

I. Introduction

Neural networks (NNs) are rapidly evolving in the field of pattern recognition. There are multiple reasons for this, but the two most important ones are new mathematical models of NNs and the constantly increasing processing power of computers [1], [2]. NNs have significantly improved the state-of-the-art in pattern recognition, especially in the fields of image processing, speech recognition, and natural language processing [2], [3], [4]. NNs with multiple layers and a complex layer structure have been developed. Among them, convolutional NNs have been very successful in processing of two or more-dimensional input data [5], [6], [7].

Long short-term memory (LSTM) NNs represent another complex NN architecture. They are recurrent NN and can solve the convergence problems that simple recurrent NNs suffer from. This makes LSTM NNs suitable for use in speech recognition, handwriting recognition and text-to-speech synthesis, among other applications [8], [9].

Dense NNs have been extensively used in the analysis of surface electromyograms (EMG) and in human-machine interfaces for classification of different movements from EMG signals [10], [11]. On the other hand, their use in decomposition of EMG signals into contributions of individual motor units (MUs) is still relatively scarce, at least to the best of our knowledge.

Indeed, identification of MU firing from surface EMG proved to be a difficult computational problem. First, reliable MU identification requires recording of EMG signals by several tens of uptake electrodes – so called high-density surface EMG (HDsEMG) [12]. Second, existing HDsEMG decomposition methods are based on relatively complex mathematical models and iterative optimizations of MU filters. When applied to the HDsEMG, these filters yield spike trains of individual MUs, with spikes denoting MU firings. In fact, each MU filter defines a linear combination of spatio-temporal HDsEMG samples that yields a MU spike train. As the spike train of an individual MU is more sparse that the merged trains of two or more MUs, the MU filters get iteratively optimized by increasing the non-Gaussianity of estimated MU spike train [13]. Thus, MU filters can also be estimated by NN, but require reinforced learning as MU firings are not a priori known.

When building a MU filter, the existing decomposition techniques take HDsEMG samples from local temporal support of a few ms. This support is defined by the so called extension factor F and is equal for all the HDsEMG channels, regardless of the actual information in the channel [13]. On the other hand, recurrent NN are theoretically capable of combining HDsEMG samples from much larger temporal supports [14] but it is not known whether this improves the estimation of MU spike trains.

In this study, we compared recurrent and non-recurrent NNs, namely dense NN, LSTM and convolutional NN and assessed their capability to learn MU firings from the HDsEMG signals in supervised fashion. For this purpose, we assumed that several (but not all) firings of individual MU are a priori known. To ensure the controlled test conditions and reliable reference data, we limited our study to synthetic HDsEMG signals that were generated by the advanced cylindrical volume conductor model described in [15].

II. Neural Networks

Three types of NNs were evaluated on HDsEMG signals that were temporarily extended by adding F-1 delayed repetitions of each HDsEMG channel [13]:

$$\boldsymbol{y}(n) = [y_1(n) \dots y_1(n-F+1) \dots y_M(n-F+1)]^T \quad (1)$$

By following the recommendation in [13], the extension factor F was set to 10. Afterwards, the correlation matrix of extended HDsEMG measurements was calculated

* This study was supported by the Slovenian Research Agency (projects J2-7357, J2-1731 and L7-9421 and Programme funding P2-0041) and by the European Research Council (project NaturalBionicS #810346).

F. Urh, D. Strnad, and A. Holobar are with the Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia (corresponding author: filip.urh1@um.si).

A. Clarke, and D. Farina are with the Department of Bioengineering, Imperial College London, London, United Kingdom.

$$\mathbf{C}_{\boldsymbol{y}} = E(\boldsymbol{y}(n)\boldsymbol{y}^T(n)) \quad (2)$$

where $E(.)$ stands for mathematical expectation. The extended measurements were whitened [13]:

$$\boldsymbol{z}(n) = sqrtm(\mathbf{C}_{\boldsymbol{y}}^{-1})\boldsymbol{y}(n) \quad (3)$$

with *sqrtm(.)* function denoting the matrix square root. The generated HDsEMG signals had 90 channels. After the extension with a factor F=10, each sample of $\boldsymbol{z}(n)$ comprised 900 values.

