Inhomogeneously Stacked RNN for Recognizing Hand Gestures from Magnetometer Data

Philipp Koch*, Mark Dreier†, Martina Böhme*, Marco Maass*, Huy Phan‡, and Alfred Mertins*
*Institute for Signal Processing, University of Lübeck, Germany
†University of Lübeck, Germany
‡School of Computing, University of Kent, United Kingdom
{koch, boehme, maass, mertins}@isip.uni-luebeck.de,
mark.dreier@student.uni-luebeck.de, h.phan@kent.ac.uk

***Abstract*—Hand gesture recognition systems relying on biosignal data exclusively are mandatory for a variety of applications. In general, these systems have to meet requirements such as affordability, reliability, and mobility. In general, surface electrodes are used to obtain signals caused by the contraction of underlying muscles of the forearm. These data are then used to decode hand gestures. In this work, we evaluate the possibility of replacing the electrodes by magnetometers that are cheap and can be easily implemented in mobile devices. We propose an inhomogeneously stacked recurrent neural network for classifying hand gestures given magnetometer data. The experiments reveal that the comparably small network significantly outperforms state-of-the-art hand gesture recognition systems relying on multi-modal data. Furthermore, the proposed network requires significantly shorter windows and enables a quickly responding classification system. Also, the experiments show that the performance of the proposed system does not vary much between subjects and works outstandingly for amputees.**

***Index Terms*—hand movement classification, magnetometer, recurrent neural network, hand prosthesis**

I. Introduction

In recent years, the interactions between humans and computers or robots have become more and more important in both the industry as well as the private life. Consequently, the field of human-machine-interfaces (HMI) has gained increasing interest. Often humans can interact with a machine in a very intuitive fashion via hand gestures [1], [2]. To enable such interaction, a robust hand gesture recognition system is required. To this end, typically, data gloves and camera based hand tracking systems are used. However, both techniques often have significant disadvantages such as high cost, the lack of robustness and accuracy, as well as lack of mobility and availability. Furthermore, these systems cannot be used by people with an amputation or a paralysis of the hand. Thus, they are not suitable in the medical field where such systems are required, e.g., to control hand prosthesis [3] or exoskeletons [4], [5]. Hand gesture recognition systems for medical applications usually rely on surface electromyography (sEMG) signals only. However, sEMG systems are complicated and expensive. Consequently, a cheap but reliable hand gesture recognition system that can be included in embedded/mobile systems would be desirable. Moreover, such a system should be able to recognize a vast variety of different hand gestures with small time delay and to detect even intentional hand movements of amputees.

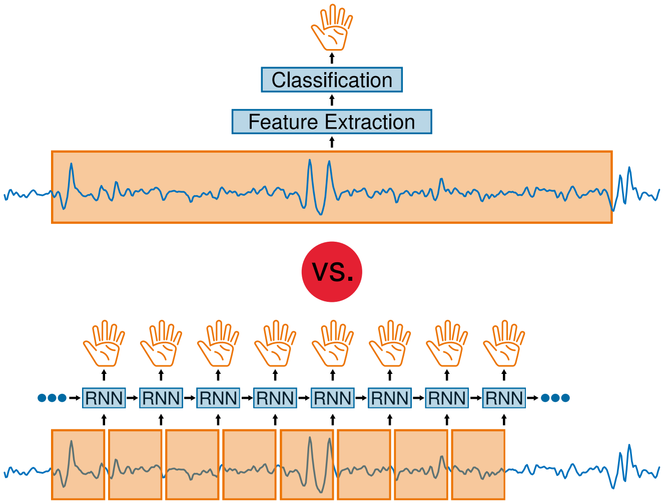


Fig. 1. Comparison of the standard classification approach (above) and the classification with an inhomogeneously stacked RNN (below).

In this work, we evaluate the possibility of solely using magnetometer data for classifying hand gestures. Magnetometers are cheap and can be easily integrated into an embedded system. Furthermore, they can be placed around the forearm like electrodes to measure changes in the magnetic field.

Considering sEMG signals, the most hand gesture recognition approaches follow a standard classification pipeline, i.e., the extraction of hand-crafted features followed by the classification with a conventional classifier such as a support vector machine or a random forest [6]–[8]. Lately, more enhanced techniques have been used such as deep learning including convolutional neural networks (CNNs) [9], [10] and recurrent neural networks (RNNs) [11], [12]. CNNs are designed to learn a feature extractor and a classifier within one network in a data-driven fashion. Consequently, they are suitable to avoid a hand-crafted feature extraction and enable the detection of hand gestures from raw data. For sEMG data, end-to-end trained CNNs have been identified as promising approaches for hand gesture classification systems [9], [10].

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

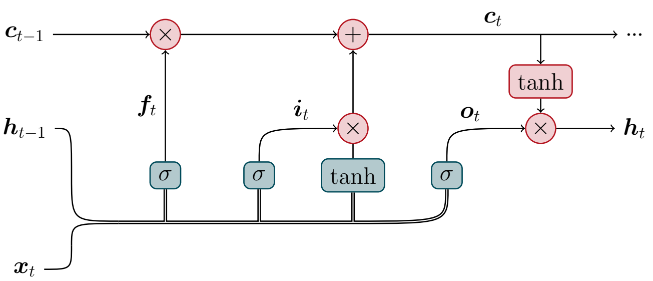


Fig. 2. Illustration of an LSTM cell.

Following the success of CNNs, RNNs have been shown to hold a great potential for recognizing hand gestures given sEMG signals [11], [12]. The main advantage of RNNs is that they can make use of the sequential nature of signals. They have been reported to achieve state-of-the-art performance with comparatively small network architectures.

To leverage the respective advantages of CNNs and RNNs within one network, we propose an inhomogeneously stacked RNN consisting of standard long short-term memory (LSTM) cells [13] and convolutional LSTM (ConvLSTM) cells [14]. This network architecture enables us to classify hand gestures based on very short windows ($5\ \mathrm{ms}$), to avoid large delays and to update the hand gesture classification quickly. In Fig. 1 the proposed method is compared with a standard classification system which usually requires at least $200\ \mathrm{ms}$ long windows.

To study whether hand gestures can be decoded from data of multiple tri-axial magnetometers placed around the forearm using this inhomogeneously stacked RNN we conducted experiments on a publicly available database containing data of both amputees and able-bodied subjects. The results indicate that this specialized network architecture is efficient for classifying hand gestures from magnetometer data. We outperform state-of-the-art systems significantly.

II. Inhomogeneously Stacked RNN Architecture

As LSTM cells overcome the problem of vanishing gradients, they can capture long-term dependencies very well. Therefore, they are suitable for the classification of sequential data. ConvLSTM cells combine the feature extraction ability of CNNs with the sequence processing ability of RNNs. For these reasons, in this work, we use both standard LSTM cells and ConvLSTM cells. In the following sections, we elaborate the LSTM cell and the ConvLSTM cell. The former is the center of the network while the later is the key element of the initial feature extraction part of the network's architecture. Finally, the whole network architecture is described in detail.

A. Standard LSTM

The general principle of an LSTM cell is illustrated in Fig. 2. The corresponding hidden layers $\mathcal{H}$ are nonlinear transformations given by

$$(\boldsymbol{h}_t, \boldsymbol{c}_t) = \mathcal{H}(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}, \boldsymbol{c}_{t-1}), \tag{1}$$

where t is the current time step regarding the given input sequence, $\boldsymbol{h}$ the output vector, $\boldsymbol{c}$ the cell state, and $\boldsymbol{x}$ the input vector. At each time step, the cell state is updated by

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tanh\left(\boldsymbol{W}_{xc}\boldsymbol{x}_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c\right) \tag{2}$$

with $\odot$ being the Hadamard product, the input gate defined as

$$\boldsymbol{i}_t = \sigma\left(\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i\right), \tag{3}$$

and the forget gate being

$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right). \tag{4}$$

All trainable weight matrices are denoted by $\boldsymbol{W}_{\cdot}$ while the trainable bias vectors are represented by $\boldsymbol{b}_{\cdot}$. The actual output of an LSTM cell is calculated via