A. Dense NN

The feedforward dense NN used in this study received 1 sample of $\boldsymbol{z}(n)$ per iteration as an input. The NN comprised 5 dense layers, each followed by a dropout layer, except for the last one. The detailed structure of dense NN is presented in Fig. 1.

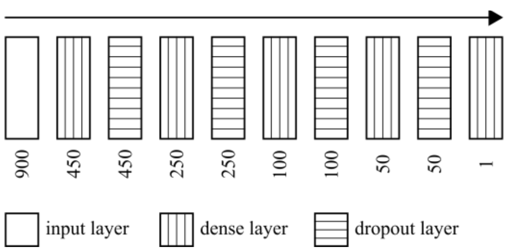

Figure 1. Layers of dense NN, named as in the Tensorflow machine learning package [16]. The output size of each layer is written below its graphical representation.

B. LSTM NN

The LSTM NN received 15 samples of $\boldsymbol{z}(n)$ per iteration (a matrix with dimensions 900 × 15) as an input. The NN comprised 2 LSTM layers and 3 dense layers, each followed by a dropout layer, except for the last one. The detailed structure of LSTM NN is depicted in Fig. 2.

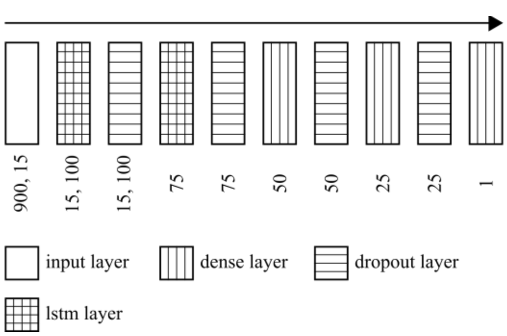

Figure 2. Layers of LSTM NN, named as in the Tensorflow machine learning package [16]. The output size of each layer is written below its graphical representation.

C. Convolutional NN

Convolutional NN received 15 samples of $\boldsymbol{z}(n)$ per iteration (a matrix with dimensions 900 × 15) as an input. It comprised 2 convolutional layers, each followed by a max pooling layer and a dropout layer. It also contained a flatten layer and 3 dense layers, each followed by a dropout layer, except for the last one. The detailed structure of the convolutional NN used in this study is presented in Fig. 3.

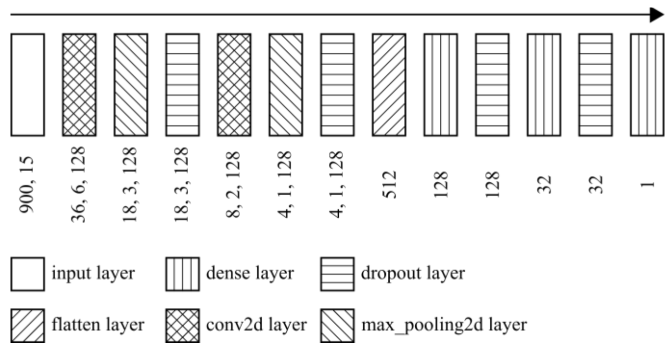

Figure 3. Layers of convolutional NN used in this study, named as in the Tensorflow machine learning package [16]. The output size of each layer is written below its graphical representation.

In all the presented NNs and each NN layer we used hyperbolic tangent activation function. The output values of this function are from the interval (-1, 1), so after the last layer the round function was used to translate the NN outputs to binary values that indicate MU firings. Adam optimizer [17] was used for training. Its initial learning rate was set to 0.001, whereas parameters β_1, β_2 and ε were set to 0.9, 0.999 and 10^{-7}, respectively.

To measure the error between the real and predicted MU spike train, mean squared error was used as a loss function. The batch size was set to 128. The number of training epochs was set to 2000, but the training was interrupted for less epochs if there was no improvement in the loss metric on the validation dataset for 50 consecutive iterations.