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh\left(\boldsymbol{c}_t\right) \tag{5}$$

with $\boldsymbol{o}$ being the output gate which is defined as

$$\boldsymbol{o}_t = \sigma\left(\boldsymbol{W}_{xo}\boldsymbol{x}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o\right). \tag{6}$$

B. ConvLSTM with Specialized Padding

The proposed network is used to recognize hand gestures in a sequence of small windows containing data of multiple tri-axial magnetometers placed on the subject's arm. In order to exploit the spatiotemporal structure of the magnetometer data, ConvLSTM cells are used. A ConvLSTM cell takes multi-dimensional data as input and allows to incorporate the spatial relations between sensors into the feature extraction. The spatial and temporal information are exploited by convolution. The ConvLSTM's hidden layer $\mathcal{H}^{\text{conv}}$ is given by

$$(\boldsymbol{H}_t, \boldsymbol{C}_t) = \mathcal{H}^{\text{conv}}(\boldsymbol{X}_t, \boldsymbol{H}_{t-1}, \boldsymbol{C}_{t-1}) \tag{7}$$

and can be described by the following five equations:

$$\begin{aligned}\boldsymbol{C}_t =& \boldsymbol{I}_t \odot \tanh\left(\boldsymbol{W}_{XC} * \boldsymbol{X}_t + \boldsymbol{W}_{HC} * \boldsymbol{H}_{t-1} + \boldsymbol{B}_C\right)\\ &+ \boldsymbol{F}_t \odot \boldsymbol{C}_{t-1},\end{aligned} \tag{8}$$

$$\boldsymbol{I}_t = \sigma\big(\boldsymbol{W}_{XI} * \boldsymbol{X}_t + \boldsymbol{W}_{HI} * \boldsymbol{H}_{t-1} + \boldsymbol{B}_I\big), \tag{9}$$

$$\boldsymbol{F}_t = \sigma\big(\boldsymbol{W}_{XF} * \boldsymbol{X}_t + \boldsymbol{W}_{HF} * \boldsymbol{H}_{t-1} + \boldsymbol{B}_F\big), \tag{10}$$

$$\boldsymbol{O}_t = \sigma\big(\boldsymbol{W}_{XO} * \boldsymbol{X}_t + \boldsymbol{W}_{HO} * \boldsymbol{h}_{t-1} + \boldsymbol{B}_O\big), \tag{11}$$

$$\boldsymbol{H}_t = \boldsymbol{O}_t \odot \tanh\big(\boldsymbol{C}_t\big). \tag{12}$$

In the above equations, $*$ denotes the convolution operation and all trainable filter kernels and bias matrices are denoted by $\boldsymbol{W}_{\cdot}$ and $\boldsymbol{B}_{\cdot}$, respectively.

In ConvLSTM cells, zero padding is usually used to avoid shrinking of the convolution results. Since we process small matrices, the kind of padding is important. The electrodes are padded in a cyclic fashion. For the time dimension of the input

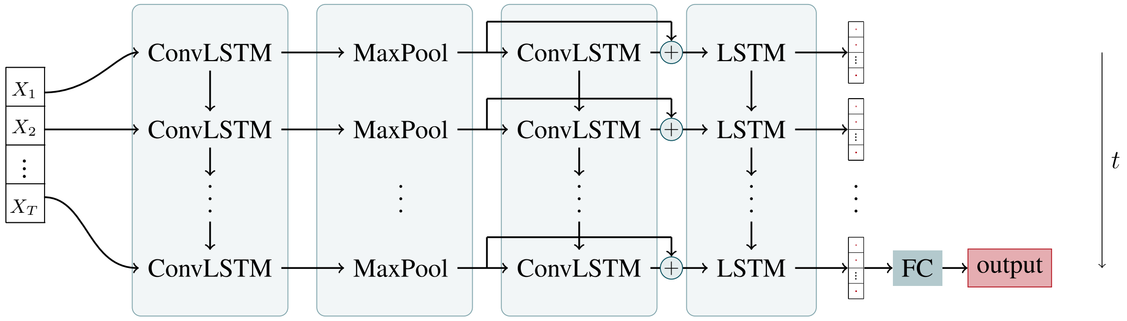


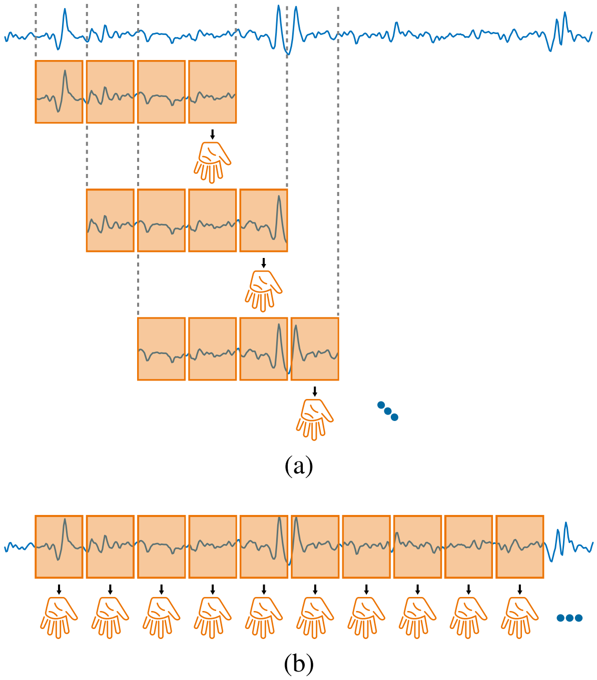
Fig. 3. Illustration of the inhomogeneously stacked RNN enrolled over time in training configuration. The first ConvLSTM cell is followed by a two-dimensional max-pooling layer (MaxPool). The actual hand gesture classification is accomplished using a fully-connected layer (FC).

data a symmetric padding is used. Consequently, the padded input for a 3×3 convolution kernel has the form:

$$\left[\begin{array}{c|cccccc|c}
x_{1,12} & x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,12} & & x_{1,1} \\
\hline
x_{1,12} & x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,12} & & x_{1,1} \\
x_{2,12} & x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,12} & & x_{2,1} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\
x_{10,12} & x_{10,1} & x_{10,2} & x_{10,3} & \cdots & x_{10,12} & & x_{10,12} \\
\hline
x_{10,12} & x_{10,1} & x_{10,2} & x_{10,3} & \cdots & x_{10,12} & & x_{10,12}
\end{array}\right].$$

C. Network Architecture

The presented architecture is an inhomogeneously stacked RNN comprised of three RNN cells and a fully-connected layer as depicted in Fig. 3. The first layer is a ConvLSTM cell taking a sequence of three-dimensional matrices as input. Thereby, the first dimension corresponds to the samples of the windows, the second to the signals from the different magnetometers, and the third one to the three axes of the magnetometers. Within this ConvLSTM cell, 24 filter kernels with a kernel size of 3×3 are learned. In order to reduce complexity and to add nonlinearity to the network, the output of the first cell is fed to a max-pooling layer, where a kernel of size 2×2 and strides of 2 are used. This pooling layer is followed by another ConvLSTM cell associated with 24 filter kernels of size 3×3. We implement a residual connection [15] by adding the output of the max-pooling layer to the output of the second ConvLSTM cell. The result of the summation is transformed into a vector and fed to a standard LSTM cell with a state size of 256. To perform the final classification, the last layer in the network is a fully-connected layer with a softmax activation function.


(a)

(b)

Fig. 4. Illustration of the training and test data generation. In (a) the extraction of overlapping sequences for the training procedure is illustrated. During training, only the hand gesture corresponding to the final window of the training example sequence is predicted. In (b) the test case is shown, where for each window a hand gesture is predicted.