III. HDsEMG Simulation and Data Analysis

A. Simulated HDsEMG Signals

Selected NNs were evaluated on synthetic signals with known MU firing patterns. The biceps brachii muscle with randomly distributed 200 active MUs was simulated [18], using the cylindrical volume conductor model [15]. The simulated muscle with elliptical cross-section was 130 mm long, 30 mm wide and 15 mm deep. Fiber density was set to 20 fibres/mm^2 [18]. All the muscle fibers belonging to the same MU shared the same conduction velocity, which was normally distributed (4.0 ± 0.3 m/s) across simulated MUs. The size of MUs ranged from 24 to 2408 fibers and was distributed according to the Henneman's size principle [19]. The innervation zone was in the longitudinal center of the simulated muscle, with the spread set to 5 mm. The simulated skin and subcutaneous fat layer were 1 mm and 4 mm thick, respectively.

Firing pattern of each MU was generated with the model proposed in [20]. Muscle excitation level was set to 30 %, which resulted in 155 active MUs. The 50 s long HDsEMG signals were sampled at 2048 samples/s and detected by an array of 10×9 circular electrodes with 1 mm diameter and 5 mm interelectrode distance. The array was centered over the simulated muscle.

Four simulations of HDsEMG signals were conducted. In each simulation, MUs were randomly distributed in the muscle tissue. In order to evaluate NNs at different noise levels, noise with SNR of ∞, 30 and 20 dB was added. Generated HDsEMG signals are exemplified in Fig. 4.

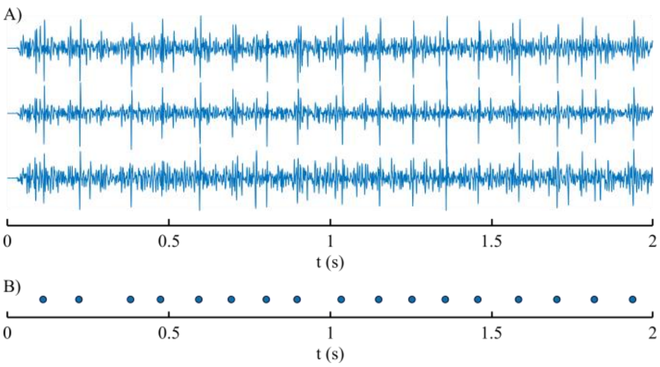

Figure 4. A) Representative example of synthetic HDEMG signal at excitation level of 30 % and SNR of 30 dB, used on the input of tested NNs. B) Representative example of MU firings at the output of the tested NN. For clarity reason, only 2 s of each signal are depicted.

Training, validation and test datasets were created from HDsEMG epochs with the length of 25 s, 12.5 s and 12.5 s, respectively. On average, there were 710 ± 195, 355 ± 97 and 351 ± 98 MU firings in these datasets, respectively.

To limit the computational time and increase the efficiency of NN learning, the spike trains of simulated MUs were first assessed by Linear Minimum Mean Square Error (LMMSE) estimator [13]. This estimator is Bayesian optimal among the linear estimators, but can only be applied to synthetic HDsEMG data, as it uses the simulated MU firing patterns to learn MU filter. Quality of MU spike trains, estimated by LMMSE was assessed by the previously proposed Pulse-to-Noise Ratio (PNR) [21]:

$$PNR\left(\hat{t}_{\mathrm{j}}(n)\right) = 10 \cdot \log\left(\frac{E\left(\hat{t}_j^2(n)\big|_{\mathrm{t_j}(n)=1}\right)}{E\left(\hat{t}_j^2(n)\big|_{\mathrm{t_j}(n)=0}\right)}\right), \tag{4}$$

where $\hat{t}_{\mathrm{j}}(n)$ is the estimated j-th MU spike train, $\hat{t}_j^2(n)\big|_{\hat{t}_j(n)=1}$ is the energy of MU spikes and $\hat{t}_j^2(n)\big|_{\hat{t}_j(n)=0}$ is the energy of noise in the spike train. Only MUs that were estimated from HDsEMG signals by LMMSE and exhibited the PNR value above 28 dB [21] were used for NN training. As a result, 49 ± 5, 9 ± 2 and 5 ± 3 MUs per generated HDsEMG signal were used for NN training at SNR of ∞, 30 and 20 dB, respectively.