III. Network Training and Validation

In order to have a sufficient number of training examples, the training data were augmented using overlapping sequences as can be seen in Fig. 4(a). Each sequence was subdivided into T windows of 5 ms duration. For each training example sequence, a single label was predicted by the network. To calculate the loss needed for the optimization of the network parameters, only the classification of the last window of the training sequence was taken into account. As a loss function the cross-entropy was chosen which is given by

$$E\left(\Theta|\boldsymbol{X}, \boldsymbol{y}_T\right) = -\boldsymbol{y}_T \log\left(\hat{\boldsymbol{y}}_T\left(\Theta|\boldsymbol{X}\right)\right), \qquad (13)$$

where $\hat{\boldsymbol{y}}_T$ denotes the network's classification result of the last window T and $\boldsymbol{y}_T$ the ground-truth corresponding to the final time step T represented in a one-hot encoded vector. The parameters of the RNN are given by Θ and the input sequence is represented by $\boldsymbol{X}$. In addition, we applied a ℓ_2-norm regularization R_{ℓ_2} to the convolution kernels of the two ConvLSTM layers to prohibit the uncontrolled growth of their weights. The overall loss function reads

$$C\left(\Theta|\boldsymbol{X}, \boldsymbol{y}_T\right) = E\left(\Theta|\boldsymbol{X}, \boldsymbol{y}_T\right) + \alpha R_{\ell_2}\left(\Theta^{Kernels}\right), \qquad (14)$$

TABLE I
RESULTS FOR THE INHOMOGENEOUSLY STACKED RNN. THE ACHIEVED RESULTS ARE COMPARED WITH THOSE OF STATE-OF-THE-ART METHODS DESCRIBED IN [18] THAT RELY ON ALL IMU MODALITIES AND ON A COMBINATION OF IMU AND SEMG SIGNALS, RESPECTIVELY. IN ADDITION, THE PERFORMANCE OF AN RNN CONSISTING OF A SINGLE LSTM CELL WITH A STATE SIZE OF 256 IS SHOWN.

Modality / Method	able-bodied	amputated
IMU / [18]	81.7 %	77.7 %
IMU & sEMG / [18]	82.7 %	77.8 %
Magnetometer / single LSTM cell	86.5 %	82.6 %
Magnetometer / proposed RNN	89.2 %	83.1 %

where α represents a weighting factor and $\Theta^{Kernels}$ the convolution kernels of the ConvLSTMs.

For training, we chose sequences of $1000\ \mathrm{ms}$ and a window size of $5\ \mathrm{ms}$. As optimizer we used the Adam [16]. Furthermore, we applied dropout [17] with a dropout probability of $50\ \%$ and used mini-batch training with a batch size of 200.

Unlike for training, during network testing, the test examples were not subdivided into shorter sequences of fixed length but processed as entire sequences (see Fig. 4(b)). Consequently, the sequence length varied for the test examples. In the test case, the network was used to predict the hand movement category for each window of the presented sequence. We chose this setting for evaluating the network's performance to keep it as similar as possible to the actual application of a hand gesture classification system. In such a system, from the moment it is deployed, the classifier has to recognize the hand gestures in each window (of a steadily growing sequence). The performance of the network was evaluated using the accuracy calculated by comparing the predicted class of each window with its corresponding ground-truth.

IV. EXPERIMENTS

A. Database

To evaluate the possibility of recognizing hand gestures from magnetometer data and to validate the inhomogeneously stacked RNN, experiments on the database DB7 of the Ninapro project were conducted [18]. This is a publicly available database that includes data of sEMG and magnetometer. Using this database allows us to compare the performance of the proposed system with other approaches relying, e.g., on sEMG or inertial measurement unit (IMU) signals.

The database contains recordings of 20 able-bodied subjects and 2 amputees. During the recording sessions, the subjects were asked to perform 40 (excluding rest position) hand gestures by presenting them on a screen. The subjects had to repeat every hand movement six times. Between the repetitions, the subjects were asked to place their hand in a resting position. During the experiments, the Delsys® Trigno™IM Wireless System was used for recording the signals. The system's sensors included both an electrode and an IMU. The sEMG data acquired by the electrodes were sampled with $2\ \mathrm{kHz}$ while the IMU data were recorded with a sampling frequency of $128\ \mathrm{Hz}$. To meet the sampling frequency of the sEMG signals, the IMU data were upsampled. The IMU in the sensor includes a tri-axial magnetometer whose signals are used in this work.

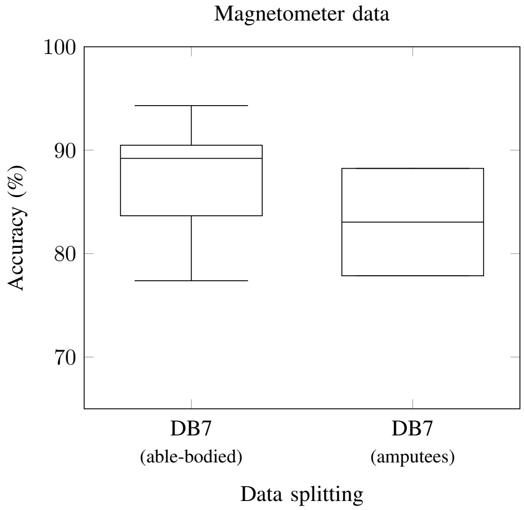


Fig. 5. Boxplot of the obtained results.

B. Data Preprocessing

For comparability with other publications, we followed the suggestions in [18] regarding the generation of training and test sets. In order to prepare the magnetometer data for classification with the proposed RNN network, the data were preprocessed by performing a channel-wise normalization of the magnetometer data. To this end, the mean and the standard deviation were calculated for each axis of the magnetometer individually. Each channel was normalized by subtracting its corresponding mean and dividing it by the standard deviation in order to achieve zero mean and unit standard deviation. Note that all necessary statistics were obtained using training data only. As mentioned previously, for network training, overlapping fixed-length sequences of $5\ \mathrm{ms}$ long windows were generated. For the test case, sequences of different length were produced.

C. Results

For the purpose of evaluation, we split the subjects of DB7 into two groups: the able-bodied subjects and the amputees. For each group we report the median accuracy achieved across all subjects of the group.

In Table I the results of the proposed network are reported and compared with results for the state-of-the-art methods that use signals of multiple modalities and a simple RNN. The state-of-the-art methods follow the standard classification pipeline featuring hand-crafted features and a linear discriminant analysis classifier. The RNN is similar to the inhomogeneously stacked RNN but has a single LSTM cell with a state size of 256. As can be seen, the proposed approach outperforms the IMU based approach [18] by $7.5\ \%$ and $5.4\ \%$ for the amputees and able-bodied subjects, respectively.

This is even more remarkable knowing that this approach not only takes the magnetometer but also the gyroscope and the accelerometer signals into account and, moreover, requires $256\ \mathrm{ms}$ instead of $5\ \mathrm{ms}$ long windows. Note, that the results obtained on magnetometer data by [18] are even worse. Besides, the proposed approach even outperforms the IMU and sEMG based state-of-the-art classification system—having similar properties to the proposed approach—significantly by $6.5\ \%$ and $5.3\ \%$ for able-bodied and amputated subjects, respectively. The state-of-the-art system relying on sEMG data exclusively is even surpassed by about $29\ \%$ for able-bodied subjects and $40\ \%$ for the amputees. As can be seen in the boxplot in Fig. 5, the performance of the proposed system does vary much across subjects. In contrast, the performance of state-of-the-art hand gesture classification systems appears to depend more on the individual subject. The individual classification accuracy for several subjects is roughly $20\ \%$ lower then the median accuracy calculated across all subjects. Consequently, our approach is potentially more robust to inter-individual differences.