B. Data analysis

The performance of NN was assessed by identifying true positive (TP), false positive (FP) and false negative (FN) firings with the tolerance set to 0.5 ms. Afterwards, precision (Pr) and miss rate (MR) were calculated for each identified MU:

$$Pr = \frac{TP}{TP+FP}, \quad MR = \frac{FN}{FN+TP}, \tag{5}$$

In more than 67 % of MUs, the normal distribution of Pr and MR was rejected by Lilliefors test. Consequently, Friedman test with Bonferroni correction and p-value < 0.05 was used for statistical comparison of results of different NNs.

IV. Results

The results of NNs evaluation are exemplified in Fig. 5 and reported in Table I.

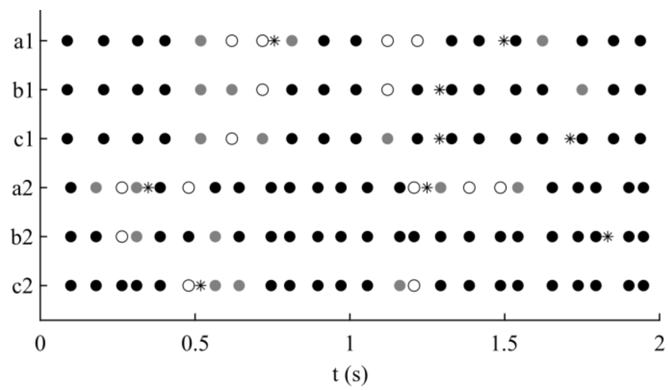

Figure 5. Firings of two MUs (1, 2), identified by dense (a), LSTM (b) and convolutional NN (c). Filled circles represent TP firings, empty circles FN firings, and stars FP firings. Gray circles represent a TP firing with identification tolerance set to 5 samples. For clarity reason, only 2 s of MU firing patterns are depicted.

V. Conclusion

All the tested NNs demonstrated relatively high precision in MU identification, however the miss rates were not negligible, especially when noise was added to the HDsEMG signals (Table I). LSTM and convolutional NN had significantly higher precision and significantly lower miss rate than dense NN. On the other hand, LSTM and convolutional NN did not produce significantly different results, except in the case of miss rates at SNR of 20 dB, where convolutional NN significantly outperformed the LSTM.

The extension factor F is commonly used in HDsEMG decomposition algorithms to improve the conditionality of the HDsEMG mixing process [13]. With its increase, the size of the input data to the NN also increases. Consequently, there are more learning parameters in the first NN layer. This potentially leads to slower learning, but potentially improves the decomposition performance [13]. In our study, the preselected temporal support of F = 10 HDsEMG samples proved to be large enough and, when compared to other two tested NNs, the recurrent connections in LSTM did not significantly increase the HDsEMG decomposition performance. On the contrary, it even seems that they increased the sensitivity of NN to noise (Table I). Further testing of different extension factors F exceeds the scope of this study.

TABLE I. PRECISION (PR) AND MISS RATE (MR) FOR DENSE, LSTM AND CONVOLUTIONAL NN AT SNR OF ∞, 30 AND 20 DB.

SNR (dB)	NN type	Metric	
		Pr (%)	MR (%)
∞	Dense	95 ± 4[b, c]	19 ± 16[b, c]
	LSTM	99 ± 1[a]	7 ± 6[a]
	Convolutional	99 ± 1[a]	7 ± 7[a]
30	Dense	92 ± 8[b, c]	27 ± 26[b, c]
	LSTM	99 ± 2[a]	16 ± 8[a]
	Convolutional	99 ± 3[a]	16 ± 10[a]
20	Dense	90 ± 6[b, c]	39 ± 31[b, c]
	LSTM	97 ± 5[a]	26 ± 20[a, c]
	Convolutional	98 ± 3[a]	23 ± 14[a, b]

[a] – significantly different from Dense NN; [b] – significantly different from LSTM NN; [c] – significantly different from Convolutional NN.

The speed of convergence was faster for LSTM and convolutional NN than for dense NN. The average number of training epochs was 1273 ± 296, 881 ± 204 and 972 ± 317 for dense, LSTM and convolutional NN, respectively.