By considering the results of the RNN based on a single LSTM cell and the ones achieved by the inhomogeneously stacked RNN it becomes clear that RNNs are highly suitable for recognising hand gestures based on analysing magnetometer data. Both networks can reliably detect hand movements even on the small $5\ \mathrm{ms}$ windows. However, the results show that the more complex RNN including two different kinds of RNN cells for feature extraction and further analysis outperforms a simple RNN.

The presented results indicate the possibility of replacing the expensive sEMG electrodes by magnetometers or at least adding magnetometers and try to reduce the number of electrodes. However, the database used in this experiments does not include recordings of multiple days, consequently it has to be studied whether a system relying on magnetometer data and an RNN based classifier performs similarly well in such settings.

Overall, the results reveal that the proposed inhomogeneously stacked RNN hold promise in classifying a vast variety of different hand gestures from magnetometer data with good robustness and accuracy. Furthermore, the results indicate that magnetometers capture similar or even more information as electrodes for hand gesture recognition.

V. Conclusions

In this work, we investigated whether hand gestures can be recognized using the magnetometer signals with an RNN. We proposed a network architecture containing ConvLSTM and LSTM cells allowing us to exploit the temporal and the spatial information within the magnetometer data. The proposed approach leads to significantly better performance than state-of-the-art systems based on multi-modal data while requiring windows that have a fraction of the length. The classification accuracy is improved by more than $5\,\%$ for both able-bodied subjects and amputees while using $5\ \mathrm{ms}$ long windows. Furthermore, the experimental results reveal that the proposed system is comparably robust to inter-individual variance and works well for all individuals. The promising results indicate the possibility of replacing the expensive electrodes by magnetometers since they are able to capture information useful for hand gesture recognition. Furthermore, adding magnetometers would be a cheap solution for acquiring more data in a hand gesture recognition system to achieve a more robust hand gesture classification.

References

[1] J. Cheng, X. Chen, Z. Lu, K. Wang, and M. Shen, "Key-press gestures recognition and interaction based on sEMG signals," in *Proc. Int. Conf. Multimodal Interf. and Mach. Learn. Multimodal Interact.*, 2010.

[2] F. Muri, C. Carbajal, A. M. Echenique, H. Fernández, and N. M. López, "Virtual reality upper limb model controlled by EMG signals," *J. Phys. Conf. Ser.*, vol. 477, 2013.

[3] C. Cipriani, F. Zaccone, S. Micera, and M. C. Carrozza, "On the shared control of an EMG-controlled prosthetic hand: Analysis of user-prosthesis interaction," *IEEE Trans. Robot.*, vol. 24, no. 1, pp. 170–184, 2008.

[4] J. Rosen, M. Brand, M. B. Fuchs, and M. Arcan, "A myosignal-based powered exoskeleton system," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 31, no. 3, pp. 210–222, 2001.

[5] K. Kiguchi and Y. Hayashi, "An EMG-based control for an upper-limb power-assist exoskeleton robot," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1064–1071, 2012.

[6] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. Mittaz Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 140053, 2014.

[7] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[8] 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, 2003.

[9] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Front. Neurorobot.*, vol. 10, no. 9, 2016.

[10] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 36571, 2016.

[11] P. Koch, H. Phan, M. Maass, F. Katzberg, and A. Mertins, "Recurrent neural network based early prediction of future hand movements," in *Proc. IEEE Eng. Med. Biol. Soc. (EMBC)*, July 2018.

[12] P. Koch, H. Phan, M. Maass, F. Katzberg, R. Mazur, and A. Mertins, "Recurrent neural networks with weighting loss for early prediction of hand movements," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, September 2018.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. IEEE Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 802–810.

[15] Y. Wang and F. Tian, "Recurrent residual learning for sequence classification," in *Proc. Conf. Empir. Methods Nat. Lang. Process.*, 2016, pp. 938–943.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[18] A. Krasoulis, I. Kyranou, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements," *J. Neuroeng. Rehabil.*, vol. 14, no. 71, 2017.