One of the limitations of this study was relatively small learning dataset used for NN training. All the tested NNs were composed of multiple layers that require large learning sets. With smaller learning sets, NN fitting becomes easier, but the knowledge generalization deteriorates significantly. In HDsEMG decomposition, the length of the learning set is always a compromise between the experimental costs and the efficiency of decomposition. Furthermore, all the evaluated NNs have one obvious drawback. Similar to LMMSE, they all need MU firings for learning. This can be avoided with the use of unsupervised type of learning, but to do this, different types of NNs are required.

In conclusion, recently developed NNs are powerful tool for addressing different pattern recognition problems. One such problem is direct estimation of MU firings from HDsEMG. In this study, we tested whether recurrent NNs outperform the non-recurrent ones. The results on HDsEMG signals that were temporarily extended by factor F=10 rejected this hypothesis and demonstrated that the recurrent NNs might be more sensitive to noise than the non-recurrent ones. Further studies are required to better understand the optimal NN architecture and the compromise between the extension factor F, efficiency of HDsEMG decomposition and the size of the learning sets in NNs.

REFERENCES

[1] N. P. Jouppi, et al. "In-datacenter performance analysis of a tensor processing unit.", *ACM/IEEE 44th Annual International Symposium on Computer Architecture*, Jun. 2017.

[2] Y. LeCun, Y. Bengio, and G. E. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436-444, May 2015.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems*, pp. 1097-1105, May 2012.

[4] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, “Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, Nov. 2012.

[5] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, High Performance Convolutional Neural Networks for Image Classification," *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Two*, pp. 1237-1242, Jul. 2011.

[6] Y. Le Cun, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, and D. Henderson, “Handwritten digit recognition with a back-propagation network,” *Advances in neural information processing systems 2*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 396–404, Jun. 1990.

[7] J. Schmidhuber, “Deep learning in neural networks: An overview,” *arXiv preprint*, arXiv:1404.7828, Apr. 2014.

[8] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke and J. Schmidhuber, "A Novel Connectionist System for Unconstrained Handwriting Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855-868, May 2009.

[9] H. Zen and H. Sak, "Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis," *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, QLD, pp. 4470-4474, Apr. 2015.

[10] S. Zhou, K. Yin, F. Fei, K. Zhang, “Surface electromyography–based hand movement recognition using the Gaussian mixture model, multilayer perceptron, and AdaBoost method”, *International Journal of Distributed Sensor Networks*, vol. 15, no. 4, Apr. 2019.

[11] A, Waris, I. K. Niazi, M. Jamil, K. Englehart, W. Jensen, E. N. Kamavuako, “Multiday Evaluation of Techniques for EMG-Based Classification of Hand Motions”, *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, Jul. 2019.

[12] D. Farina and A. Holobar, "Characterization of Human Motor Units from Surface EMG Decomposition," *Proc. IEEE*, vol. 104, no. 2, pp. 353–373, Feb. 2016.

[13] A. Holobar and D. Zazula, "Multichannel Blind Source Separation Using Convolution Kernel Compensation,” *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4487–4496, Sep. 2007.

[14] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[15] D. Farina, L. Mesin, S. Martina, and R. Merletti, “A surface EMG generation model with multilayer cylindrical description of the volume conductor,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, pp. 415–426, Mar. 2004.

[16] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: A system for large-scale machine learning,” *arXiv preprint*, arXiv:1605.08695, May 2016.

[17] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv preprint*, arXiv:1412.6980, Dec. 2014.

[18] J. B. Armstrong, P. K. Rose, S. Vanner, G. J. Bakker, and F. J. Richmond, “Compartmentalization of motor units in the cat neck muscle, biventer cervicis,” *J. Neurophysiol.*, vol. 60, no. 1, pp. 30–45, Jul. 1988.

[19] E. Henneman, “Relation between size of neurons and their susceptibility to discharge,” *Science*, vol. 126, no. 3287, pp. 1345–1347, Dec. 1957.

[20] A. J. Fuglevand, D. A.Winter, and A. E.Patla, “Models of recruitment and rate coding organization in motor-unit pools,” *J. Neurophysiol.*, vol. 70, no. 6, pp. 2470–2488, Dec. 1993.

[21] A. Holobar, M. A. Minetto, and D. Farina, "Accurate identification of motor unit discharge patterns from high-density surface EMG and validation with a novel signal-based performance metric," *J. Neural Eng.*, vol. 11, no. 1, p. 016008, Feb. 2014.